Schaum series statistics and econometrics pdf book download online full

is a poiItt estimate d U . (b) Which are the endogenous and exogenous variables of the system? AUTOCORRELATION 9.23 Table 9.19 gives fixed private investment Y , GDP X 1 , both seasonally adjusted in billions of dollars, and the commercial paper interest rate X2 for the United States from 1982 to 1999. 12.7, a multiple regression was estimated using Excel. Similarly, the area between z = 0 and z = 2 is 0.4772, or 47.72% (by looking up z = 2.00 in the table), so that the area between z = &2 is 95.44% (see Fig. (a) In order to remove the influence of X2 on Y , we regress Y on X2 and find the residual el = Y*. (b) The producer may instead be interested in testing if the breaking strength of the cable exceeds 5000 lb . (b) This is a two-tail test with nl + n 2 - 1 = 49 degrees of freedom. (a) The relative frequency or empirical probability of the 3 is given by the ratio of the number of times 3 comes up (106) out of the total number of times the die is rolled (600). If the parent population is normal, the sampling distributions of the mean are also normally distributed, wet h sillall samples. Some examples are the IQs (intelligence quotients), weights, and heights of a large number of people and the variations in dimensions of a large number of parts produced by a machine. Med = L+- n/2 - F c = $3.90 fm +-25125 - 9 (0.10) = $3.90 + $0.07 = $3.97 as compared with the true median of $3.95 found from the ungrouped data (see part a). (c) Moving down the columns headed 0.05, 0.025, and 0.005 in App. Unit-root testing suffers from being a low-power test in that it seldom rejects a unit root when it should. 3) ~ = 0.0081 P(4) = 4!(4 - 4) ! (. A Y. The disadvantages of the (c) The advantages of the mode are the same as those for the median. 6 7 6.26 (a) What is meant by the best unbiased or efficient estimator? (b) The exact linear relationship in Eq. (6.1) can be made stochastic by adding a random disturbance or error term, ui, giving (c) Most observed values of Y are not expected to fall precisely on a straight line (1) because even though consumption Y is postulated to depend primarily on disposable income X, it also may depend on SIMPLE REGRESSION ANALYSIS [CHAP. (b) The standard error of the mean a is given by the standard deviation of the parent population a divided For finite populations of size N, a3nite by the square root of the samples' size Jt;:that is, ag = a/,&. Descriptive statistics is concerned with g and describing a body of data. Thus STATISTICAL INFERENCE: ESTIMATION CHAP. Why is this important? No Table 5.34 Democrats and Republicans below and above Age 40 Age Group 16 14 30 122 STATISTICAL INFERENCE: TESTING HYPOTHESES CHAP. 21 21 DESCRIPTIVE STATISTICS Thus 7 calculated from the grouped data is only a very good approximation for the true value of 7 calculated for the ungrouped data. A n s (a) P(Xl < X < X2) =:J f(x) d~ (6) f(X) = (l/=) exp[-(l/2)[(-- / a 2] (c) E(X) = :J Xf (X) dX and a2=:J [X - E (x)]~f (x) dX (d) f (X) = (11 6)exp[-(112)~] (e) E(X) = p = 0 and a2 = 1 3.65 Find the area under the standard normal curve (a) within z k 1.64, (b) within z = k1.96, (c) within z = k2.58, (4 between z = 0.90 and z = 2.10, (e) to the left of z = 0.90, (f) to the right of z = 2.10, (g) to the left of z = 0.90 and to the right of z = 2.10. It is bounded by 0 and 1, but achieves 1 only in the limit, and rarely takes on large values. b b b 6 6 6 6 6 b L. How: More than 4? 1 8 = 19.980~ As noted in Prob. 61 137 SIMPLE REGRESSION ANALYSIS (d) Taking the sum of the absolute deviations avoids the problem of having the sum of the deviations equal to 0. Quartile deviation measures the average range of one-fourth of the data. This is given by where X is the mean number of occurrences for the interval of interest and e-A can be obtained from App. (4.2a), (4.9a), and (2.8b)l X where X j is a sample or column mean, is the grand mean, r is the number of observations in each sample, and c is the number of samples. 11.4 TESTING FOR UNIT ROOT Stationary and nonstationary series can follow different patterns, many of which look similar when graphed. In simple random sampling, not only each item in the population but each sample has an equal probability of being picked. (6.25) x) C yi = n ci C xi = (b) C xi - ci) Subtracting Eq. (6.25) from Eq. (6.24), we get [illegible] decreases with X,. (4 We saw in part c that Cui = 0 and E c : = 1 / E x : . (b) What is the meaning of the slope? [illegible] = 72 SSE = SST - SSA - SSB = 266 - 156 - 72 = 38 [illegible] This most popular instrumental variable is the lagged value of the explanatory variable X itself. The book is aimed at college students in economics, business administration, and the social sciences taking a onesemester or a oneyear course in statistics andlor econometrics. The occurrence of one does not preclude the occurrence of the other(s). 8 The data are first transformed into natural log form, as shown in Table 8.7, and then the OLS method is applied to the transformed variables as explained in Sec. However, dropping either In K or In L from the multiple regression leads to a biased Table 9.6 Output, Labor, and Capital Inputs in Original and Log Form [illegible] PROBLEMS IN REGRESSION ANALYSIS OLS slope estimate for the retained variable because economic theory postulates that both labor and capital should be included in the production function. This may seldom be the case in the real world, thus requiring a more flexible lag scheme (see Prob. [illegible] (b) Draw a normal curve in the f and z scales and show the percentage of the area under the curve within 1,2, and 3 standard deviation units of its mean. If n(1 - p) < 5, success and failure should be redefined so that np < 5 to make the approximation accurate. (a) Table 6.5 shows the calculations to find land/or the data in Table 6.4. [illegible] Thus the equation for the estimated consumption regression is [illegible] (b) Using a table of random numbers, obtain a random sample of 10 from the 95 employees of a plant that were sick over a particular day. 77 Consistency, 1:483 Consistent estimators, 134, 146:149, 186:187 Contingency-table tests, 9&92 Continuity, correction for, 92 4142, 5762,6 Correlation, coefficient of: rank. (g) The odds of picking a blue ball are given by the ratio of the number of ways of picking a blue ball to the number of ways of not picking a blue ball. SIMPLE REGRESSION ANALYSIS A A Y, = bo [CHAP. [illegible] 8.33 Table 8.20 gives the per capita disposable income Y in thousands of dollars and the percentage of college graduates in the population 25 years of age or older X for the eastern United States in 1998. [illegible] 8), there is evidence of autocorrelation. EXAMPLE 3. [illegible] 5.49 A random sample of 25 cities in the United States indicates the number of rainy days during the month of June given in Table 5.32. 5-7 5.12 (a) Draw a figure for the answers to Prob. If the null is rejected, the data are stationary and can be used without differencing. 21 2.25 DESCRIPTIVE STATISTICS Compute the covariance from Table 2.26 using the alternate formula. What is the probability that a component picked at random from the production line will have a lifetime (a) Between 1120 and 1180 h? [illegible] The estimated coefficients [illegible] tha? What is the minimum sample size required? That is, E(bo) = bo, E(bl) = bl, and E(b2) = b2, and sg0' sg,, and sb2- .are lower than for any other unbiased linear estimator. (b) Three trainee positions eventually if the positions differ? Thus Ho is rejected and H I is accepted, so that the firm can claim at the 95% level of confidence that the detergent from plant 1 contain more than 1000 g of detergent. (a) Does the DJIA have a unit root? (c) Find the expected value and standard deviation for [illegible] (a) If the parent population is normally distributed, the theoretical sampling distributions of the mean are also normally distributed, regardless of sample size. [illegible] The regression line represents the best fit to the random sample of wmsumption/disposable income observations in the sense that it minimizes the sum of the squared (vertical) deviations from the line. [illegible] "statistically significant at customary levels," and that the econometric assumptions of the model are satisfied 7 CHAPTER 2 A Descriptive Statistics 1 FREQUENCY DISTRIBUTIONS to organize or arrange a body of data into frequency distribution. The moving-average process, on the other hand, carries forward E,, the random component of y,, so previous observations are not perpetuated. (a) Calculate the covariance between X and Y. 10, indicate (a) what the first step is in econometric research and (b) what the a priori theoretical expectations are of the sign and possible size of the parameters of the demand function given by Eq. (1.4). For example, the relative frequency or empirical probability might be 0.517 or 1000 tosses, 0.508 for 10,000 tosses, and so on. [illegible] 7 for a! = 0.05 and c - 1 = 3 df in the numerator and (r - 1)c = 28 df in the denominator is 2.95. [illegible] Two events, A and B, are i n d e p e n d e n t if the occurrence of A is not connected in any way to the occurrence of B. 8-3, the maximum point of the log-likelihood function is at bl = 0.001, which was the estimated value from Example 8.5. 8.23 Estimate Example 8.5 using the logit model. [illegible] However, with 2SLS estimation (as opposed to ILS), we also get the standard errors of the estimated structural parameters directly. [illegible] (d) Test for the presence of heteroscedasticity. Table 2.17 Calculations for the Average Deviation for the Grouped Data In Table 2.12 Hourly Wage, Class Midpoint Frequency [illegible] DESCRIPTIVE STATISTICS [CHAP. [illegible] FURTHER TECHNIQUES AND APPLICATIONS IN REGRESSION ANALYSIS Table 8.16 Inventories, Sales, and Z Values in U.S. Manufacturing, 1981-1999 (in Billions of Dollars) [illegible] The kurtosis for a mesokurtic curve is 3. [illegible] For example, on 1 roll of a fair die, we have 6 mutually exclusive outcomes (1, 2, 3, 4, 5, or 6), each associated with a probability occurrence of 116. [illegible] (b) How many defective items do you expect out of the 1600 items produced each day? (b) The classical or a priori approach to probability can only be applied to games of chance (such as tossing a balanced coin, rolling a fair die, or picking cards from a standard deck of cards) where we can determine a prior, or without experimentation, the probability that an event will occur. [illegible] the 5% level of significance. [illegible]

8.38 85.71 84.13 1.58 100 99.61 0.39 The maximum difference is 26.30% (0.2630), which is less than the critical value of 0.410; therefore we accept the null hypothesis that sales are normally distributed with a mean of 10 and standard deviation of 3. Should the hypothesis be accepted or rejected at the 5% level of significance that the production of acceptable items is independent of the hour of the morning in which they are produced? The minimum acceptable number of observations for most practical purposes is 30. 11-4, however, gives a picture that does not look stationary. 51 STATISTICAL INFERENCE: TESTING HYPOTHESES preceding analysis implicitly assumes that the effects of the two factors are additive (i.e., there is no interaction between them). = (8 - 2)(8)(34) c e: 2 = (n - k) C x: i0 6.44 to = - = -= S2.08 ha - 12.9416 and = (8 - 2)(34) - 3.10 il 0.82 t l = - = -% 3.28 sdl 0.25 s$=2.67 " 0.06 and and s - % 2.08 % 0.25 ha sil and is significant at the 5% level and is also significant at the 5% level s~g1.63 Therefore, the 95% confidence or prediction interval for Y F is given by Y F = 14.64 f 2.45(1.63), where to,o2s= f2.45, with n - k = 8 - 2 = 6 df, so that we are 95% confident that 10.65 5 Y F 5 18.63. The economist uses it to test the efficiency of alternative production techniques; the businessperson may use it to test the product design or package that maximizes sales; the sociologist to analyze the result of a drug rehabilitation program; the industrial psychologist to examine workers' responses to plant environment; the political scientist to forecast voting patterns; the physician to test the effectiveness of a new drug; the chemist to produce cheaper fertilizers; and so on. 9 (4 Since ESS2/ESSI = 7.08 exceeds Fll,ll= 2.82 at the 5% level of significance, heteroscedasticity is present. Joint moment. When this is done, bl is consistent. Table 2 9 Frequency Distribution of Grades Grade 1.52.4 2.S3.4 3.54.4 4.S5.4 5.56.4 6.S7.4 7.58.4 8.S9.4 9.S10.4 Class Midpoint Absolute Frequency Relative Frequency Cumulative Frequency 2 3 4 5 6 7 8 9 10 3 3 5 5 6 4 4 2 4 0.075 0.075 0.125 0.125 0.150 0.200 0.100 0.100 3 6 11 16 22 30 34 38 -2 o.050 40 40 1.000 8 (c) See Fig. 8), there is no evidence of autocorrelation. Wilcox. 61 Fig. (a) What is the probability of 3 customers stopping in any hour? The fourth assumption is E(u,uj) = 0 for i j; and the fifth assumption is E(X,u,) = 0 . 5.19 Table 5.1 1 gives the distribution of the number of acceptances of 100 students into 3 colleges. The only difference is that now we are dealing with In addition, before the distribution of the x s rather than with the distribution of the xs. (a) ln L = -36.59 (b) ln L = -9.70 (c) ln L = -6.91 8.40 Calculate the log-likelihood values for the logit model in Prob. 71 7.28 MULTIPLE REGRESSION ANALYSIS For the regression in Prob. (b) Is the intercept significantly different in wartime than in peacetime? The expected value and variance are (a) For the statement of Prob. (c) Chebyshev's theorem? EXAMPLE 6. However, with measurement errors in the explanatory variables, the fifth of the OLS assumption of independence of the explanatory variables and error term is violated (seeprob. 6.30. Find an interval estimate for p such that we are 90% confident that the interval includes p. (b) Does X Granger-cause Y with six lags? 11 COINTEGRATION AND ERROR CORRECTION 11.15 (a) What is cointegration? b6 Fig. P (3) = 0.08928 (from part a) Thus P (X 5 3) = 0.00248 + 0.01488 + 0.04464 + 0.08928 = 0.15128 (c) The expected value, or mean, of this Poisson distribution is X tion is f i = 4 % 2.45 customers. It is the unbiased estimator with the mast m p a c t or least spread-out distribution. However, deductive and inductive statistics are complementary. + + (d) Partial correlation coefficients are used in multiple regression analysis to determine the relative importance of each explanatory variable in the model. 3-13). (b) Draw a figure for Eqs. This involves OLS estimation of the R,reduced-form equation and then use of R, to estimate the Y structural equation. As a result, Y and the sampling distribution of the parameters of the regression are also normally distributed, so that tests can be conducted on the significance of the parameters (see Secs. 4.43 if n < 30? Then the hypergeometric distribution is used. 8.11 (a) Dcrivc thc Koyck distributcd lag modcl. 4-1). If the variables suggested by economic theory do not provide a satisfactory explanation, the researcher may experiment with alternative formulations and variables suggested by previous tests or opposing theories. The first stage is The second stage is GI = 0.1 153 is a consistent estimate of a l . (b) Blue? (b) In what way are t values interpreted differently from z values? 1 7 9 ~=~0.45 ~ f 2.179(0.10) = 0.45 f 0.22 So b2 is between 0.23 and 0.67 (i.e., 0.23 5 b2 5 0.67) with 95% confidence. For example, when Xi = 114, Y i = 2.30 0.86(114) = 100.34. Thc sum of thc probabilitics associated with all the values that the discrete random variable can assume always equals 1. When Xi = 178, Y i = 2.30 0.86(178) = 155.38. By convention, this is done whenever n < 0.05N. 5.23(c)] SST = C C(Xij - F)' = (51 - 55)' + (47 - 55)' = SSA SSE = 112 + 726 = 838 + + . You're Reading a Free Preview Pages 105 to 121 are not shown in this preview. 11 As can be seen, an AR(1) process contains some part of each previous error term. (a) The binomial distribution is used to find the probability of X number of occurrences or successes of an event, P (X) , in n trials of the same experiment when (1) there are only 2 mutually exclusive outcomes, (2) the n trials are independent, and (3) the probability of occurrence or success, p, remains constant in each trial. ESTIMATION: TWO-STAGE LEAST SQUARES 10.13 When can 2SLS be used? From past experience, the firm knows that the amount of detergent in the boxes packed in the two plants is normally distributed. Then -- - covariance of X and Y and 6.22 Find R2 for the estimated consumption regression of Prob. (b) One-half hour will pass? Most of the applications in econometrics involve first rather than second- or higher-order autocorrelation. 8.29 + + Fit a double-log function to the data in Table 6.12. (c) A green or white ball from the urn of Prob. z=- x-py - ax x-p a/Jt; and is found in App. (b) Answer the same question if the population standard deviation had not been known, but the sample standard deviation was calculated to be 8. (a) Table 7.10 gives the calculations required to fit the linear regression. (b) If we had not been told that the proportion voting for the candidate was 0.30, we could estimate the largest value of n to achieve the precision required no matter what the actual value of p is. Although this is only a necessary rather than a sufficient condition for identification, it usually gives the correct answer (see Prob. c c e: x?, (12.9416)(546) % 4.33 s2 = ha (n - k) n C x? 3 1 a n d s ~ l ~ l . ~ m . Since the calculated value of F = 5.69 (from Table 5.24) is less than the tabular value of F, we accept Ho, that the means for factor B (pesticide) populations are also equal. (a) b, and (b) b2 in Prob. (8.9) FURTHER TECHNIQUES AND APPLICATIONS IN REGRESSION ANALYSIS [CHAP. How can the agency conduct the test at the 5% level of significanceif it knows that the amount of detergent in the boxes is normally distributed? 8.15). (a) We can find the coefficient of skewness for the data in Table 2.9 using the formula based on the third moment with the aid of Table 2.22: This indicates that this distribution is negatively skewed, but the degree of skewness is measured differently than in Prob. If the purchaser knows that the lifetime of the components is normally distributed, should she accept the hypothesis that they last less than 100 h at (a) A 95% level of confidence? 't e report the data from Chap. This, in turn, affects Y,in the sewnd equation. 9.19, (a) correct for heteroscedasticity. The predicted values of the endogenous variables are obtained by substituting the observed values of the exogenous variables into the reducedform equations. A confidence interval can be constructed similarly for the population proportion (see Example 7) where p - - - 1.=6 p P-n gp = (the proportion of successes in the population) $ 9 (the standard error of the proportion) EXAMPLE 5. The only way to avoid this and reduce both os and p is to increase the sample size, n. 10.7(c) and are where The value of bl can be calculated from These two estimates of bl will generally be different, reflecting the fact that the supply equation is now overidentified. These differences require special methods of analysis (such as the inclusion of a disturbance or error term with the exact relationships postulated by economic theory) and multivariate analysis (such as multiple regression analysis). (a) 00.082, or 40.82% (b) 0.0918 (c) 0.228 3.68 The IQs of army volunteers in a given year are normally distributed with p = 110 and a = 10. y (a) What is the MSE? It is expected that Y will be directly related to both XI and X2. (b) Regress C o n Yd only. The htakes a random sample of n = 25 and finds that H = 520g and s = 75 g. (3.19) and (3.20)]: S E(X) = p Var X 3.32 = I I = a2 = Xf (X) dX [X - E (x)] ~f (x) dX (a) What is a normal distribution? (a) mutually exclusive, (b) not mutually (a) Two or more events are mutually exclusive, or disjoint, if the occurrence of one of them precludes or prevents the occurrence of the other(s). 10.9, estimate the supply equation by (a) 2SLS and (b) OLS. Since the exogenous variables of the system are uncorrelated with the error terms, OLS gives consistent reduced-form parameter estimates. With p = 8 years and a = 1 year, 6 years of schooling is 2 standard deviations below p and 10 years of schooling is 2 standard deviations above p. c 3. (b) How do these results compare with those in Prob. (a) Using the formula, determine the probability of picking 2 men in a sample of 6 selected at random without replacement from a group of 10 people, 5 of which are men. (b) Assuming that the error variance is proportional to s2and reestimating the regression using the transformed variable to correct for heteroscedasticity, we get PROBLEMS IN REGRESSION ANALYSIS CHAP. For factor A, F = 9.78 (from App. corrected? X and Y are assumed to be errorfree. (c) What is the standard normal distribution? 5.21; and (3) testing to see if the assumptions of the basic regression model are satisfied or, if not, how to correct for violations. 5-7 shows the values of p for various values of p > po. r~rr.,.?.,*nn - * !---,IT,' hL+ -r? Standard deviation. P = nP (3.11) The standard deviation is If p = 1 - p = 0.5, the binomial distribution is symmetrical; if p p > 0.5, it is skewed to the left. Since p could be equal to, larger than, or smaller than 1000, the firm should set the null and alternative hypotheses as Since n w 30, Ihe sampling disiribuiion of Ihe mean is approximaiely normal (and we can use s as an eslimak ofa). (4 Find the t value for 5,2.5, and 0.5% of the area within each tail for a sample size, n, that is very large or infinite. Assume that a population is composed of 900 elements with a mean of 20 units and a standard deviation of 12. A random sample of 100 construction workers in New York has an average weekly salary of $400 with a standard deviation of $100. When the number of independent lagged terms k is small, the model can be estimated with OLS, as done in Chap. b12 3.35 Assume that family incomes are normally distributed with p - $16,000, and a - $2000. 61 SIMPLE REGRESSION ANALYSIS Table 6.7 Consumption Regression: Calculations to Test Significance of Parameters Year Yi Xi t. P (R) .P (R) =-.-.-=10 10 10 3.16 125 1 = 0.125 1000 -s Past experience has shown that for every 100,000 items produced in a plant by the morning shift, 200 are defective, and for every 100,000 items produced by the evening shift, 500 are defective. 8 Utilizing the data in Table 7.9 and applying the OLS method to this demand function transformed into double-log linear form, we get In Y = 1.96 - 0.26 In X I (- 3.54) + 0.39 In X2 R 2 = 0.97 (6.64) where -0.26 and 0.39 are, respectively, unbiased estimates of the price and income elasticity of demand (see Prob. The fourth assumption is that the value which the error term assumes in one period is uncorrelated or unrelated to its value in any other period: E(uiuj) = 0 for i +j; i,j = 1,2, . What is the probability that 10 of the items picked are acceptable? What is the lowest IQ score acceptable for the advanced training? By setting os at I%, the firm would face a greater probability of accepting POTHESES ABOUT THE POPULATION MEAN AND PROPORTION I The formal steps in testing hypotheses about the population mean (or proportion) are as follows: t p equals some hypothetical value po. 8.5). If "Fixed width" is selected, the next box allows the selection of columns. 5.3 TESTING HYPOTHESES FOR DIFFERENCES BETWEEN TWO MEANS OR PROPORTIONS In many decisionmaking situations, it is important to determine whether the means or proportions of two populations are the same or different. Equation (7.18) for var bo is very cumbersome and also for that reason is seldom given and used: Table 7.6 Worksheet CHAP. This will k done in this chapter and in Chap. 8-2 Find the 95% confidence interval of YF for Xlp = 35 and XZF= 25 in 1981, given that = 31.98 + 0.65Xli + 1.11X,, & = 18, & = 12 (from Example 7. 8.14 Derive the Almon transformation for (a) a three-period lag taking the form of a second-degree polynomial and (b) a four-period lag taking the form of a third-degree polynomial. The critical value from App. indicates the sum of squared residuals for the unrestricted regression (i.e., the usual residuals). (b) A lower grade point average? Since for f, < 5, diseases C and D, we combine these two classes (see Table 5.10): STATISTICAL INFERENCE: TESTING HYPOTHESES CHAP. A 8.6 INTERPRETATION OF BINARY CHOICE MODELS The interpretation of bl changes in a binary choice model. High, simple, or partial correlation coefficients among explanatory variables are sometimes used as a measure of multicollinearity. [CHAP. 8 This can also be interpreted as the marginal effect of X on the expected value of Y. 62 4.48 (a) Write down the expression for n to solve Prob. while The variances of iland i2(or their estimates) are required to test hypotheses about and construct confidence intervals for bl and b2. Therefore A X , is a stationary series which can be used in a regression. 2 (40)(36) - 1,632 - 1,440 40 40 - N 192 40 - -= 4.8 points squared thc samc as in Prob. The methodology of this book and much of its content has been tested in undergraduate and graduate classes in statistics and econometrics at Fordham University. 71 MULTIPLE REGRESSION ANALYSIS (a) Using R 2 = 0.6932 found in Prob. i , , ~ j l ~ ~ ~bbL.i,,b ~ ~~ ~6 6 6 6 0 0 0 0 0 0 0 0 0 0 0 o w w m m 0 O O O O O O O O O O m w4OoOow 6 6 6 6 6 6 6 6 6 6 2 W 0 0 0 0 0 0 O b b b 6 6 6 6 6 6 6 6 6 6 6 L1bbbb 6 6 6 6 6 6 6 6 6 6 bbbbL1bbbb 4 0 4 0 0 0 0 0 0 0 5 O 0 5 0 0 0 0 0 W L.66 6 6 6 6 6 6 6 6 6 6 6 - 6 6 6 6 6 6 6 6 6 6 - P 4 P bi,,- 0 o o o o o o o o o o o o 2%- 8 8 8 8 8 8 8882; I P O 0 W W 0 I - 0 0 0 0 0 0 0 0 0 0 0 0 O N N I O o 0 0 0 0 0 0 O O O O P 4 bLlbi,,b 6 6 6 6 6 6 6 6 6 6 W 0 0 0 0 0 O O N W O o P o P u w $ I o W W o o o o o o o o o o ~6 ~ ~ ~ 0 0 0 0 0 0 0 0 8 8 8 8 8 8 E y 0 O O O O W O o P bLlbi,,b 6 6 6 6 ~ W N O W O o 0 0 0 0 0 w u w w 4 W W W 4 O O O O ~ ~ 2 8 8 8 8 8 8 8 8 8 2 3 i z 2 S i 8 8 8 8 8 Z g i S !?i% N W O o 0 0 0 W O W I O - 0 0 V I 4 4 O bbbbbbL1 m w 0 0 0 0 0 0 0 0 0 N 6 6 6 6 66bbL.bbb 3 3 $ 8 8 8 8 8 8 S S g g S W $ W z d 8 8 8 8 8 8 S G S 3 % W z % g8 8 8 8 8 S Z 8 % % , : 8 WI- w o w m u 0 6 6 6 6 4 w o o W wO N o WY o 4 bbL.66 6 6 6 6 6 b66L.b o o0o- oD 0 wO g & C j h 698bS3 8 8 8 E n L U Z E e 8 Q z i D 8 8 5 8 z $ $ y S Uip o o o o o - w O o P o k d- 6w 6w 6o 6 o6 o6 o6 o6 z b66L.L. bL.L.66 6 6 6 6 o w m w w N w - P o o o o o 6QbL.L. o PoOokkt b b b b b bbL.' .- L . To identify the values for the calculation, either individually enter each cell (Al, A2) or use a colon to indicate a range of cells (Al:AlO, all cells from A1 to A10). Regressing GPDIT on GDPT and PT, we get Since d = 1.77 > dU = 1.53 at the 5 % level of significance with n = 18 and k 1 = 2 (from App. 1 to 5 deal with the statistics required for econometrics Chapter 12 is concerned with using the computer to aid in the calculations involved in Copyright 2002 The McGraw-Hill Companies, Inc. (a) 0.56, or 56% (b) 0.84, or 84% (a) 1.5 Estimation inference is one of the most important and crucial aspects of the decision making process and science. This would not show up in the model. Bayes' theorem is applied in business decision theory, but is seldom used in the field of economics. This is to be contrasted with nonlinear regression analysis (discussed in Sec. , ~ x ' + n 7 ~ 2 ~ ~ X , 2 - n ~ 2 + n2 ~~ 2~ X , = a,, ~ aio r= a,, - a,, nEx' nEx' nEx' since in part a we saw that 6.15 E x! = E x?, -n ~ ~ 2 . (b) Which gives a better fit for the data in Table 6.12, the linear form of Probs. 6.1) il=)ci Yi =)ci(bo+ bl Xi + ui) = bo)ci + bl)ciXi +)ciui C xiui =b1+): c.u. - bl + C x: since C ci = C xi/ C x: = 0 (because C xi = 0) and 1)cix, I - - , - -xzxi -xzxi EX,' - EX,? 32 1 Copyright 2002 The McGraw-Hill Companies, Inc. (a) s2 = 19,760,000 dollars squared (b) s % $4,445.22 (b) the standard deviation for the 2.44 Find the coefficient of variation V for (a) the data in Table 2.29 and 2.30. P(G) . 11.27 (a) Calculate the Q statistic for the autocorrelations in Prob. This is obtained from App. No, it is rejected (H = 14.25) 5.57 Repeat the test from Prob. (a) Since d = 0.49, there is evidence of autocorrelation at both the 5 and 1% levels of significance. (c) Use instrumental variables to obtain consistent parameter estimates, on the assumption that X, is correlated with u,. (a) Estimator io s 2.30 is the Y intercept, or the value of aggregate mnsump$on, in billions of dollars, when disposable income, also in billions of dollars, is 0. 11 EXAMPLE 3. 81 FURTHER TECHNIQUES AND APPLICATIONS IN REGRESSION ANALYSIS 203 DUMMY VARIABLES 8.30 For the data in Table 8.2 (a) run regression Eq. (8.6). For an overidentified equation, more than one numerical value can be calculated from some of the structural parameters of the equation from the reduced-form parameters. The correct procedure is then to take first differences of Y before using it in a regression. ECONOMETRICS EXAMINATION is an appropriate estimation technique for the model? R2 = 1 when all sample points lie on the estimated regression line, indicating a perfect fit. These are mutually exclusive events. 12 MTCROSOFT EXCEL 12.4 Using the data from Example 1, (a) Use the data analysis tools to graph the histogram and ogive of test scores. We set the null and alternative hypotheses as The steps for the signed rank test are listed in Table 5.7. Since 4 < W < 32, we accept Ho: Med = 5 at the 5% significance level. (a) p = $1.15 (b) Median = $1.16 (c) Mode = $1.17 2.29 Find (a) the mean, (b) the median, and (c) the mode for the frequency distribution of incomes in Table 2.30. (a) Find the mean output for each fertilizer and the grand mean for all the years and for all four fertilizers. 41 (instead of n = 157 when we were told that p = 0.30). The semilog function is appropriate when the dependent variable grows at about a constant rate over time, as in the case of the labor force and population (see Prob. k b L . 11.21 and 11.22. Data on Px,Y, and PZ are then regressed against data on Dx and estimates of parameters bo, b,, bZ,and bj obtained. Another continuous probability distribution is the exponenticrl &triBution (see Prob. In these simple regressions, both In L and In K are statistically significant at much more than the 1% level with R2 exceeding 0.96. Similarly, in order for each estimated coer~ciento be "statistically significant," we would expect the dispersion of each estimated coefficient about the true parameter (measured by its standard deviation; see Sec. 3, we get See Fig. Since our data do not wnstitute a time series, we enter 1 as start date and 10 as end date to clear enough space for 10 obswvations. (b) Show the second-stage result of 2SLS estimation. We want to determine if housing starts is a leading indicator of consumption using Granger causality. 7 extends the corn production table to add insecticide use. 203. 8.7 Table 8.1 1 gives the quantity of milk (in thousands of quarts) supplied by a firm per month Q at various prices P over a 14-month period. (a) bo and b l are the parameters of the true but unknown regression line, while io and ilare the parameters of the estimated regression line. (a) Run a regression of Y on X and on dummies to take regional effects into account. In addition to Statistics and Econometrics, the Schaum's Outline Series in Economics includes Microeconomic Theory, Macroeconomk Theory, International Economics, Mathematics for Economists, and Principles uf Ecunuwrics. Massey, pr., "Kolmogorov-Srnirnov Test for Goodness-of-Fit." J. If there are 6 children in the family, what is the probability that half of them will have blond hair? How could this be corrected? © 1996-2014, Amazon.com, Inc. For X = 100 Looking up this value in App. 51 STATISTICAL JNFERENCE: TESTINO HYPOTHESES Table 2 Calculations to Find Sample Mean, Median, md Mode Unemployment Rate, % Frequency f Midpoint X Table 3 Calculations to Find the Variance, Strmudrmrd Deviation, and Coetlident of Variation Unemployment Class Rate, % Midpoint X 7.&7.4 7.5-7.9 8.H.4 8.58.9 9.&9.4 9.59.9 7.2 7.7 8.2 8.7 9.2 9.7 Frequency f Mean R 2 4 5 4 3 2 Cf=n=20 8.4 8.4 8.4 8.4 8.4 8.4 S = a (X - -1.2 -0.7 -0.2 0.3 0.8 1.3 m2 1.44 0.49 0.04 0.09 0.64 1.69 f (x -a2 2.88 1.96 0.20 0.36 1.92 3.38 f (X- R2= 10.70 ~ f (-q2 x 10.70 =r 0.56% squared n-1 19 Sk =-3 (X - X - med s =3 8.40 - 8.32 0.75 S 0,32 (see Fig, 1) Unemploymentrate Fig. (c) The probability of rain tomorrow is 20%. (b) What is the relationship between correlation and regression analysis? 111 TIME-SERIES METHODS 265 11.33 Using the data from Table 11.16, (a) What is the F statistic used to test if X Granger-causes Y with six lags? The levelof conjidence (given by 1 - a) refers to the probability of accepting a true hypothesis. A EXAMPLE 7. (c) What is the expected value and standard deviation of the distribution? (b) At the 1% level of significance? L~xP[-~ (~) ~] = d 1x M (the total area under the normal curve from minus infinity to plus infinity) (b) The normal distribution is the most commonly used of all probability distributions in statistical analysis. 46 (195I), pp 6&78. (a) 2.60 (b) No (c) Calculate the AIC for different number of lags and use model with lowest AIC CHAPTER 12 existing source (rather than collected by the researcher), they often come t. (b) Two aces from a deck in 2 picks without replacement? (The reader without knowledge of calculus can skip this problem.) Normal Eq. (6.4) is derived by minimizing C e: with respect to (b) Normal Eq. (6.5) is derived by minimizing 6.7 io: c e: with respect to il: Derive (a) Eq. (6.6) to find Ll and (b) Eq. (6.7) to find Lo. multiplying Eq. (6.5) by n and Eq. (6.4) by C X,.] (a) Multiplying Eq. (6.5) by n and Eq. (6.4) by [Hint for part a: Start by C Xi, we get ~ C X ~ Y ~ = ~ ~ ~ C X , + ~ ~ ~(6.24) CX,? However, because of common trends in time-series data, we would expect more than 50 to 70% of the variation in the dependent variable to be explained by the independent or explanatory variables for the model to be judged satisfactory. (c) One advantage of 2SLS over ILS is that 2SLS can be used to obtain consistent structural-parameter estimates for the overidentified as well as for the exactly identified equations in a system of simultaneous equations. (a) The OLS method gives the best straight line that fits the sample of XY observations in the sense that it minimizes the sum of the squared (vertical) deviations of each observed point on the graph from the straight line. (b) We can find the sample mean for the grouped data in Table 2.12 with the aid of Table 2.14: Note that in this case CJX = $98.75 # C X = $98.65 (found in part a) since the average of the observations in each class is not equal to the class midpoint for all classes [as in Prob. The rejection region for Ho lies to the right 1.64 standard normal deviation units for a! = 5%. 2-5. 5), we conclude that both bo and bl are statistically significant at the 5% level. Table 11.5 Randomly Generated, Standard Normal Distributed Variable The calculations for the first four parts are given in Table 11.6. To carry out the calculations for an AR process, we will use as an example the calculations for part a since parts b, c, and d use the same method with only a change of yl. 6-5? Statistical inference refers to estimation and hypothesis testing stirnation is the process of inferring or estimating a population parameter (such as its mean from the corresponding statistic of a sample drawn from the population. 111 253 TIME-SERIES METHODS For the first series: The correlation is high for the first lag, declines for the second, but is still positive, and then is close to zero at the third lag, indicating an AR process. First, since the purpose of theory is to generalize and simplify, economic relationships usually include only the most important forces at work. negative covariance indicates that they move in opposite directions. (a) This problem involves the binomial distribution. Problems 5.10 and 5.11 show how to find the operating-characteristic curve (OC curve), which gives the value of /3 for various values of c(. r' % 0.90 (% r % 0.88) (h) r. 6.4. (c) An interval estimate refers to the range of values used to estimate an unknown population parameter together with the probability, or con3dence level, that the interval does include the unknown population parameter. Fit the data to the Cobb-Douglas production function Table 8.6 Output and Labor and Capital Inputs of 14 Firms in an Industry FURTHER TECHNIQUES AND APPLICATIONS IN REGRESSION ANALYSIS [CHAP. - 2.17 Starting with the formula for a2 and s2 given in Sec. (c) What is the value of bl? 203, 255 CHAP. method (OLS) is a technique for fitting the "best" straight line to the Tt involves minimizing the sum of the squared (vertical) deviations of Min x(Yi - fi)' (6.3) observations, and firefers to the corresponding fitted values, so that residual. 1.18? We then take a random sample from the on the basis of the corresponding sample characteristic, we either accept or reject the articular degree of confidence. 4-5, we have a normal curve in the X scale and a standard normal curve in the z scale. (a) Regress Y on X and test the coefficient on X at the 5% level of significance. In what follows, we assume simple random sampling. 2 4 MEASURES OF CENTRAL TENDENCY Find the mean, median, and mode (a) for the grades on the quiz for the class of 40 students given in Table 2.7 (the ungrouped data) and (b) for the grouped data of these grades given in Table 2.9. (a) Since we are dealing with all grades, we want the population mem: E X 7+5+6+--.+5 N 40 f i = = 240 40 -= 6 points That is, p is obtained by adding together all the 40 grades given in Table 2.7 and dividing by 40 [the three centered dots (ellipses) were put in to avoid repeating the 40 values in Table 2-71. No 5.36 Define the acceptance region for Prob. Econometrics is subdivided into theoretical and applied econometrics. Cf) The area to the left of z = -1.60 and to the right of z = 2.55 is equal to 1 - 0.9398 (see part c). If Y is above (resp. In this and similar cases, trying to get an actual estimate of p does not greatly reduce the size of the required sample. MATRIX NOTATION 7.25 (a) Why is matrix notation used? The probability of getting a total of more than 4 equals 1 minus the probability of getting a total of 4 or less. We run the model restricted and unrestricted, then use the F test to test whether housing starts are a statistically significant predictor of personal consumption. (c) What do the answers to parts a and b indicate? 4.13(b)], and the sample proportion p is an unbiased estimate of p (the proportion of the popula,tion with a, given chara.cteristic). APPENDIX 4 4 Table of Random Numbers 309 Copyright 2002 The McGraw-Hill Companies, Inc. (a) AP, = -99.04 + 0.45 A X l , - 0.91 AX2, (6.02) (-0.18) R 2 = 0.73 d = 1.51 (b) There is now no evidence of autocorrelation at either the 5% or the 1% level of significance. (b) An inspection engineer picks a sample of 15 items at random from a manufacturing process known to produce 85% acceptable items. How is this different from the process of finding the probability that a normally distributed random variable assumes a value within a given interval? < 5000 lb), with a particular degree of confidence 1 - a. (b) Would the conclusion change if the sample had been 124, but with the sample proportion of the firms meeting the pollution standards the same as before? + + SIMPLE REURESSION ANALYSIS CHAP. (b) A continuous probability distribution refers to the range of all possible values that a continuous random value can assume, together with the associated probabilities. (a), (c) The Durbin-Watson statistic can be calculated in the basic regression procedure, "proc reg," but can also be calculated in "proc autoreg" with the added benefit of a p value which eliminates the need for supplementary critical value tables, and can be used for longer lags of autoregression. Thc spec$c linear relationship corresponding to the qneral !near relationship in Eq. (6.1) is obtained by estimating the values of bo and bl (representd by bo and bl and read as "b sub zero hat" and "b sub one hat")). 5000 lb). - = - = 9 1 13 13 P(Ac and S) = P(Ac n S) = P(Ac) P(S/Ac) = - - = - - - 52 51 2652 C H A P . 51 STATISTICAL INFERENCE: TESTING HYPOTHESES The calculated value of t exceeds the tabular value of to,,, = 1.71 with 24 degrees of freedom. Then, bo = -bA/bj and h, = 1 /hj, where ho and h, are consistent estimates of the intercept and slope parameter of the regression of Y, on X,. The relationship between X and Y in Fig. 5-13); that is, we accept the null hypothesis, Ho, that pl = p2= p3 = p4. A high observation at time t will be carried forward indefinitely in smaller and smaller proportions. If two or more explanatory variables are perfectly linearly correlated, it will be impossible to calculate OLS estimates of the parameters because the system of normal equations will contain two or more equations that are not independent. Note that the more p exceeds po, the greater is the power of the test (i.e., the greater is the probability of rejecting a false hypothesis). However, the sampling distribution of the ean, given by the symbol p r , and a standard deviation or stanhrd error, 52. (b) What is the importance of the answer to part a? Therefore the correct model of consumption and GDP is an error-correction model: Apt = 0.01+0.73 A X , - 0.55etPl (9.49) (-4.42) R2 = 0.76 The results reveal that for a 1 % increase in income there is a 0.73% increase in consumption (note that this is a double-log model). Since the calculated F = 70, we reject Ho and accept H I , that the population means of firstyear earnings for the 5 schools are different. For this reason, Chebyshev's theorem is seldom used to find confidence intervals for the unknown population mean. Ewnostrongly identified with regression analysis. (d) Test for the overall significance of the regression. SUMMARY PROBLEM 6.44 Table 6.13 gives data for a random sample of 12 couples on the number of children they have Y , and the number of children they had stated they wanted at the time of their marriage Xi. Regress Y , on Xi and report your results in summary form. Assume that the miles per gallon for each octane is normally distributed with equal variance. Also, the regression would violate the assumption of no heteroscedasticity because of the discrete nature of the dependent variable. Both these slope coefficients remain highly significant before and after the transformation, as does R ~ .The new constant is -8.45 instead of -6.17. (a) 39 to 61 (using Chebyshev's theorem and s as a rough estimate of a) (b) The 95% confidence interval here is much wider than those found in Prob. (b) Government agencies have large amounts of public, computer-readable data (Census, Bureau of Labor Statistics, Federal Reserve, etc.). Accept Ho Table 5.33 Acceptable and Nonacceptable Components Produced Each Hour of the Morning Acceptable Nonacceptable 90 5.51 100 110 100 400 The number of people voting Democrat or Republican below the age of 40 and 40 plus in a random sample of 30 voters in a city is given in contingency Table 5.34. (c) Why would the estimation of W and P equations by OLS give biased and inconsistent parameter estimates? (b) What are the advantages? (c) What is the symbol conventionally used for the probability of type I1 error? (a) P(C) = 13/52 = 114 (b) P(A) = 4/52 = 1/13 (c) P(Ac) = 1/52 (d) p(C1) = 314 (e) P(C) p(C1) = 1 + 3.45 An urn contains 12 balls that are exactly alike except that 4 are blue, 3 are red, 3 are green, and 2 are white. (a) To economics, business, and other social and physical sciences (b) Summarizing and describing a body of data (c) Drawing inferences about the characteristics of a population from the corresponding characteristics of a sample drawn from the population. (b) The a priori economic criteria refer to the sign and size of the parameters of the model postulated by economic theory. Thus the best results are those given in part a. It is important to annotate programs and give variables descriptive names so the program is easily debugged, if necessary, and others can read your program. In Q on 1nK only. E{X) = p = np = (6)(1/4) = 312 = 1.5 blond children ~ ~ 1.06blondchildren Bewuse p < 0.5, the probability distribution of blond children is skewed to the right. [If we had not treated the number of defective items as a continuous variable, we would have found that P (X 5 4) = 0.287.1 3.39 If events or successes follow a Poisson distribution, we can determine the probability that the first event occurs within a designated period of time, P (T I t) , by the exponential probability distribution. First, if u, in Eq. (8.9) satisfies all the OLS assumptions (see Prob. (e) ryxl.x2= 0.71 and ryx2.xl= -0.87; thus X2 contributes more than X I to the explanatory power of the model. income of 195 U.S. dollars while holding X2 constant. (c) The size of the total confidence interval in Prob. bo b3 = 1 - a0 and bo - bl b2 = a l (l - b3) - bl = a l a0 Substituting into the original model yields Grouping terms, we obtain bo AY, = bo + b l AX, --(Y,-, a0 Since YrPI- a lXIPI= a. The number of lags may be chosen using the AIC, adjusted R2, or one may include the highest feasible number of lags. The reader without knowledge of calculus can skip this problem.) c But in the OLS process)eixli = 0 and)eixli = 0 a c e : = -)eixzi = 0 and)eix2i = 0 ace' - axil a c i 2 MULTIPLE REGRESSION ANALYSIS [CHAP. (b) Estimate the population variance from the variance between the means or columns. PROBABILITY AND PROBABILITY DISTRIBUTIONS CHAP. An.r. (a) 1/36 (h) 118 (c) 5/36 (4 116 (e) 116 3.52 What is the probability of obtaining the following from a deck of cards: (a) A diamond on the second pick when the first card picked and not replaced was a diamond? It is also convenient (and needed for plotting the frequency polygon) to find the class mark or midpoint of each class. In the second equation, X and Y are uncorrelated with u2 (i.e., YI is correlated with u1 but not with u2), SO that OLS gives unbiased parameter estimates for the second equation. Since the ADF statistic is greater than the critical value, we accept the null of unit root. 1057-1072. 2 INTRODUCTION [CHAP. THE POISSON DISTRIBUTION 3.28 (a) What is the difference between the binomial and the Poisson distributions? 7.37) but n = 100, find Ans. The data step is as follows: d a t a example ; i n f i l e "c:\examplel.txt"delimiter= i n p u t t e s t score: ","; run ; quit ; The SAS Log window reports the following information: 1 2 d a t a example ; 3 i n f i l e "c:\examplel.txt~' i n p u t t e s t score; 4 run; delimiter = ","; NOTE: The i n f i l e " c : \ e x a m p l e l . The 95% confidence interval for the unknown population mean is /A =X f1 .96~~ since n > 30 since n > 0.05N 60 dm Jm 1000 - 1 = 100 f 1.96- using J as an estimate of f Thus p is betwew 90.89 and 109.11with a 95% degree of wddence. (8.5) to (8.7) showing a consumption function for peace FURTHER TECHNIQUES AND APPLICATIONS IN REORESSION ANALYSIS [CHAP. 2.3a) N') ' ~(X-p)'=1+1+0+16+4+1+0+1+9+9+16+4+1+1+4+0+1+4+4+16 +9+1+0+1+9+4+16+4+1+9+4+0+1+4+9+0+1+9+16+1 = 192 (a) a '= a - N 192 40 'I2 a = /C(X - -= - PI' 4.8 points squared = = N rn % 2.19 points (b) We can find the variance and the standard deviation for the grouped data of grades with the aid of Table 2.18: '= a and a= (x -'I2 N ?!!-40 - = 4.8 points squared & J48s 2.19~oints = the same as we found for the ungrouped data. (a) Yes (b) -0.0166 (c) -0.0118 Table 11.16 Industrial Production Index and S&P 500 Index: United States, Feb. (c) Why do we not simply take the sum of the deviations without squaring them'? (a) $0.29 (b) $10,000 to $29,999, or $20,000 2.38 Find the interquartile range and quartile deviation for the data in (a) Table 2.29 and Ans. (We assume here that EO = 0 to get a starting value. If n c 30 but either a is not given or the population from which the random sample is taken is not known to k normally distributed, we should use Chebyshev's inequality or increase the size of the random sample ta n r 30 (ta enable us to use the normal distribution). First, on the basis of the sample uld reject a hypothesis that is in fact true. 12.12 (a) At the 5% level of significance? Consumption theory tells us that, in general, people increase their consumption expenditure C as their disposable (after-tax) income Yd increases, but not by as much as the increase in their disposable income. EXAMPLE 7. What is the general procedure? (c) Derive the formula for the structural coacients from the reduced-furin coefficiet~ts. (c) What is the usefulness of the coefficient of variation? (b) What is meant by the mean and standard error of the sampling distribution of the mean? Note that the range was extended from $3.50 to $4.30 so that the lowest wage, $3.55, falls within the lowest class and the largest wage, $4.26, falls within the largest class. Thus the estimated OLS regression equation of Y on XI and X2 is (b) The estimated OLS equation indicates that the level of per capita income Y is inversely related to the percentage of the economy represented by agriculture XI but directly related to the literacy rate of the male population (as might have been anticipated). 5.32 Repeat the test from Prob. 3-39). Thus 1 1 1 P (H and H) = P (H n H) = P (H) . 51 STATISTICAL JNFERENCE: TESTINO HYPOTHESES Table 5.21 One-Way ANOVA Table for Fertilizer Experiment Variation Sum of Squares Degrees of Freedom Mean Square Explained by fertilizer (between columns) SSA= 112 c-1=3 MSA = 37.33 Error or SSE = 726 (r - l)c = 28 MSE = 25-93 SST = 838 rc-1=31 F Ratio MSA/MSE = 1.44 unexplained (within columns) Total -- Acqtance region - Rejection region Fig. The variances of ioacd il(or their estimates) are required to test hypotheses about and construct confidence intervals for bo and bl. One must also take into account the long-run relationship between the variables. 7.5(a)], C yxl = -28 and 7.4), and y: = 40, we get c C yx2 = 38 (from Table This value of R2 differs slightly from that found in parts a and b because of rounding errors. When all the numbers are equal, pG equals p; otherwise pG is smaller than p. These are probably very common in view of the way most data are collected and elaborated. We have also received very useful advice from Professors Mary Beth Combs, Edward Dowling, and Damodar Gujarati. (d) How can the problems created by the existence of errors in variables be corrected? (a) What is the value of the new intercept and the slope coefficients associated with X , and X,? Does the urn contain an equal number of green, white, red, or blue balls? Table 7.1 extends

Table 6.1 and gives the bushels of corn per acre, Y, resulting from the use of various amounts of fertilizer X1 and insecticides X2, both in pounds per acre, from 1971 to 1980. (b) What does this show? 1.69 1.26 1.27 1.29 1.30 1.32 1.33 1.44 1.45 1.46 1.47 1.48 1.17 1.19 1.21 1.22 1.24 1.26 1.54 1.55 1.55 1.55 1.56 1.56 1.08 1.10 1.12 1.14 1.16 1.18 1.66 1.66 1.66 1.65 1.65 1.65 0.99 1.01 1.04 1.08 1.10 1.79 1.78 1.77 1.76 1.76 1.75 0.90 0.93 0.95 0.92 1.01 1.03 1.92 1.90 1.89 1.88 1.86 1.25 1.23 24 25 26 27 28 1.02 1.04 1.05 1.07 1.09 1.10 1.19 1.20 1.21 1.22 1.23 1.24 0.94 0.96 0.98 1.00 1.02 1.04 1.29 1.30 1.30 1.31 1.32 1.32 1.33 ... 1.62 1.62 1.64 1.61 1.65 ? (c) Find the concept of determination. (c) For a histogram in Excel, choose Tools-Data Analysis. 3 CHAP. (a) The correlation coefficient measures the degree of association between two or more variables. Arrange data from smallest value to largest value. (The values for y: are obtained by squaring y) from Table 7.1 and are used in Sec 7.3.) Using the values from Table 7.2 and 7.1, we get ... The coefficient of determination R2 is defined as the proportion of the total variation in Y "explained" by the regression of Y on X. As a result, there are only three independent normal equations and four coefficients to estimate, and so no unique solution is possible. The disadvantage of the subjective or personalistic approach to probability is that different people faced with the same situation may come up with completely different probabilities. (a) The hypotheses to be tested are H0 : μ = μ0 and H1 : μ ≠ μ0 ... FURTHER TECHNIQUES AND APPLICATIONS IN REGRESSION ANALYSIS ... FORECASTING ... TESTS OF SIGNIFICANCE OF PARAMETER ESTIMATES ... MULTICOLLINEARITY ... HETEROSCEDASTICITY ... AUTOCORRELATION ... DISTRIBUTED LAGS ... SIMULTANEOUS-EQUATIONS METHODS ... TIME-SERIES METHODS ... PROBABILITY AND PROBABILITY DISTRIBUTIONS ... STATISTICAL INFERENCE: TESTING HYPOTHESES ... DESCRIPTIVE STATISTICS ... THE TWO-VARIABLE LINEAR MODEL ... MULTIPLE REGRESSION ANALYSIS ... SAMPLING AND SAMPLING DISTRIBUTIONS ... ESTIMATION ... [illegible]

Note that in this problem we had to use the t distribution because s was given (and not a, as in Prob. (a) The best unbiased or efficient e s t h t o r refers to the one with the smallest variance among unbiased estimators. 6.4), leading to biased and inconsistent OLS parameter estimates. The firm takes a random sample of n = 100 of its lightbulbs and finds that the sample mean X = 980h and the sample standard deviation s = 80 h. (b) Getting a heart at a club an a single pick from a well-shuffled deck of cards also constitutes two mutually exclusive events. Note that the class midpoints are obtained by adding together the lower and upper class boundaries and dividing by 2. Only if these assumptions are satisfied will the estimated coefficients have the desirable properties of unbiasedness, consistency, efficiency, and so on (see Sec. Models of this nature are recursive rather than simultaneous. 7.8 (c) s = 1.8. A simple random sample is obtained from a population, and a spot- observed value is called an estimate. Estimate the population variance from the variance within the samples (MSE in Table 5.4) 3. 6, recall that pi = 27.12 + 1.66 Xi, n = 10, where 2 = 18, C(Xi - 2) 2 = 576 (from Example 6.2), and s2 = - e i / (n - 2) % 47.3118 % 5.91 (from Example 6.3). (a) The t value for 10% of the area within each tail is obtained by moving down the column headed 0.10 in App. For the four models the estimation yields TIME-SERIES METHODS [CHAP. 12 is significant at the 10% level, but the R 2 indicates that 42% of the variation in Y is explained by the model even though none of the independent variables stand out individually. 7.15(a), we get c This compares with = = -12.2730 found in Table 7.7. (The small difference in the value of = found by these two methods is obviously due to rounding errors.) Note, however, that finding C is an approach that yields the null and alternative hypotheses in testing the overall significance of the regression. Since the calculated value of F = 8.21 (from Table 5.24) is less than the tabular value of F, we accept Ho, that the means for factor A (fertilizer) populations are equal. (a) Yes. (b) t0.05 = 1.699 (c) to.025= 2.045 (d) to.005= 2.756 4.50 Find the value of t for the area as the following data ...

STATISTICAL INFERENCE: TESTING HYPOTHESES 89 performed we have seen from previous classes, we first note that this problem involves the binomial distribution. (a) Find the OLS regression equation of Y on X1 and X2. / y?, (b) R 2 = 1 - = ? (d) Degrees of freedom (do refer to the number of values we can choose freely. In this excel file, weighted average, of a population, p,., or sample, Xu,. the weights, w, have the same function as the frequency in finding the mean for the grouped data. Excel uses a spread, variance, which means that the user enters data and function in certain locations on the spreadsheet (called cells). 11) INTRODUCTION (c) Inferential statistics (both estimation and hypothesis testing) refers to the drawing of generalizations about the properties of the whole (called a population) from the specific or a sample drawn from the population. t,l C yxl - L2 C yx2. 6), we reject Ho, that age is not a factor in the type of car bought (and conclude that younger people seem more likely to buy foreign cars). (d) We can find a2 and a for the grouped data in Table 2.9 with the aid of Table 2.12 - 2.2 - 1,440 - 192 = JaZ N 40 = J48% 2.19 points 40 40 = 8 points squared the same as in part a and Prob. (a) See Table 2. 1 can be written with matrices as Y = Xb+u where e 'e (X 'X): (symmetrical, so lower and upper triangle are identical) 160 MULTIPLE REGRESSION ANALYSIS EXAMPLE 7. / (-k 1)[(C e?/(n - k)] = [(R 2 / (k - 1)] / [(1 - R2)/(n - k)]. If the test, we expect the 3 to come up 100 times in 600 tosses of the die as compared with the actual, observed, or empirical 106 times. C e2 = (1 - R 2)c y2 = (1 - 0.95)(40) = 2. If the bank takes a random sample of 100 accounts, what is the probability that the average savings for these 100 accounts will be below $2800? (a) If ~ < = . If the calculated F ratio is greater than the tabular value of F at the specified level of significance and degrees of freedom (from App. (Note: R2 is not well defined in regression through the origin and is therefore not comparable with the previous regressions. Thus p is between 68.24 and 91.76 with 95% level of confidence. 4.14 If a random sample of 49 with X = 50 and a standard deviation of 20 is taken from a population of 800. Since the is interested in testing if p > 500 g, we have Since the population distribution is normal but n i30 and u is not known, we must use the t distribution (with n - 1 = 24 degrees of freedom) to define the critical, or rejection, region of the test at a 5% level of significance. If Y and X are cointegrated, then it is not enough to simply difference the variables in the regression to remove the problem of nonstationarity. Since the errors created the problem of Heteroscedasticity, we can have nonstationary data. The estimated regression is CAFE 10.2 Reduced-form equations are obtained by solving the system of structural equations so as to express each endogenous variable of the system as a function only of the exogenous variables of the system. These first three assumptions about the error term are: (a) Exogenous variables are those variables which are determined from outside the model. COMPUTER APPLICATIONS IN ECONOMETRICS CHAP. A distribution is positively skewed if the right tail is longer. (a) 0.0123 (b) 0.242 (c) 0.409 (d) 0.5 (e) 0.591 (f) 0.758 5.43 How can we get the power curve for Prob. (c) Is there any test to detect the presence of errors in variables? However, with k large (in relation to the length of the time series), an inadequate number of degrees of freedom may be left to estimate the model or to be confident in the estimated parameters. (b) Of the probability of hitting a target on a single shot is 0.3, what is the probability that in 4 shots the target will be hit at least 3 times? Calculate the ACF up to four lags. 11.21 The data in Table 11.12 report housing starts Y in thousands and personal consumption X in billions of 1996 US dollars. 4.47 If the manufacturer did not know that p = 0.27 (b) By letting p = 0.5 and n = 9? Ans. The sum is true for two m s s i v e t o w s of a pair of dice or picks of two cards from a deck with replacement. CHAP. 12 (b) Because u 1s is not a constant. (c) if the distribution is normal. 10, the critical region. 8 (a) The test is satisfied since the sample is a random sample of 1 the pattern of X. the value of test of significance has distribution for the null hypothesis. (b) Variance of nb, and that ~n = 0.9995 ~ 0.025. 3.5 Sinc we have that the test is not normal. What would be the distribution of the sample mean, of the sample data ... 37.33 3 MSA 37.33 F = ... 1.44 MSE ... 25.93 (b) See Table 5.21 (c) The hypotheses to be tested are Ho: pl = p2= p3 = p4 versus H1: pl, p2, p 3 , p 4 are not equal Since the calculated value of F = 1.44 is smaller than the tabular value of F = 2.95 with df = 3 and 28, we accept Ho (see Fig. 3 ie) (0.2)=(0.8)(0 01 ~ 01 = 0.1074 (looking up = 10,X = 0, and p = 0.2 in App. (a) Testing a hypothesis refers to the acceptance or rejection of an assumption made about an unknown characteristic of a population, such as a parameter or the shape or form of the population distribution. 10.8 and using the data in Table 10.2 and trend values T = 1,2,3, . 51 STATISTICAL INFERENCE: TESTING HYPOTHESES 109 females in the ?1-plus age group. Sampling from an infinite parent population (or from a finite parent population with replacement) would have required taking an infinite number of random samples of size n from the parent population (as obviously impossible task). (4) Two successive tosses of a coin, by a die. When Y is above the level indicated by X, we would expect Y to fall, and vice versa. The minimum required sample size (n > 30) is found as follows: since the total confidence interval, X n = 67.24, p, is 3 min or 68 (rounded to the next higher integer) EXAMPLE 7. (a)), (b), (c), (d)We start with the Excel file from Prob. (a) Estimator i0? Should the hypothesis be accepted at the 1% level of significance that the average income for those 10 families is higher than 50? (c) Find the median, and (c) the the advantages of the mean are the three (a) The mean and the median only. We have a (b) the fact that the sample is taken from a population of three... these are the median and the mode are the same. However, the results may vary well differ in other cases. T(X): a(X) = a (X) = 0.9, Since we have: Why is the probability that after a decrease in the price of a car bought of (foreign or domestic) is independent of the age of the buyer, the dealer constructs a table of expected frequencies. (a) 0.0048 dollars spend on Y/s is $0.0093 that is deducted for the 2.43 Using the easier computational form~/alias, find (a) the variance and family incomes in Table 2.30. What is the probability that after a defective item is drawn that another one? Note that increasing n (and holding everything else the same) increases the probability of accepting an equivalent structural equation. Ede Ed Qbleetty = s w From Q i ck l V W ~ ~ o cl s l ~ e dPr~d al ~ m aFRIQX I: ... on ws a s a do ~ Help 1 ~p ul ~ a r n ~ = labet ef ~ Ed P u n s e Granger Causality Tests Date: 02/8/01 Time: 11:17 Sample: 1961 1935 Lags 6 Obs Null Hypothesis: X does not Granger Cause Y Y does not Granger Cause X 29 F-Stat ~st ~ (Probab~ility 2 59831 0 5353 0.05957 0.77853 12.13 (a) Save the variables from Prob. 8.1: (a) Under the Yi= to also since the coefficient of X Y p1 i.e., & Is statistically significant at the 5% level and its value exceeds the value of bo, this model does not fit the data well. (1. , : . f; 1) -1 ~ - K . 2.16). What is the relationship between Durbin and G. Note that the 95% we have to is ? r 1 and important? r.,~rr~g Fv' II. The same is true for the third equation. Durbin and G. Note that the 95% are n H e m bands are clogst at XF = X. Since the sample size of 2 is greater than 5% of the STATISTICAL INFERENCE: TESTING HYPOTHESES [CHAP. For the sample in plant 2, the mean is 1024 g and the standard deviation is 60 g. Collection of data on the variables of the model and estimation of the coefficients of the function with appropriate econometric techniques (presented in Chaps. (We assume again that ED = 0 to get a starting value.) (b) The statistical criteria are satisfied only if a "high" proportion of the variation in the dependent variable is "explained" by the regressors. The estimated equation is that the parameters are "sufficiently narrow." There is no generally accepted answer as to what is a "high" proportion of the variation in Y "explained" by Px,Y, and PZ. This is accomplished by moving down the z column in the table to 1.0 and then across until we are below the column headed .00. 71 159 MULTIPLE REGRESSION ANALYSIS [CHAP. (c) Why is the probability theory required? 31 E(X) = p = np = (4)(0.3)(0.7) = n a) 0.92 hits Because p < 0.5, the sampling distribution is skewed to the right. (b) Type I error refers to the rejection of a true hypothesis. 1) Thus Ho Here n 15, N = 10, p = 0.85, and 1 - p = 0.15. X,, refer to the n (or N) observations. Simple Regression Analysis (1) THE NO-VARIABLE LINEAR MODEL table linear model, or simple regression analysis, is used for testing hypotheses about the relationship between a dependent variable Y and an independent or explanatory variable X and for prediction. Linear regression analysis usually begins by plotting the set of XY values on a scatter ... (a) Suppose Ho: Y = Yb ... Y = b0 + bX (6.1) ... if we let b = (b0, b1) be the constant, then we have ... to test the significance of b1 in this regression ... using the t distribution with n - 2 degrees of freedom. 14.7 ... b(1) = b + ... Have Ec = (z) 1 = ... 74 ... (a) Find the probability that ... is at least one ... (b) ... the number of ... (c) The sum of the assumptions of the multiple regression linear model ... 0.4946 ~ 0.0054, or 0.54%, is the area in the right tail of z = 2.55 (see Fig. Should the producer accept the hypothesis that the cable has a breaking strength of 5000 lb at the 5% level of significance? a ... Generally by declining weights and 0 < b < 1 gives bi=bbo Substituting Eq. (8.16) into Eq. (8.9), we obtain Lagging by one period and multiplying by λ, we get 8.2 How can be obtained? 1.9 In what way is the MSE (b) is it to the case ... (a) We must, since if E(u) = 0 ... of interest. It is not ... the unbiased ... (d) ... are normally distributed ... Since the ... lb ... and we reject Ho ... We find that ... (b) ... the 1% level of significance, we accept Ho (6.2) ... is not significant. ... the ... 15th percentile (b) ... 30th percentile (c) 65th percentile ... of the 20.5th item in the median = Q3 (third quartile) = 7.5 = the value of the 30.5th item D3 (third decile) = the value of the 12.5th item P60 (sixtieth percentile) = the value of the 24.5th item = $3.90 + 2(125 ? (0.10) = $3.90 + 0.07 = $3.97 = 4.1 (a) ... (a) 0.0003 = 2.22 DESCRIPTIVE STATISTICS [CHAP. (c) The estimate ... 0.0342 dollars squared CHAP. 37 ... 3. I want to expand our definition of a parameter ... (d) Y ... indicated by ... as an additional explanatory variable. (c) ... 6 (see Table) ... 0.09 ... 0.8: PROBABILITY AND PROBABILITY DISTRIBUTIONS P(X) = (CHAP. However, the income, Y, equation is underidentified because it excludes no exogenous variable. (4 How is the value of pF found? However, 4 heads in 20 tosses is very likely to mean an unbalanced coin. 5 Table 5.28 Ratings Signed Rank Test 5.29 x,4 - X B Ordered Rank -0.5 -1 0 -3 -2 0 4 -1.5 -1 -3 0 0 -0.5 -1 -1 -1.5 -2 -3 -3 4 N/A N/A 1 2.5 2.5 4 5 6.5 6.5 8 Rank for Positive Differences = - - - - - - 8 Data from the World Bank's World Development Indicators report that for 9 Latin American countries, male illiteracy is as follows (in percent): Argentina 3, Bolivia 8, Brazil, 15, Chile 4, Colombia 9, Ecuador 7, Peru 6, Uruguay 3- Venezuela 7 (a) Test the null hypothesis that the median illiteracy rate is 8 % at the 10% significance level. For our case, we try AR(1) ... AR(1) and MA(0), and MA(1) and MA(0) since we have an idea of the lag lengths, but not the process. 4 4 1 * Durst B). Heteroscedasticity Eq 61 (c) With heteroscedasticity, the OLS parameter estimates are still unbiased and consistent, but they are inefficient (i.e., they have larger than minimum variances). The series can be named in the "Series" tab, we also switch the X and Y variable so that P is on the vertical axis. This is analogous to what was done in Sec. For example, for the estimated regression equation Y = bo bXt b2X2,ryx1,x2has the same sign as bl and ryx2,x1bas the same sign as b,, the regression line goes through point XY. ei ei2 X, Y X2 1 2 3 4 5 6 7 8 9 10 11 12 102 106 108 110 112 120 124 130 142 148 150 154 174 178 216 130 140 149 148 156 160 174 178 200 196 206 ... 0.4771 ... 1.2950 0.196 ... 4 ... -0 ... 0.9 ... 41.7316 4.0100 21.7156 2.2500 1.4 ... 0.4761 ... -0.2975 ... 1.2400 ... 2.3 ... 4.9 ... 0.21 ... -0.8 ... 1.0 ... 0.6 ... 1.3 ... ... 3. EX' = 27,112 c x = 4812 c e1 = 0 ... = 1153.2752 ... (b) Since ... EX ... 3.7616 ... (e) ... 0.17 ... 4.5 ... 39 ... ... 1.1 ... (a) State the null and alternative hypotheses ... 11 ... ... Table 6.1 (a) State the ... 2.17 ... 1.6 ... 1998 ... 1.5 ... 0.81 ... (e) ... 5 ... (a) the ... ... 11 ... 1.99 ... GDP ... (b) ... 10 ... 0.96 ... X ... GDP ... 0.00 ... Price Index, 1982~1999 Year 1982 1983 1984 1985 1986 1987 1988 1989 1990 GDP 571.1 762.2 876.9 887.8 838.2 929.3 916.7 922.9 849.6 GDP 415.5 6286.8 15583.1 5806.0 5969.5 6234.4 6465.2 6633.5 6664.2 P ... 47.66 69.75 72.24 74.40 76.85 78.84 81.36 84.24 87.76 Year 1991 1992 1993 1994 1995 1996 1997 1998 1999 ... 1109.1 ... 1155.7 ... 1152.4 ... 1283.7 ... 1438.1 ... 1609.9 ... 1751.6 ... 1720.9 ... 1768.7 ... 7461.1 ... 7621.0 ... 7933.1 ... 8272.9 ... 8654.5 ... 9084 ... 9190.6 ... 7930.10 ... 8065.1 ... 9112 ... 9... 1 ... ... 108.5 ... 91.18 ... 99.82 ... 100.00 ... 101.53 ... 102.49 ... 103.05 ... 103.31 ... Source: St. Louis Federal Reserve (Bureau of Economic Analysis) (a) ... (c) ... 4.0 ... (b) Find ... the value ... in ... [CHAP. ... 5 ... (a) ... 4 ... = 1.05 at the 5 % level of significance with n = 18 and k 1 = (2 from App. Finally, since and it falls within the acceptance region, we accept Ho, that p = 500 g, at the 5% level of significance (or with a 95% level of confidence). Thus and EXAMPLE 2. 4.6(b) (a) find the mean and the standard error of the mean using the formulas for the population mean as in Eq. (8.10) and the period and the form of the lag may not be known 8.35 Table 8.21 gives the business expenditures for new plant equipment of public utilities Y and the gross national product. (c) The expected value ... and mean and standard deviation ... (a) the mean and the standard deviation of the ... (d) the the binomial distribution ... (c) the ... distributions can be derived by substituting for C and f (X) for P(X) into the formula for the expected value and variance for discrete probability distributions (Eqs. (e) Which is the "best" result? (c) Regress Y: on X,, to correct for autocorrelation, where Y, and X,. are the transformed variables. 5-61. (b) (a) i= 0.40 and 6 = -3.20 Table 8.21 Business Expenditures for New Plant Equipment of Public Utilities and the Gross National Product, United States, 1960~1979 (in Billions of Dollars) Source: Economic Report of the President, U.S. Government Printing Office, Washington, DC, 1980, pp. By assuming that our confidence interval (based on the single random sample that we have actually taken) is one of those 95% confidence intervals that include p, we take the calculated risk of being wrong 5% of the time. 6.2 The analysis of variance does this test? 2.7 7.7 TEST OF THE OVERALL SIGNIFICANCE OF THE REGRESSION The overall significance of the regression can be tested with the ratio of the explained to the unexplained variance. 2.7 A firm may ... ... the estimated ... ... 7.7 ... THE ... SIGNIFICANCE ... R ... ... b ... 2 (c) ... ... 4 ... ... in deviation form for the case where X = T = 0. 7 TEST OF THE OVERALL SIGNIFICANCE OF THE REGRESSION The overall significance of the regression can be tested with the ratio of the explained to the unexplained variance. ... Therefore, 0.5 5 5 P (R and R and R = P(R). Therefore a Y, is a stationary series which can be used in a regression. A manager wishes to estimate the mean number of minutes that workers take to complete a particular manufacturing process within 15 min and with 95% confidence. Computations are 162.495 = 10.355. (Copyright 2002 The McGraw-Hill Companies, Inc. If a nonstationary series is used for the Y value, we ... and R ... = ... Log ... then we ... 10 = ... lkelihood, and the Durbin-Watson statistic. = h (b l - a1) +bo = 10.9 n2 %(:-$) b2 = -n5(bl - a1) = n5 and bo = -n3(bl - a1) With reference to the demand-supply model given below, (a) determine if the demand and/or supply equation is exactly identified, overidentified, or underidentified. Table 2.16 Calculations for the Average Deviation for the Grouped Data in Table 2.9 Grade Class Midpoint X Frequency f Mean p 1.2-2.4 2.5-3.4 3.5-4.4 4.5-5.4 5.5-6.4 6.5-7.4 7.5-8.4 8.5-9.4 9.5-10.4 2 3 4 5 6 7 8 9 10 2 3 5 5 8 8 4 2 6 6 6 6 6 6 6 6 6 21 4 p-2' 12 -4 -3 -2 -1 0 1 2 3 4 PI Fx 4 3 2 0 1 2 3 4 1 2 3 4 N=40 = PI 12 9 10 5 0 8 8 12 8 C IX p =72 Find the average deviation for the grouped data in Table 2.12. 4 4 Solved Problems SAMPLING 4.1 (a) What is meant by statistical inference? How do they compare with the corresponding values before the transformation? Since used as an unbiased estimate of a2? 2 5 2 where n = number of available observations. Is voting Democrat or Republican independent of the voter being below the age of 40 or 40 plus in this city at the 5% level of significance? 3.20 and 3.31). Thus probability theory is an essential element in statistical inference. 2.15(b) 3(p-med)%6-6.17 p 2.19 N -3(-0.17) 2.19 ? Since the system has a random error term this... from observations that ... of the ... the ... are ... of ... of ... 0.95? ... if ... the ... (a) Use 95% confidence interval ... for the ... points 10 22 10 = - = 22.12 = points squared = 2.148 points Table 2.5 Calculations for the Range in Grade Y p 6 7 6 8 5 7 6 9 10 6 7 7 7 7 7 7 7 7 ... Y-Y - Pl pl 0 1 1 0 1 2 3 1 -1 0 -1 -2 0 -1 2 3 -1 0 1 (X ... (b) ... 2 ... ... 1 ... (X ... the 95% ... ... -1 ... 12 ... EXAMPLE 6. The following solution works with any number of independent variables, and is therefore extremely flexible. Picking at random a spade or a king on a single pick from a well-shuffled card deck does not constitute mutually exclusive events since we can pick the king of spades. DESCRIPTIVE STATISTICS n = C X * N * 2 = 1.62 [CHAP. 161 FURTHER TECHNIQUES AND APPLICATIONS IN REGRESSION ANALYSIS [CHAP. 61 0 7'11 ~ 118 126 134 142 150 158 166 178 Composite Industrial Stock X Reg M The wlnumpth~nagresssi~nline is plotted in Fig. 6.5, where the periods over which the sample was taken ... P ... 5. ... "Pnamesfdnaryandnj~e(cuntinuehd(tabsfor"] "Pegindatastatepandnamestemporary/datasettPr = e a d s p r e v i o u s s y s v e d d a t a:" prev alone; var=ratio; P ... procedure (n c e c a l c u l a t e ccorrelogram*select:a:selada(clowelityth"Y run :) cnit ; This produces the following output: The ARIMAProcedure Name o f W o r k i n g S e r i e s Standard Deviation Number of O b s e r v a t i o n s 0.301881 0.070289 10 Autocorrelat ions Lag Covariance Correlation -19876543210123456789 Std Error COMPUTER APPLICATIONS IN ECONOMETRICS = a g Correlation [CHAP. Is the die balanced at the 5% level of significance? ·Rejection region ·Aweptmce region Rejection region Fig. Since we were told (in Prob. 5.39? To do this, he takes a random sample of 100 bulbs from each brand and finds that brand 1 lasts 980 h on the average X with a sample standard deviation 80 h? If all or most sample observations involve opposite changes in X and Y, then (r ~oo) X and large, implying a large negative correlation. Similarly, once YZ has been CHAP. (g) Ho: p = 0 (a) 26.60 - 5.1061X1 + 0.0167X2 t values (-5.4096) Ho: = 0.9588 R2 = 0.9426 F2,12= 115.9512 (2.5692) = In p = -0.7213, -0.0377 ryx2,xl= -0.5979 l/M = 0.2624 Supplementary Problems THE THREE-VARIABLE LINEAR MODEL 7.30 Table 7.11 extends Table 6.12 and gives observations on Y , X 1, and X2. How can the best test be performed? 10); therefore we accept the null hypothesis that the percentage of correct ... lies ... the Ho the 20 ... This is a two-tailed test and the acceptance region ... 20 ... ... (b) and ... the ... the ... of ... the mean ... 21 ... (c) In ... the equation and with s2, in the supply equation. In this case, the sampling distribution is a continuous discrete distribution or the empirical sampling distribution of ... the ... (a) ... of 20 4.5 ... 4, there ... (d) Since the endogenous variable P is also an explanatory variable in both the demand and supply equations, P is correlated with the error terms in both equations, leading to a correlation between ... (b) ... Hz ... Combinations and permutations are not used in process problems because these models are usually simpler than the regression line. The regression for ... (b) ... (c) Since the calculated value of F is 9.55 at a = 0.01, the alternative hypothesis is accepted at the 1% level of significance. 8.4 There are 6 red chips in the highway (10 mi at 60 mi/h) and 40 min on local streets (10 mi at 15mi/h) for a total of 50 min, and this is the correct answer we get by using p = ~ H - 3 - - a). 10) You're Reading a Free Preview Pages 325 to 327 are not shown in this preview. Use (1) = 4 (a) Ave Lo,LL and 2 BLUE? For example, there is perfect multicollinearity between Xl and X2 if X1 = 2X2 or X1 = 5 - (1/3)X2. 11 TIME-SERIES METHODS (a) The results are listed in Table 11.9. The method for generating Table 9.3. Note that the estimated coefficients ... Y1 = = 1.4884 Y2 = Y1 ~2 ~ 1.4884 0.2709 1.7593 Y3 = Y2 ~3 = ~ 1.7593 - 0.2714 = 1.4879 Y4 = Y3 ~4 = ~ 1.4879 - 3.6337 = -0.8758 etc. Other important criteria for a good estimator are discussed in Sec. The endogenous variables of the model are R and Y, while M is exogenous (determined outside the model). (7. Read in data (File/Import-Read Text, Lotus, Excel) ... a ... r*t + In the dialog box, we give the columns for Y and X variables (in this case the variables are on sheet). (b) ... to get ... Deskriptive Statistics. Multiple regression analysis is discussed in Chap. Therefore, we substitute ... the first is ... the ... 3 ... in ... (b) ... Distinguish between ... the 1% of the ... the ... to estimated. (c) ... 2.9 ... 10, ... Previous data can also be a ... To do this ... that the ... equation ... the mean ... the ... (a) ... the ... (b) ... If ... the hypothesis ... the ... at ... ... (c) ... is ... sample of 6 from ... 2 ... 1 ... that ... is not ... The regression result ... (b) ... The sample mean ... (c) ... 100 ... (see Prob. (a) ... a ... in the ... Z ... test ... 5% ... Supplementary Problems ... (a) ... (b) If ... the ... Fig. ... . 1. ... reject ... 8.19 ... 99 ... Table ... level ... 6.5 ... we use ... one ... that the ... fruit ... 11 ... 50 ... (b) ... 1.3 ... 13 ... ... the same ... Table 2 ... 2.8 ... 1999 ... 10 ... 9.0 ... of ... 1 ... (d) ... ... and ... level ... and ... 0.95 ... (a) ... 11.9 ... ... 4.3 ... 33 ... Find ... ... interval ... (c) ... (a) ... The ... Table ... ... from ... a ... 14 ... 6 ... 5 ... the ... Z ... ... consecutively to each quarter and letting Dl = 1 for the second quarter and 0 otherwise, D2 = 1 for the third quarter and 0 otherwise, and D3 = 1 for the fourth quarter and 0 otherwise, we get Table 8.19 Aggregate Reserves of U.S. Depository Institutions (in Millions of Dollars) Quarter Year I I1 I11 IV 1995 1996 1997 1998 1999 2000 57,571 54,878 47,551 45,591 43,229 39,752 57,03 1 53,742 46,606 45,094 42,331 39,217 57,162 5 1,045 46,060 44,199 41,314 39,257 57,896 51,174 47,919 45,209 41,655 Source: Federal Reserve Board of Governors. EXAMPLE 8. The outcomes of two successive tosses of a balanced coin are in&pendent events. The BLUE properties of the OLS estimator is often referred to as the Gauss-Markov theorem. 81 FURTHER TECHNIQUES AND APPLICATIONS IN REGRESSION ANALYSIS 199 (a) Since this is a probit model, we know that ui is normally distributed in the model of the underlying propensity of Y: where Yi = 1 if Y: > 0 and Yi = 0 if Y: < 0. 3, we get 0.1915 for the area from z = 0 to z = -0.5. Looking up z = 1, we get 0.3413 for the area from z = 0 to z = 1. (u) Two conditions at r e q c e d for at1 estitnator to be consistent: (1) as the satnple s k itlcitases, the estimator must approach more and more the true parameter (this is referred to as asymptotic unbiusedncss); and (2) as the sample size approaches infinity in the limit, the sampling distribution of the CHAP. (a) 3 workers (5) About 0.1 (using the Poisson approximation to the binomial distribution) CONTINUOUS PROBABILITY DISTRIBUTIONS: THE NORMAL DISTRIBUTION 3.64 Give the formulas: (a) the probability that continuous variable X falls between XI and X2, (b) the normal distribution, (c) the expected value and variance of the normal distribution, and (4 the standard normal distribution, (e) what is the mean and variance of the standard normal distribution? It should be noted that minimum variance by itself is not very important, unless coupled with the lack of bias. 12), what two data formats would be able to read the text file of the interest rate data? The most important measures of central tendency are (1) the mean, (2) the median, and (3) the mode. t x t " i s : F i l e Name=c :\ e x a m p l e l . 91 207 PROBLEMS IN REGRESSION ANALYSIS Neither ilnor i2 is statistically significant at the 5 % level. (4 Under what conditions can the Poisson distribution be used as an approximation to the binomial distribution? Two events, A and B, are not mutually exclusive if the occurrence of A does not preclude the occurrence of B, or vice versa. S., and the Longrnan Group Ltd., London, for permission to adapt and reprint Tables TII and IV from their book, Statistical Tables for W ological, Agricultural and Medical Research. (a) The probabilities of picking an item produced by the morning shift M and evening E are P (M) = 1600 = 0.625 600 P(E) =1600 = 0,375 and The probabilities of picking a defective item D from the morning and evening outputs separately are and The probability that an item picked at random from the total of 1600 items produced during the 24-h period was produced by the morning shift and is defective is P (M and D) (b) = P (M) . 2SLS involves regressing each endogenous variable on all the exogenous variables of the system and then using the predicted values of the endogenous variables to estimate the structural equations of the model. 3- 14 to the right of 78 must be 0.4. We must look into the body of App. Y = 2.62 + 27.12111 X (0.36) (5.90) Ans. The OLS estimator is widely used because it is BLUE (best linear unbiased estimator). Related to conditional probability is Bayes' theorem (see Prob. 7.23(a)]: where ryx, = simple-correlation coefficient between Y and X I , and ryx2 and rxIx2 are analogously defined. (b) Using R2 = 0.6932 from Prob. NONSTATIONARY SERIES 11.8 (a) What are the problems of a nonstationary series? (a) No (b) Yes Table 5.37 Soap Sales for Each of 3 Wrappings and 4 Formulas Formula Formula Formula Formula 1 2 3 4 Wrapping 1 Wrapping 2 Wrapping 3 87 79 83 85 78 79 81 83 90 84 91 89 NONPARAMETRIC TESTING 5.55 Using thc data from Tablc 5.35, would thc Wilcoxon signcd rank tcst rcjcct at thc 10% significance lcvcl thc null hypothesis that the median miles per gallon for type 1 gasoline is (a) 12 (b) 15? Table 1 gives the quantity supplied of a commodity Y at various prices X, holding everything else constant. CHAPTER 9 A Problems in Regression Analysis TTCOLLINEARITY nearity refers to the case in which two or more explanatory variables in the regression hly correlated, making it difficult or i~npossibleto isolate their individual effects on the e. 7. Given = 1000 h and a = 80 h and letting X , = 1120 h and X2 = 1180 h, we get ZI X I - p - 1120- 1000 =1.5 u 80 =-- and STATISTICAL INFERENCE: TESTINO HYPOTHESES [CHAP. (a) Type I1 error; type I error (b) a!; level of significance (c) /3 (d) 1 - a ! (e) /3 increases 5.34 Having set a! = 5%, when is a graduate school more likely to accept the hypothesis that the average Graduate Record Examination (GRE) scores of its entering class (a) Equal 600? ECONOMETRICS EXAMINATION 5. (b) Test the equality of the median illiteracy rate for South America and Asia at the 5% significance level. For the sample 56 1 = - =640 . Cell location in Excel is defined by the row number and column letter of each cell. All other probabilities are also affected, since there are now only 51 cards in the deck. Supply: Qt = bo + al P, + a2Y, + a3W , + ul, + b~Pt + ~ 2 , where W , is wealth and the expectation is that a3 > 0. The ARIMAProcedure Name o f V a r i a b l e = y M e a n o f w o r k i n g s e r i e s 0.033797 StandardDeviation 2.122958 Number of O b s e r v a t i o n s 220 Autocorrelations Covariance Correlation Std Error " . 6.30 is not a reliable estimate of bl. 5 Since z = (p,- E) / U ~=,(0.5 - ~- 0.4)/0.1 = 0.10/0.10 = 1, we accept Ho, thatpl =f i , with os = 0.05. Thus P(955 iX i975) = 0.2123 - 0.1217 = 0.0906, or 9.06%. n.M (6) '1'0 correct for autocorrelation, the same procedure is used as for the standard regression, except that "ar(1)""s included in the regression equation. Looking up zl = 1.25, we get 0.3944. The normal distribution often can be used to approximate other distributions, such as the binomial and the Poisson distributions (see Probs. Then and assuming n < 0.1 . I I 0 0.83 z scale Rg.44 4.12 I -1.73 I z scale 0 Rg.4 7 A small local bank has 1450 individual savings accounts with an average balance of $3000 and a standard deviation of $1200. is relatively high. 10 With reference to the demand-supply model in Prob. We owe a great intellectual debt to our former professors of statistics and econometrics: J. (8.11) and (8.12) Y,=a + b o x , + b l X r p l + b 2 X r p 2+ b 3 X r p 3+ u, bi = co + cl i + c2i2 with i = 0 , 1 , 2 , 3 and substituting Eq. (8.12) into Eq. (8.11), we get Rearranging the terms in the last expression: 3 3 3 and letting Z,, = C XI-,, Z2,= C iXrpi,and Z3,= C i 2 x r P i we , get i=O i= 1 i= I Y,=a + coZlt + C I Z +~ C, Z Z+~u,, (b) With a four-period lag taking the form of a third-degree polynomial, we have + b o x , + blXr-1 + b2Xr-2 + b3Xr-3 + b4X,-4 + u, with i = 0, 1,2, 3 , 4 bi = co + c l i + c2i2 + c3i3 Y,=a Substituting the second into the first, we get Rearranging the terms in the last expression, we have and letting the terms in parentheses equal, respectively, Z , , , Z2,, Z3,, and Z4,,we get 8.15 Using the data from Table 8.15 and assuming a three-period lag taking the form of a seconddegree polynomial, (a) Prepare a table with the original variables and the calculated Z values to be used to estimate the Almon lag model. (b) In Fig. 8.4 FORECASTING Forecasting refers to the estimation of the value of the dependent variable Y F given the actual or projected value of the independent variable XF. (c) To test the statistical significance of bo and bl, the t distribution (from App. If Y , = Y t p 1 E Y , and X, = Xtp1 EX,, we see that Y and X have a unit root. / (a) Using ~ j =:27.727 from Prob. Some examples are the number of telephone calls to the police per hour, the number of customers arriving at a gasoline pump per hour, and the number of traffic accidents at an intersection per week. 8.9, 8.26, and 8.27). 64 Critical region (see R Cross-sectional anal Cross-sectional data, Cumdative nor-ma1 [&on model), 184, 199 I Copyright 2002 The McGraw-Hill Companies, Inc. (b) E(X) = 4 students, SD X = 1.79 students, and the distribution is skewed to the right. The result fails to reject the null that both means are the same at the 5% level of significance. 8 Table 8.20 Disposable Income and Percent of College Graduates in the East in 1998 Disposable income, % Percent with college degree, % State 19.76 24.99 20.77 26.72 23.02 30.22 26.06 28.31 22.79 19.2 26.6 27.1 31.0 27.8 31.4 26.8 30.1 22.1 ME NH VT MA RI CN NY NJ PA Region Mid-Atlantic New England Disposable income, % Percent with college degree, % State 24.96 24.9 23.0 17.12 20.49 18.52 21.27 22.06 25.1 31.8 30.3 16.3 23.3 21.3 20.7 22.5 DE MD VA WV NC SC GA FL Region South Atlantic Source: Statistical Abstract of the United States (b) lo= 8.16 for New England and South Atlantic states, while lo= 10.99 for Mid-Atlantic states. (b) What are some possible problems with delimited data? 21 25 DESCRIPTIVE STATISTICS Note that the average deviation takes every observation into account. (b)Why are A R and M A processes referred to as stationary processes? 4 CONFIDENCE INTERVALS FOR THE MEAN USING THE t DISTRIBUTION 4.49 Find the t value for 29 df for the following areas falling within the (right) tail of the t distribution: (a) lo%, (b) 5%, (c) 2.5%, and (d) 0.05%. + (59 - 55)' = 838 The df of SSA = c - 1 = 4 - 1 = 3; df of SSE = (r - 1)c = (8 - 1)(4) = 28; and df of SST = rc - 1 = 32 - 1 = 31, which is the same as the df of SSA plus the df of SSE. 0 1 ~ ~R2 = 0.94 (-9.84) (10.42) DUMMY VARIABLES 8.6 (a) Write an equation for peacetime and one for wartime for Eqs. STANDARD NORMAL DISTRIBUTION [APPENDIX 3 : Areas reported below: * P z Proportions of Area for the Standard Normal Distribution *Example: For z - 1.96, ahnded area ia 0.4750out of the total area of 1.0000. X could move before Y in anticipation of Y moving. Similarly, if the proportion of defective items is greater for the evening than for the morning shift, the probability that an ikm picked at random from the evening output is defective is greater than for the morning output. 4.51 4.54 Indicate which distribution to use in order to find confidence intervals for the unknown population mean from a random sample taken from the population in the following cases: (a) n = 36 and s = 10, (b) n = 20 and s = 10 and the population is normally distributed, and (c) n = 20 and s = 10 and the population is not normally distributed. IDENTIFICATION 10.5 (a) What is meant by identification? Similarly, we need data on the average price of coffee, consumers' income, and the price, of say, tea (a substitute for coffee) per year from 1960 to 1980. That is, among all unbiased linear estimators, it has the lowest variance. These tests are inefficient if the distribution is known or the sample is large enough for a parametric test. 6-2). Am. St&. Specifically, E(v,vrp1) 0 because v, and vrplare both defined with u r 1in common (i.e., v, = u, - hutpl and vrpl = urpl- hurp2). 11.26. Find thc gcomctric mcan of thc inflation ratcs (thc geometric mean, pG or XG,of a sct of n positive numbers is the nth root of their product and is used mainly to average rates of change and index numbers): CHAP. 3.29, find the probability that starting at a random point in time, the first customer stops at the gasoline pump within a half hour. Table 2.27 Calculations for Covariance Employee Number Hourly Wage X , $ Years of Schooling Y 1 2 3 4 5 6 7 8 9 10 8.50 12.00 9.00 10.50 11 .OO 15.00 25.00 12.00 6.50 8.25 12 14 10 12 16 16 18 18 12 10 - X = 11.775 - Y = 13.8 (X - 7) (Y - 7) -3.275 0.225 -2.775 -1.275 -0.775 3.225 13.225 0.225 5.275 -3.525 -1.8 0.2 -3.8 -1.8 2.2 2.2 4.2 4.2 1.8 -3.8 (x-X)(Y -7) 5.895 0.045 10.545 2.295 - 1.705 7.095 55.545 0.945 9.495 13.395 C(X -X) (Y - 7)= 103.55 CHAP. (b) What is the long-run effect of X on Y? = 5000 lb, the more likely the producer is to accept the hypothesis at the specified level of significance, a. Answers 1. Other frequently used wddence intervals are the 90 and 99% levels, wrresponding to the z values of 1.64 and 2.58, respectively (see App. (a) Should the recruiting oficer accept the hypothesis? 7.30, 8.1 1): Y , = a(l - h) + b o x t + h Y t p 1+ v , (8.10) where 0 < h < 1 and v, = u, - hutPl. However, Eq. (8.10) violates two assumptions of the OLS model and results in biased and inconsistent estimators that require adjustment (see Sec. (a) The demand equation is underidentified because it does not exclude any exogenous variable. 5 with hypothEXAMPLE 1. Custom class intervals may be typed into Excel and indicated as the "Bin Range." The results are as follow: Parts 6 and c are reported in the image below as both numerical results and Excel formulas. (a) < 0.1 1, or 11% (using Chevyshev's theorem) (b) at least 0.89, or 89% 3.73 According to Chebyshev's theorem, at least what proportion of the observations fall within standard deviations from the mean, (b) 2.5 standard deviations from the mean? (a) Here Ho:p = 0.80 and H I :p > 0.80. = p. (a) PROBLEMS IN REGRESSION ANALYSIS CHAP. 3 deals with probability, Chap. I 1,' ' I ;r,r,,+ +.r ~ r P " 5 & ct .eg.e< c mqn '18m m * I U ~ . This large-sample property of consistency is used only in situations when smallsample BLUE or lowest MSE estimators cannot be found. (b) Graph the unit-root series and the original error term on the same axis. (a) Consistent parameter estimates of the regression of I.; on X, are Lo = -91.43 and Ll = 0.90. The outcomes are therefore mutually exclusive. What is the probability that an item picked at random from the total of 1600 items produced during the 24-h period (a) Was produced by the morning shift and is defective? 117, to the nearest whole number 3.69 Past experience indicates that 60% of the students entering college get their degrees. R2 = 0.596 7.39 For R2 = 0.40, n = 10, and k = 5, find R2. (a) Find PG. 6.14 and s ~ are, , respectively, the standard deviations of S" 4l Prove that (a) mean = b l , and (c) mean bo = bo, and (a) bl=-- - Cxiyi C4 = (b) var 4l = oil (4 var bO= a : (C.Xi(Yi - 7) - C4 CxiYi - Cx' x? 61 -1 5 r 5 + l . Several sources of online data are used, and Web addresses are given for the student's and researcher's further use (App. Alpha is thc dcsircd lcvcl of significance. The continuous buildup of the errors creates the problem that a nonstationary series will tend toward an infinite variance. The reason for this is that if we take many samples from a population, the average of the sample variances does not tend to equal population variance, a 2 , unless we use n 1 in the denominator of the formula for s2 (more will be said on this in Chap. 9.21 Table 9.18 gives the level of gross fixed capital formation Y, sales XI, both in thousands of dollars, and a productivity index X2, for 35 firms in an industry. 2 NorthCarolina, 7.7 NorthDakota, 0.6 Ohio, 11.3 (a) Fixed format, state in columns 1 to 14, population in columns 16 to 19. 10.3(a)] are bl = = ,-I' n3 A 0'0039 = 1.5000 [see Prob 10.7(c)] 0.0026 where ho and hl are consistent estimators of ho and h l , respectively, and the structural supply equation (estimated by ILS) is (c) Regressing Q, on P, directly, we get CHAP. Data may be read from external files (see Example 2) or typed directly into the cells by clicking on the cell and typing the text or number desired. 1 (the table of binomial probabilities), (b) p = n p = (30)(0.3)= 9 persons, and r r = d w = d (3 0) (0 . 3 8 3 . 4.15, where the standard normal distribution was used. (a) What is the relative frequency of the 3? (a) Taking South Atlantic as the base, D l = 1 for New England states and 0 otherwise and D2 = 1 for Mid-Atlantic states and 0 otherwise, we get FURTHER TECHNIQUES AND APPLICATIONS IN REGRESSION ANALYSIS [CHAP. M~= -201 -80 (-6.48) + 0.14 GDP, (29.44) 2 = 0.98 d = 0.54 Since d = 0.54 idL = 1.20 at the 5% level of sigdicance with n = 20 and k' = 1 (from App. The values $3.595, 3.695, 3.795, etc. we get 0,1217. (b) Unbiased estimator? (c) for unrelated data? (a) Regressing I on S and R for the entire sample of 35 firms, we get To test for heteroscedasticity, we regress I on S and R for the first 14 and for the last 14 observations, leaving the middle 7 observations out, and we get = 2.82 at the 5% level of significance (see App. nonstrike, male vs. The approximation is sufficientlygood for n 30. The fust reduced-form equation can be derived by substituting the sewnd equation into the first and rearranging: ; The sewnd reduced-form equation wn be derived by substituting the first equation into the sewnd and rearranging: Copyright 2002 The McGraw-Hill Companies, Inc. The small sample precludes the use of the chi-square goodness-of-fit test but the nonparametric KolmogorovSmirnov test may be used to test Ho: normally distributed c(. (b) If cointegration exists, the long-run process should be used to explain the dependent variable. Starting at t = 1, y, = 1.4884. (The reader without knowledge of calculus can skip this problem.) (a) The definition of price elasticity of demand is The derivative of the Q function with respect to P is Substituting the value of dQ/dP into the formula for q p , we get (b) The definition of income elasticity of demand is The derivative of the Q function with respect to Y is Substituting the value of dQ/dY into the formula for q y , we get 8.3 Table 8.6 gives the output in tons Q, the labor input in hours L, and capital input in machinehours K, of 14 firms in an industry. Then P(X)= n! ,px(l X!(n- X) . Xacale P EXAMPLE 3. using (a) ~ j ? If there is no unit root in the error term from the regression Y , = bo bl X, u,, then Y and X are cointegrated. 12 The wmlogsam is found in Quick-Series Statistics-Correlogram. What is the minimum sample size required if other polls indicate that the proportion voting for this candidate is 0.30? (b) Just as in other probability distributions (see Secs. Another disadvantage is that in actual empirical work, neither the period nor the form of the lag may be suggested by theory or be known a priori. Aceeptance region for H, -- Rejection region for H, Fig. 3-1 1). 4.8 A population of 12,000 elements has a mean of 100 and a standard deviation of 60. 5) ~ 14 12 6 5 + + + Since the calculated x2 is larger than the tabular value of x2 with w = 0.10 and df = 1, we reject Ho, that males and females over 65 continue to work in this town independently of whether they are above or below 70 years of age. (B) For descriptive statistics of a series, choow Quick-Series Statistics-Histogram and Stats. r > 0 indicates that X and Y change in the same direction, such as the quantity supplied of a commodity and its price. 121 COMPUTER APPLICATIONS IN ECONOMETRICS SAS 12.24 From the estimationin Prob. 7.29 and X I F = 2 and X2F = 1250 for 2000 (a) find s$ and (b) the 95% confidence interval for Y F , given that p = 82.27 - 5.11XI + 0.02X2, X I = 6, X2 = 1100, s2 = z e 2 / n - k = 226.32112 r 18.86, s; % 1.41, s; % 0.01, s? Regressing Y on X, we get Pr = 126.06 + 1.03Xr (16.68) R~ = 0.94 d = 0.58 Table 9.3 Inventory and Sales (Both in Billions of Dollars) in U.S. Manufacturing 1979-1998 Source: Economic Report of the President. (b) Three tails in 3 flips of a coin? 9-2b, there is negative fistorder autowrrelation. (c) Derive the formula for the expected value and variance of a continuous probability distribution. Testing for unit root in levels yields Since ti1 = -2.48 > -3.33 (from App. Acceptance region for Ho Rejection region for Ho Rg.5-2 In the past, 60% of the students entering a specialized college program received their degrees within For the 1980 entering class of 36, only 15 received their degrees by 1984. Text format is flexible since any statistical package and brand of computer can read it. 5.23, find the value of SSA, SSE, and SST; the degrees of freedom for SSA, SSE, and SST; and MSA, MSE, and the F ratio. Therefore, the partial correlation coefficient is merely the simple correlation coefficient between the residuals Y* and X ; (that is, ryxl.x2 = rye,;). 7.30, find (a) ryxl.x2and (b) variable contributes more to the explanatory power of the model? TESTING HYPOTHESES ABOUT THE POPULATION MEAN AND PROPORTION 5.4 A producer of steel cables wants to test if the steel cables it produces have a breaking strength of 5000 lb. X would Granger-cause Y, but it is the movement in Y which is the true cause. (b) If so, what is the short-run magnitude of the causality? (a) We are interested here in finding the number of combinations of 8 people taken 3 at a time without concern for the order In general, the number of arrangements of n things taken X at a time without conern for the order is a combination given by nCx = where n! (read n factorial) =n (2 = X!(n n! - X)! . : I ..- a- 22 ,,.. A car dealer has collected the data shown in Table 5.2 on the number of foreign and domestic cars purchased by customers under 30 years old and 30 and above. Source: World Bank World Development Indicators. This may makc it difficult or impossiblc to isolatc thc cffcct that each of the highly collinear explanatory variables has on the dependent variable. It is important that all the independent variables are in a continuous range of columns. 4.17? (a) How many different committees of 3 members each can be formed from the club? The remaining 3% is attribute$ to factors included in the error term. 4.5(b). 9.1 5). There are nonparametric tests appropriate for most scales of measurement, and for nonstandard functional forms and distributions. When k > 1, (n - l)/(n - k) > 1 and R2 > R2. Rejection region Aweptance region Fig. However, the OLS estimated coefficients are still unbiased (if the model is properly specified). 3.20 and 3.31) is equal to the population parameter being estimated. ' 1 , .-I C ITIITF 1 r I I n ,-- .- , , . 6) ~ = 116 - 112.36 fi = % = 3.64 applications squared 1.91 applications Table 3.4 Calculations to Find the Expected Value and Variance Number, X Days, f p(x) xp(x) x2 x2p(x) 7 8 10 11 12 14 10 10 20 30 20 10 0.1 0.1 0.2 0.3 0.2 0.1 0.7 0.8 2.0 3.3 2.4 1.4 49 64 100 121 144 196 4.9 6.4 20.0 36.3 28.8 19.6 N=Cf=100 CP(X)=1.0 CXP(X)=10.6 E(X)=p 3.22 = cX'P(X) = 116.0 C X P (X) = 10.6 applications (a) State the conditions required to apply the binomial distribution. (a) For the initial regression in levels, we obtain [CHAP. Find (a) the 90%, (b) the 95%, and (c) the 99% confidence intervals for the unknown population mean. 4.2 What is meant by random sampling? (a) For each value of p r po,the probability of rejecting Howhen Ho is false is given by 1 - p, where /3 was found in Prob. 24 DESCRIPTIVE STATISTICS Kurtosis = Kurtosis = and ' ' [CHAP. we X ~ need , to find first r y x 1 ,ryx2, and r x I x 2 . However, it is not easy to calculate the standard errors of the structural parameters, nor can ILS be used in cases of overidentification. When j E 1, autocorrelation can be corrected by rerunning the regression in difference form and omitting the intercept term (see Prob. 4.27 A random sample of n = 25 with X = 80 is taken from a population of 1000 with a = 30. To do this, once again, the producer takes a random sample of the cable produced (i.e., ~ c > and tests the mean breaking strength X. Unbiased estimates of the variance of Lo and ilare then given by SIMPLE REURESSION ANALYSIS CHAP. We select to add a linear trendline; under the "Options" tab wc can scloct to havc thc rcgrcssion quation and 132 rcportcd. These are now the reduced-form equations. P (R / R) . However, the number of pwple is a discrete variable, In order to use the normal distribution, we must treat the number of pwple as if it were a wntinuous variable and fmd P(X 1 9.5). 6-5. Moreover, the lagged explanatory variables in a distributed lag model are likely to be strongly correlated, so it may be difficult to adequately separate their independent effects on the dependent variable [see Prob. = 6 - - X=9 STATISTICAL INFERENCE: TESTING HYPOTHESES CHAP. 4.15(b) and 4.25(b)? (a) 1312652 or 11204 (b) 61132 or 1/22 (c) 1/22 (d) 1/11 (e) 611320 or 11220 3.54 Suppose that the probability of rain on a given day is 0.1 and the probability of my having a car accident is 0.005 on any day and 0.012 on a rainy day. 4.17), we get z a / (X - p) = f i . , n E(uiuj) = 0 E(Xiui) = 0 - (See Prob. To do this, it takes a random sample of 25 recruits and finds that the average weight for this sample is 85 kg. (d) Below 60? (a) The classical or a priori approach. (b) Create a variable for the ferti1izer:eorn ratio. Given the population standard deviation or its estimate, and given that the population is normal or that a random sample is equal to or larger than 30, we can find the 95% confidence interval for the unknown population mean as This states that in repeated random sampling, we expect that 95 out of 100 intervals such as Eq. (4.4) include the unknown population mean and that our confidence interval (based on a single random sample) is one of these. As we move further away from the mean in both directions, the normal curve approaches the horizontal axis (but never quite touches it). The sample standard deviations [as defined in Eqs. Then Since the calculated t value falls within the rejection region, the agency should reject Ho and accept the consumers' complaints, H I ,Note that if ar had been set at 1%, the rejection region would lie to the left of t = -2.896, leading to the acceptance of Ho . Ncxt, thcsc diffcrcnccs arc includcd as an additional variablc Since all variables in the error-correction model are stationary, OLS may be used. A child is born either a boy or a girl. In a simple toss of a die, we get one and only one of six possible outcomes: 1,2, 3 , 4 , 5 , or 6. Of course, in the real world it is rarely known what error of measurement might be present (otherwise, the errors could be corrected before running the regression). In all problems of this nature it is helpful to sketch a figure. Skcwncss can also be measured by the third moment [the numerator of Eq. (2.14a,b)] divided by the cube of the standard deviation: Sk = C f (X- 1'' d and for populations ~k = C f (x- g3 for samples 2 For symmetric distributions, Sk = 0. Since the hospital should accept H,, that p = 0.90, at the 1% level of significance. 3-8 Tree Diagram for Rdling Two Dice Simultanewsly (c) 3.15 Since 2 balls, one of which was red, were already picked and not replaced, there remains a total of 8 balls, of which 4 are red, in the urn. Student's Distribution I A r w rep& below: Proportions of Are I for the t Distributions ed area to reprmnt 0.05 o f the total area o f 1 .O, value o f t with 10 degrees o f freedom is 1.812 o f Fisher and Yates, Statistical TablesforBiological, Agricdtural and Medical Research, 6th ed., 1974, roup Ltd., London (previously published by Oliver & Boyd, Edinburgh), by permission of the authors Copyright 2002 The McGraw-Hill Companies, Inc. Since in this case cov(X, Y) > 0, X and Y move together relative to their means. (a) Parameter bo is the constant term or intercept of the regression and gives the estimated value of Y,, when Xli = X2i = 0. 111 TIME-SERIES METHODS Supplementary Problems ARM A 11.24 Using the random variable from Table 11.13, and an AR(2) process for y, with yl = 0.4 and y2 = -0.3, (a) calculate y4 (b) y8 (c) Y ~ 0 (4y30. What do they depend on? 111 TIME-SERIES METHODS 247 running a regression. = number of e ~ p l ~ n a t o vnrioble3 ry e~oludingthe mnstont term. (a) Specification of the theory in stochastic equation form and indication of the expected signs and sizes of estimated parameters (b) Collection of data on the variables of the model and estimation of the coefficients of the function (c) Economic, statistical, and econometric evaluation of the estimated parameters 1.23 What is the first stage of econometric research? Since a discrete random variable can take on only a finite number of values, we can list the values and their associated probabilities. Schaum's Outline of Theory and Problems of Econometrics. Such specification problems may be detected and corrected through the various econometric methods (see Chaps. 6 and 7). Dynamic models are more difficult to estimate and interpret than static models, and the interpretation of the estimated parameters depends on the specification of the model. (c) We determine the expected number of defective items in the sample by summing over the possible outcomes.

cannot reject the null hypothesis that there is a unit root for e,. Nnte that in SAS "I*" and 'I" enclose comments which are not read by SAS. & A. We get the dialog box &low. Hypothesis testing? 3-3 Probability Distribution of Heads in Two T o w of a Balanced Coin EXAMPLE 10. 4.17 if we had not been told that the population was normally distributed? (b) Which are the endogenous and exogenous variables? (b) At the 1% level of significance, the calculated z would fall within the acceptance region for Ho (see Fig. 9, the critical values at the 5% significance level are 41 and 78. (a) Normal distribution (invoking the central limit theorem and using s as an estimate of a) (b) The t distribution with 19 df (c) Chebyshev's theorem CHAPTERS 4 Statistical Inference: Testing Hypotheses otheses about population characteristics (such as p and a) is another fundamental aspect rence and statistical analysis. 9-lb, increases with X,. An unbiased estimate of the forecastwhere n is the number of observations and a error variance $ is given by where .? Because an interval estimate also expresses the degree of accuracy or confidence we have in the estimate, it is superior to a point estimate. The exogenous variables and the lagged endogenous variables are sometimes called predetermined variables. 1 F-F a 4 e "I cr'?.(,?q , -,: *.c~I: I -.pr 3s. They are called parameters. 6.3 TESTS OF SIGNIFICANCE OF PARAMETER ESTIMATES In orderAto tes; for the statistical significance of the parameter estimates of the regression, the variance of bo and bl is required (see Probs. (b) The endogenous variables of the model are Q and P. How large a sample would the firm take? The relative frequency of heads is 53/100, or 0.53. + With constant returns to scale, the CobbDouglas prduction function can be rewritten as Expressing this production function in double-log form and rearranging it, we get Setting In Q* = In Q - 1nK and In L* = In L - In K and then regressing InQ* on In L*, we get HETEROSCEDASTICm 9.4 (a) What is meant by heteroscedhticity? 5 A more concise way of expressing the above is F= variance between sample means variance within samples - 37.33 25.93 - The value of F from App. (a) 0.899, or 89.90% (b) 0.95 (c) 0.9902 (4 0.1662 (e) 0.8159 (,f) 0.0179 (g) 0.8338 3.66 A random variable is normally distributed with p = 67 and a = 3. Fit an OLS regression line to the data in Table 8.8. Table 8.8. Millions of Persons Employed in the United States from 1980 to 1999 Source: Bureau of Labor Statistics. It is thus better than the range, but it is not as widely used as the other measures of dispersion. Find the 90% confidence interval for the mean weight of the population of army recruits. (a Table 6.10 Per Capita Income (Y, $QO) and Percentage of the EEonomy in Agriculture (X) *Key: (1) Argentina; (2) Bolivia; (3) Brazil; (4) Chile; (5) Colombia; (9 Dominican Republic; (7) Ecuador; (8) El Salvador; (9) Honduras; (10) Mexico; (1 1) Nicaragua; (12) Panama; (13) Peru; (14) Uruguay; (15) Venezuela. 3) 0.4750 of the area under the standard normal curve from p = 0 to z = 1.96. = 0 and (2.16) c is the width of the class intervals). We then test for reverse causality: Since the coefficient on AY,-, is not significant at the 5% level of significance, we conclude that Y does not Granger-cause X . 2000 (base year = 1982-1984). In the two independent or explanatory variable multiple regression, k = 3. 7.5(a),find (a) ryxl.x2and (b) r y x 2 . - p)n-x where n! (read "n factorial") = n . (a) In Y = ln bo b l X u (b) Y = bo b l R + u, where R = ln X (c) Y = bo - b o Z + u, where Z = l / X (d) Y = b o + b l X - b 2 W + b 3 T + u , w h e r e w = x 2 a n d ~ = x 3 + + In Y = 2.64 + 0.72 In X (14.69) (6.31) Fit a semilog function of the form Y = bo R2 = 83.26% + b l 1nX + u to the data in Table 6.12. 8.23 for bl = 0.0018 and bo = (a) -3.8 (b) -3.6 (c) -3.4. Ans. It takes the absolute value (indicated by the two vertical bars) because C (X - p) = 0 (see Example 5). The average deviation, variance, standard deviation, and coefficient of variation for the frequency distribution of weights (grouped data) given in Table 2.3 can be found with the aid of Table 2.6 (X = 20.08 oz; see Example 4): c f(x -X12 S = n-1 2.9520 2 0 . 2 to 30, 2001. (b) Test X , from Prob. 4.21). (c) Perform an ANOVA test of the null hypothesis that all three wrappings have the same average sales. (b) What is its usefulness? 3WO. 11 are used to find the critical value for the ADF test. The total variation in I' or total sum of squares TSS = C(I: - T)2 = y'. (4 What is the distinction between a probability distribution and a relative-frequency distribution? Another important advantage is that 2SLS (but not ILS) gives the standard error of the estimated structural parameters directly. 10.3(u): SIMULTANEOUS-EQUATIONS METHODS CHAP. Since the average value of u is assumed to be 0, Eq. (6.1) gives the average value of Y. 9 8 and Sb1- - -4 i X r 0 . It also refers to the presentation of a body of data in the form of tables, charts, graphs, and other forms of graphic display. No (maximum difference = 0.391) Is the null hypothesis of equality of Statistics Examination 1. Thus the coefficient of variation can be used to compare the relative dispersion of two or more distributions expressed in different units, as well as when the true mean values differ. 31 PROBABILITY AND PROBABILITY DISTRIBUTIONS If P(A') represents the probability of nonoccurrence of event A, then P(A) + P(A') = 1 EXAMPLE 1. However, an equation of a system is overidentzjied (or underidentzjied) if the number of excluded exogenous variables from the equation exceeds (or is smaller than) the number of endogenous variables included in the equation minus 1 (see Example 3). 4.17 A researcher wishes to estimate the mean weekly wage of the several thousands of workers employed in a plant within plus or minus $20 and with a 99% degree of confidence. The F test therefore requires two regressions to be run Unrestricted: + A i r = 6922.06 - 0.58XrPl 134.17t (-2.77) (2.77) + 0.33 A X r P l+ 0.28 A X r P 2 (1.54) (1.31) R 2 = 0.29 ESS = 26,483.44 Restricted: A% = 219.57 + 0.05 A X r P l- 0.02 A X r P 2 (0.22) (-0.09) R~ < 0.01 ESS = 37,225.56 Calculating the F statistic, we obtain Since F = 4.46 < 7.24, we accept the null that X follows a unit-root process. 4.4 and 5.2). In order valid, the sample must be representative of the population and the probability of error also statistics is discussed in detail in Chap. As a result, the regression line passes through point XY (see Fig. In Eq. (6.1), bo and bl are unknown constants called parameters. Thus the relative frequency or empirical probability of the 3 is 106/600 % 0.177 in 600 rolls. 7.14) Since the inclusion of additional independent or explanatory variables is likely to increase the RSS = C j ? The expected numkr of heads in 6 flips = p = np = (6)(1/2) = 3 heads. 71 155 MULTIPLE REGRESSION ANALYSIS Estimator i1measures the change in 7 for a unit change in XI while holding X2Aconstant.Ai2 is analogously defined. March April May 1566 1510 1520 1546 1390 1442 1494 1437 1492 Y 1457 1355 1486 5530.0 X 5342.1 5351.2 5358.7 5368.2 5361.5 5397.4 5454.0 5464.9 5467.3 5484.8 5506.5 1998 Year 1998 1998 1998 1998 1998 1998 1998 1998 1998 1998 Dec. 24(f241 Inductive reasoning, 2 Inferential statistics, 1-3 (See also Estimation; Hypothesis testing) Infinite population, 74, 84 Instrumental variables, 209-210, 22 1-222 Interquartile range, 13, 24 Interval estimates, 69-70, 7G81 Inverse least squares, 222 Joint moment, 16 Joint probability, 38 Kolmogorov-Smirnov test, 9 4 9 5 , 118-119, 123 Koyck lag model, 183, 193-194, 204 Kruskal-Wallis test, 94, 117-1 18, 123 Kurtosis, 1S16, 31, 35 Lack of bias, 147-149 Lagged variables, 209-210, 221-222 (See also Simultaneous-equations methods) Left-tail test, 88, 98, 102 Leptokurtic curve, 15, 31 Likelihood function, 199 Likelihood ratio index, 185-186, 200 Linear regression analysis, 128, 134 (See also Regression analysis) Linear relationship, 154 Log-likelihood function, 184, 198-199 Logistic function (logit), 1 8 4 185 Logit model (logistic function), 184185, 199-200 Marginal effect, 185, 200, 205 Mathematics, 1, 4, 7 Matrix notation, 159-160, 173-175, 179 Maximum likelihood, 184, 199 Mean(s), 11, 12 and analysis of variance, 92 in binomial distribution, 39, 5 4 5 5 confidence interval for the, using r distribution, 7(f71, 81-84, 86 in descriptive statistics, 1S16, 19-24 Mean(s) (Conr.): of error term in simple regression analysis, 128 hypothesis testing for differences between two proportions or, 89-90, 101-104, 120 in normal distribution, 4 1 4 2 of normal distribution as a continuous probability distribution, 57-58 in Poisson distribution, 40, 57-58 balanced coin. The arithmetic mean or average, of a population is represented by p (the Greek letter mu); and for a sample, by 7 (read "X bar"). (e) A nonred ball? In the resulting dialog box, select "Histogram" and click "OK." We then choose the options we want for our histogram in the following box: Our data are in wlumn B, from row 1 to row 10. 5.4 CHI-SQUARE TEST OF GOODNESS OF FIT AND INDEPENDENCE The X2 (chi-square) distribution is used to test whether (1) the observed frequencies differ "significantly" from expected frequencies when more than two outcomes are possible; (2) the sampled distribution is binomial, normal, or other; and (3) two variables are independent. 0.04804 to 0.051960 5.37 A easy recruiting center knows from past experience that the height of recruits is normally distributed with a mean of 180 cm (1 cm = 1/100m) and a standard deviation of 8 cm. The benefit of the logistic distribution is the ease of calculations since there are no tables required to find the cumulative probability. (b) How can you estimate p, the coefficient of autocorrelation? (b) Draw a figure showing positive and negative firstorder autocorrelation.] - F) 2 (between-column variations) + + + + = 3[(19 - 14) - (16 - 14) + (13 - 14) - (12 - 14)] - (10 - 14)-] = 3(25 + 4 + 1 + 4 + 16) = 150 = 5[(- (1 6 - 1 4) 2 + (1 4 - 1 4) 2 + (1 2 - 1 4) 2] = 5 (4 + 0 + 4) = 40 ESS = (Y - F)' = TSS - SSR = 194 - 150 - 40 = 4 These results are summarized in Table 5.26. (c) Estimate the population variance from the variance within the samples or columns. (b) Structural or behavioral equations describe the structure of an economy or the behavior of some economic agents such as consumers or producers. 3.42 What is the probability that in tossing a balanced coin we get (a) a tail, (b) a head, (c) not a tail, (d) a tail or not a tail? Wilcoxon and R. 5.25, construct an ANOVA table similar to Table 5.4. (b) Test at the 1% level of significance the hypothesis that the means for factor A, populations (fertilizers) are identical. The default, . -, -, = , (q = ~ ~ ~ I F I E I 4.1' . ~=s:. E(X) = p SD x = = np = d (150.85) = 12.75Acceptable items m = J(150.85) (0.15) = 1 = 1.38 acceptable items Because p > 0.5, the probability distribution is skewed to the left. 10.12(d) 43.53 17.53 35.73 38.33 22.73 30.53 27.93 33.13 12.33 46.13 47.38 33.38 33.35.73 46.13 32.47 -7.53 82.7 8.67 -2.17 21.53 32.53 -2.47 -14.13 -14.53 -39.93 -14.13 -4.33 -2.13 12.47 -7.53 -4.33 23.27 -0.53 ... 0.291 0.27 0.24 0.23 1.36 -0.363 0.356 0.32 0.29 0.27 1.63 J; = 1 2 3 4 5 6 7 8 9 10 11 30 35 = 35 E(Y) Source: F. (b) What are the conditions required in order for statistical inference to be valid? = 493 and a = 80.72. What is the lowest grade point that can be designated an A on the midterm? 11.11 For a unit root using the F-test form of the ADF with a trend. 61 SIMPLE REGRESSION ANALYSIS Table 6.4 Aggregate Consumption (Y) and Movable Income (X) O*110 Year X YI XI 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 ... When we wish to test the hypothesis that the population means are equal. Appropriate Statistical Procedures, American Chemical Company, 1964 For larger sample sizes, standard normal tables can be used for the test statistic. APPENDIX 10 -3 ~~atmo~ovov~mirnov Critical Values Kolmogorov-Smirnov~Critical Values for Various Significance Levels ... 0.388 0.368 0.352 0.338 0.325 0.314 0.304 0.295 0.286 0.05 0.975 0.842 0.708 0.624 0.565 0.521 0.486 0.457 0.432 0.410 0.391 0.375 0.361 0.349 0.338 0.318 0.301 0.995 0.929 0.828 0.733 0.669 0.618 0.577 0.543 0.514 0.490 0.468 0.450 0.433 0.418 0.404 0.392 0.381 18 0.278 U.3W 0.371 19 20 25 0.272 0.61 0.274 0.202 0.21 0.27 ... 810.5409 41.9609 c: = 2872.8535 c x ' = 2257 c y = 5879.7335 SIMPLE REGRESSION ANALYSIS CHAP. (b) We can use the t distribution (for the given degrees of freedom) when n > 30 but r is not given and the population from which the sample is taken is normally distributed. Some Approximate Statistical Procedures, American Cyanamid Company, 1964 For larger sample sizes, standard normal tables can be used for the test statistic. ... 158 166 178 D i s p d I e income X Rg.4d when t refers to t-tests with one degree of freedom ... the proportion of households in the population; (c) Write the summary results on the data in Table 6.4 in the answer sheet. ... (Y - 1) (d) The relationship between ... 9.3 AUTOCORRELATION When the error term in one time period is positively correlated with the error term in the previous time period, we face the problem of (positive first-order) autocorrelation. ... of significance? (d) Simultaneous-equations models refer to relationships among economic variables where one variable affects another and such that the economic variables in the various equations are jointly determined. This can be conveniently subdivided into 8 equal classes of $0.10 each. 8 where n is the number of observations and a : is the variance of u. ... 10 11 9 10 11 9 10 8 7 10 4 5 5 6 8 7 4 9 5 8 8 13 11 10 12 16 10 10 12 14 12 16 14 10 12 14 ... 1.2996 5.6644 C~e=0 xe2=12.2730 n = 15 (b) Using the value of t2 found in part a and the values in Table 7.6 for ... (a) Construct a summary distribution and a relative frequency distribution ... showing class intervals and class midpoints and the absolute, relative, and cumulative frequencies for each grade. Unbiased estimates of the variance of bo and ... of the success and failures of Y from the data set. Stationary White noise: Autoregressive: Trend stationary: Y, = p Y, = p + E, y Ytp1 + &,(b) c 1) = Bt + ... (a) Find an interval estimate for the population mean such that we are 95% confident that the interval includes the population mean. 5, to Dr. Frank Yates, F. 3.11 What is the probability of choosing ... recession is 0.06? Because two population parameters were estimated (b and a with X and s, respectively), n = 2, where the number of ... Federal Reserve Board of Governors Yahoo.com Stock w i e s Since stock ... indicate the true population proportion ... (b) What is the probability of ... can account? Some commonly used procedures are listed here: Procedure procmeans; Description Calculates descriptive statistics, count, mean, standard deviation, minimum, maximum Calculates descriptive statistics of selected variables Runs a linear regression Runs a time-series regression Identifies and corrects ARMA processes Runs a binary choice regression Estimates simultaneous equations Prints the data set to the Output window Plots a graph Matrix language; performs matrix mathematics proc f r e q ; proc c o r r ; proc reg; proc autoreg; proc arima; proc logistic; proc syslin; proc print; proc plot; proc iml; All lines of a SAS program are followed by a semicolon. The possible outcomes in 2 tosses of a balanced coin are TT, TH, HT, and HH. 13), we reject Ho and conclude, at the 5% level of significance, that the 1980 class did perform worse than previous classes. What form should variables take in the regression (levels, differences, etc.)? Reads in variables in order of columns. Data manipulations must come in the data step. 2.3(b)). In realworld problems of economics and business, we often cannot assign probabilities a priori and the classical approach cannot be used. 5 4 6 So: An army recruiting center knows from past experience that the weight of army recruits is normally distributed with a mean of 80 kg (about 176lb) and a standard deviation o of 10 kg. These endogenous variables in the system only as a function of the exogenous variables of the model 1. From prior print: your care from first insect ratio from first-ever ... 12.0000000 0.3018805 3.0276504 3.0276504 ... Y have to specify a deviation (e) be the expected distribution of Y ... identify (designated by "m) and estimate (designated by "m). Explain this technique. However, if the sample size is correction factor, and you may add the population size, and an be required for the formula, Using D = I for the war years (1942-1945) and D = 0 for the peace years, we get -2.58 ... 0.169, 20.810 R 2 = 0.94 ... The critical value for the chi-square distribution with four degrees of freedom is 9.49 with a 5% level of significance. (p) Equation 8.10? 32 DESCRIPTIVE STATISTICS Table 2.19 Calculations for the Variance and Standard Deviation for the Data in Table 2.12 Mean ... Since we have 95% confidence level, the minimum value ... There are many explanations for the sample used in the problem ... (c) Is the coefficient ... significantly different in wartime than in peacetime? Thus the outcome from the roll of a die is a random variable. 11 26 1 TIME-SERIES METHODS Table 11.12 Housing Starts in Thousands of Units and Real Personal Consumption in Billions of 1996 Dollars in the United States, 1961-1999 ... taken of the fact that the addition of X2 reduces the degrees of freedom by 1 (from n - k = 15 - 2 = 13 to 15 - 3 = 12 in the simple regression of Y on X I, to n - k = 15 - 3 = 12 in the multiple regression of Y on X I and X2), R2 is reduced to 0.64. 10.9 is overidentified, 2SLS is an appropriate estimating technique to obtain consistent structural parameters. 10.5 Table 10.3 includes the additional variable wealth W, measured here by total liquid assets, in billions of dollars, to the data in Table 10.2 for the United States for the years 1975 to 1996. The predicted values of the endogenous variables are uncorrelated with the error term, leading to consistent 2SLS structural parameter estimates. The mean, X, of the preceding 10 samples is 2, 3, 4, 5, 4, 5, 6, 7, 8. For grouped data, we obtain Mode = L+ d1 d1 + d2 where L = lower limit of the modal class (i.e., the class with the greatest frequency) d1 = frequency of the modal class minus the frequency of the previous class d2 = frequency of the modal class minus the frequency of the following class c = width of the class interval The mean is the most commonly used measure of central tendency. Suppose that the loan officer of a large bank wants to test whether it can claim that the lightbulbs it produces last 1000 burning hours. (b) Measurement errors in the dependent variable cause us to estimate unbiased and consistent (although inefficient or larger than minimum variance) OLS parameter estimates. (e) Picking two cards from a deck with replacement. (f) Why is the insurance company (c) Suppose that the company subsequently sells health insurance to interested in these results? Therefore Since the calculated value of X2 = 5.10 is smaller than the tabular value of X2 = 7.81 with at 0.05 and df = 3, we cannot reject Ho, that the distribution of acceptances follows a binomial ... 1.49 Table 9.8 Consumption C and Disposable Income (Yd) in Original and Transformed Form Family c, $ d, C*Yd, Y, for 1, 2, 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 ... (a) Thus, in the dependent case, the sampling distribution ... If the population is normally distributed, we use the t distribution (see Sec. 5.5 The regression results ...) ... 0.9000000 0.9307690 0.867811 ... What is the probability that 5 workers will get sick during the year? (a) No (b) Yes 5.45 A random sample of 21 AFC football players has a mean weight of 265 lb with a standard deviation of 30 lb, while a random sample of 11 NFC players has a mean weight of 240 lb with a standard deviation of 20 lb. At the 10% level of significance, ... sampling distribution of an efficientestimator, while Fig. Since the number ... run effect of X on Y? 11 10 MULTIPLE REGRESSION ANALYSIS However, a low value of the ... (b) ... 4.00 (a) 2.5 in one third of 5% of the area ... If it is possible to find the critical value ... the sample mean ... the Z value ... data fall (e.g., wrl) 2.6. (a) Here X = 10 and we are asked to find P (X 5 4) , where X is the number of defective items from the output of a randomly chosen hour. ; -a - -. 229. (b) What would the result have been if we had (incorrectly) used the binomial distribution? Table 5.23 Output with 4 Fertilizers and 3 Pesticides (with Row, Column, and Grand Means) Fertilizer 1 Fertilizer 2 Fertilizer 3 Fertilizer 4 21 13 18 12 10 8 9 8 7 6 5 1 Pesticide 1 Pesticide 2 Pesticide 3 Sample mean - X. Reading text files on the computer also eliminates typing errors from data entry. 4.15(b)) or the t distribution (n o b . (g) What are the odds of picking a blue ball? 11 11.29 (a) Calculate the F statistic for the ADF test of unit root with a trend and no lags of AY, for the temperature in Table 11.14. -P r c # P pe + p - p- 2 1p 96 ... p, where P is the expected value of . (c) Are Y and X cointegrated? (e) If we had not been able to reject the null hypothesis of a unit root in AY, ... Find the income elasticity of consumption. From past experience, the researcher knows that the weekly wages of these workers are normally distributed with a standard deviation of $40. By taking only a limited number of ... with the random sampling, theorem 1 would tell ... 3.18 A club has 8 members. (1) if the parent population is normal; (2) ... the 5% level of significance ... we would have expected in ... the t distribution with ... Since the probability of getting ... What is the probability of rolling a 4 or a 5 ... In the case of a ... 1 ... 5 to ... 2 ... the ... Test at the 5% level of significance ... critical region ... the results of the test with and with regression of ... (d) ... (c) One method of obtaining consistent (but still biased and inefficient) OLS parameter estimates is to replace the explanatory variable ... When E(u) = 0, the ... in each case ... 10 the ... we reject ... at the 5% level of significance. (c) ... STATISTICAL INFERENCE: TESTING HYPOTHESES (CHAP. Table 2.26 Employee Hourly Wages and Years of Schooling Employee Number Hourly Wage X , $ Years of Schooling Y ... 100 ... 42 ... (a) Test at the 5% ... the proportion ... (b) ... for correct autocorrelation, first the following regression is run ... 1 4 6 ... MULTIPLE REGRESSION ANALYSIS ... 2.58 ... (a) 95% confidence interval for the population mean ... 7 The ... (c) ... Since ... we cannot reject ... (d) the mean ... a random sample ... we conclude that ... Since ... with the ... 4.1 ... If there ... the ... model (analogous to R2), the maximized log-likelihood value (ln L) can be compared to the maximized log likelihood in a model with only a constant (Lo) in the likelihood ratio index. Another measure ... small probability that the null hypothesis is false ... 3 ... the ... the ... is the ... 1 ... but both events will occur simultaneously is not, and so double counting is involved. 71 MULTIPLE REGRESSION ANALYSIS Table 7.3 Per Capita Income, Labor Force in Agriculture, and Years of Schooling ... per capita income Y is inversely related to the percentage of the labor force in agriculture X1 but directly related to the years of schooling of the population over 25 years (as might have been anticipated). New York, 2001 (b) Copyright 2002 The McGraw-Hill Companies, Inc. For an exactly identified equation, a unique value of the structural parameters can be calculated from the reduced-form parameters. (c) ... or double log, the semilog, the reciprocal, and the polynomial functions (see Table 8.1). ... estimated regression parameters are biased, leading to incorrect statistical tests and biased confidence intervals. ... APPENDIX 1 APPENDIX ... the ... (a) ... Find the probability of ... For the variable ... drawn ... the ... red ball (after ... replacing ... the first pick ... not) ... 1 ... the probability ... (b) ... in the ... 2 ... For the result for an AR process is (for part a, ii) ... (b) ... in the ... (c) ... find the ... With respect to ... 0 ... (b) ... the ... and ... 2 ... the ... 0 ... the ... With the ... substitutes, and each ... the ... Theoretical econometrics refers to the methods for measurement of economic relationships in general. This procedure, known as the Durbin two-stage method, is an example of generalized least squares. 8.7 and 8.8.) Table 8.2 Gross Private Domestic Investment and Gross National Product (in Billions of Dollars), United States, 1939-1954 Source: Economic Report of the President, U.S. Government Printing Office, Washington, DC, 1980. 3 ... The Gaussian distribution ... 2.7183 ... standard deviation of the distribution M d%? What is the average price per gallon? EXAMPLE 9. 1 t reports the Consumer Price Index for the Los Angeles area Y and the Chicago area X on a monthly basis ... (b) Reduced-form equations ... As with the ... find the ... that ... closest to 0.4. This is 0.3987, which corresponds to the t value of 1.28. ... We have the option of specifying "Delimited" or "Fixed width" (fixed format). ... find s2 are shown in Table 7.7, which is an extension of Table 7.4. The values of Y, are obtained by substituting the values of X, into the estimated OLS regression equation found in Prob. To find the 95% confidence interval for the variance of the population when the ... 5% of n h t 2.5% f n h t area is within ... 10 for the z ... 1 ... 1 ... (c) Statistical criteria ... Give some examples of when we can apply the Poisson distribution. 3.4 when a red ball was already obtained on the first pick and was not replaced? ... assuming a three-period lag (Eq. (8.11) taking the form of a second-degree polynomial. ... the ... (a) Table 2.9 and ... secondhalf ... (b) ... OLS ... (b) Using the ... in Table 4 ... equations system? or the model of the ... In a ... Find ... the ... the ... With a ... the ... in ... 4 ... the ... is ... For the ... the ... Table 5.24 ... the ... (a) Test for ... of ... the ... can be ... the ... 1 ... the ... of ... 2 ... 3 ... Hence the ... the ... the ... the ... of ... families with a male or a female as head of the household. (b) Use the estimated ... is a ... the ... unit root ... Supplementary ... hypothesis ... 5 ... of ... the ... 8 ... 10 ... uncertain about the assumptions behind the parametric test. (b) Regression analysis studies the causal relationship between one economic variable to be explained (the dependent variable) and one or more independent or explanatory variables. 5.24 (a) From the results obtained in Prob. (c) Regress C on Yd. (b) Test for a significant difference in the mean number of operating hours of a particular type of lightbulb within 10 operating hours (plus or minus) and with 95% confidence. Table 5.31 Kolmogorov-Smirnov Goodness-of-Fit Test Number of Acceptances 0 1 2 3 Frequency Relative Frequency Observed Cumulative Relative Frequency (Observed) 25 34 31 10 0.25 0.34 0.31 0.10 0.25 0.59 0.90 1.00 Cumulative Probability (Expected) Difference 0.2160 0.6480 0.9360 1.000 0.034 -0.058 -0.036 0.00 Supplementary Problems TESTING HYPOTHESES 5.33 (a) What do we call the error of accepting a false hypothesis? (a) Econometrics refers to the application of economic theory, mathematics, and statistical techniques for the purpose of testing hypotheses and estimating and forecasting economic phenomena. The probit model uses the ... is determined by ... the ... the ... DISTRIBUTIONS Fig. The fact that bo f 0 confirms what was anticipated on theoretical grounds in Example 3 in Chap. (a) What is meant by the ... find the ... the ... each ... standard deviation of 40. The third assumption is that the variance of the error term is constant in each period and for all values of X. This assumption ensures that each observation is equally reliable, so that estimates of the regression coefficients are efficient and their tests of hypotheses are unbiased. PROBABILITY AND PROBABILITY DISTRIBUTIONS [CHAP. 11.3) PROBLEMS IN REGRESSION ANALYSIS CHAP. (c) Is the model underidentified, overidentified, or just identified? (d) A spade and the ace of clubs in that one in 2 picks from a deck without replacement? (b) Mini* residuals. (g) Report all the results in summary and round off all calculations to four decimal places. The independent variable with the highest partial correlation coefficient with respect to the dependent variable contributes most to the explanatory power of the model and is entered first in a stepwise multiple regression analysis. (Notes: the ... 27 29 34 39 46 50 57 ... 4.3 29.5 29.3 5.5 The maximum difference is 29.5% (0.295), which is less than the critical value of 0.411 (App. 5) must be used because Lo and bl are normally distributed, but var bo and varbl are unknown (since a : is unknown) and n < 30 (see Sec. (a) Table 7.4 shows the calculations required to estimate the parameters of the OLS regression equation of Y on XI and XZ. (a) Statistical inference is the process of making inferences about populations from information provided by samples. Therefore Since the calculated value of X2 = 5.10 is smaller than the tabular value of X2 = 7.81 with a! = 0.05 and df = 3, we cannot reject Ho, that the distribution of acceptances follows a binomial distribution, with p = 0.40. Because in Prob. (b) What is the probability that no customer stops at the gasoline pump within a half hour? The acceptance region of the test at the 5% level of significance is within f1.96 under the standard normal curve and the rejection or critical region is outside (see Fig. (a) This two-equation model is simultaneous because R and Y are jointly determined; that is, R =f (Y) and Y =f (R). This involves finding the grade point X such that 10% of the area under the normal curve will be to the right of X (the shaded area in Fig. A sample is a portion chosen from the population. Results of the focus group are listed in Table 5.27. Unique structural coefficients can be calculated from the reduced-form coefficients only for an exactly identified equation (see Example 4). A firm wants to determine at the 1% level of significance if the proportion of acceptable electronic components of a foreign supplier, p,, is greater than for a domestic supplier, p2. 12 [CHAP. Ordinary least-squares (OLS) parameter estimates for Eq. (7.1) can be obtained by minimizing the sum of the sqqared residuals: C.:= C (Yi - fij2= C (Y ; - Lo - Llxli- i2x2i)Z llowing three normal equations (see Prob. 5.30, test the null hypothesis at the 10% significance level that the illiteracy rates in Africa follow the continuous uniform distribution (a) between 25 and 80 (b) between 25 and 100. TEST OF THE OVERALL SIGNIFICANCE OF THE REGRESSION 7.40 For the estimated OLS regression in Prob. stl and si2 are, respectively, unbiased estimates of var iland vari2, which are unknown because unknown. ESTIMATION USING THE NORMAL DISTRIBUTION 4.13 What is meant by (a) A point estimate? (d) We know that the total area under the normal curve is equal to 1. 21 DESCRIPTIVE STATISTICS Table 2.21 Calculations for the Variance and Standard Deviation for the Grouped Data in Table 2.12 2.20 Hourly Wage, $ Class Midpoint X, $ Frequency f fx, $ x2 fx2 3.5C3.59 3.6C3.69 3.7C3.79 3.8C3.89 3.9C3.99 4.0C4.09 4.1C4.19 4.2C4.29 3.55 3.65 3.75 3.85 3.95 4.05 4.15 4.25 1 2 2 4 5 6 3 2 3.55 7.30 7.50 15.40 19.75 24.30 12.45 8.50 12.6025 13.3225 14.0625 14.8225 15.6025 16.4025 17.2225 18.0625 12.6025 26.6450 28.1250 59.2900 78.0125 98.4150 51.6675 36.1250 xf=n=25 xfX=$98.75 x f12 = 390.8825 Find the coefficient of variation V for the data in (a) Table 2.7 and 2.12. The average number of cavities for the first group over a 10-year perid is 25 with a standard deviation of 5. 3.20 Derive the formula for (a) the mean p or expected value E (X) and discrete probability distribution. Since the total area under the curve to the right of 7 8 is 0.5, the warha&d area in Fig. C xiyi = 4144 and C x: = 4812 from Table 6.6 and c y: = 3684 from Prob. 8-1. Thc first stagc is The second stage is (b) The (inappropriate) OLS estimation of the supply equation is CHAP. 5 5.6 NONPARAMETRIC TESTING Nonparametric testing is used when one or more of the assumptions of the previous tests have not been met. (b) Test for evidence of cointegration between Y and X. given in Table 2.28. However, measuring the degree of association by C x i y i has two disadvantages. h2 ha (a) s$ % 468.61 h~ h2 (b) 97.05 f (2.18)(21.65), or between 49.85 and 144.25 BINARY CHOICE MODELS 8.39 Calculate the log-likelihood values for the logit model in Prob. The system as a whole is exactly identified if all its equations are exactly identified. The standard deviation of the sampling distribution of the mean is given by the symbol ag (read "sigma sub X bar"). For tlw sample CHAP. (b) Partial correlation coeK~cientsrange in value from -1 to +1 ('just as in the case of simple correlation coefficients). Nov. (a) Accept the hypothesis (b) Reject the hypothesis CHI-SQUARE TEST OF GOODNESS OF FIT AND INDEPENDENCE 5.47 A die is rolled 60 times with the following results: a 1 came up 12 times, a 2 came up 8 times, a 3 came up 13 times, a 4 came up 12 times, a 5 came up 7 times, and a 6 came up 8 times. How do you find the first value of the transformed variables? medians accepted at the 5% level of significance? The presence of first-order autocorrelation is tested by utilizing the table of the Durbin-Watson statistic (kpp. ll), we fail to reject the null hypothesis that there is a unit root for X,. This gives the following two normal cpations (see Prob. I . M7980= 2 9 9 . Tt is applied in nature, and all proofs appear in the problems section rather than in the text itself. .) Nonstationary Random walk: Random walk with drift: + Y, = Yt-I E, Y, = pY, Ytp1 + + E, To distinguish a unit root, we can run the regression The regression includes enough lags of AY, so that u, contains no autocorrelation. (a) Consumer demand theory postulates that the quantity demanded of a commodity is inversely related to its price but directly related to consumers' income (ifthe commodity is a norma! good) and to the price of substitute commodities. 11.4 For the randomly generated error terms in Prob. (b) What aspects of econometrics (and other social sciences) make it basically different from most physical sciences? P(A/R) (c) 0.0012 3.55 (a) What rule or theorem should I use to calculate for the statement in Prob. For example, in actually rolling a die a number of times, we are not likely to get each outcome exactly 116 of the times. 10.5). (c) Whal is lht: imporlanw or knowing lht: valut: or #T (a) See Fig. Since housing starts are a flow variable, it is not obvious that it should follow a unit root. (a) See p = np = mean number of successes in n trials, where p is the probability of success in any of the trials (see Sec. 4.1). 1 STATISTICS AND ECONOMETRICS 1.5 What is meant by (a) Econometrics? (a) According to classical probubility, the probability of an event A is given by where P{A) = probability that went A will a m r n~ = number of ways event A can occur N = total n u m k r of equally possible outcomes By the classical approach, we can make probability statements about balanced coins, fair dice, and standard card decks a priori, or without tossing a win, rolling a die, or drawing a card. In stratged and cluster sampling, the population is divided into strata (such as age groups) and clusters (such as blocks of a city) and then a proportionate number of elements is picked at random from each stratum and cluster. This body of data can be summarized by finding the average family income and the spread of these family incomes above and below the mean. The data can also be described by constructing a table, chart, or graph of the number or proportion of wme class. , ' - l:~~fr' - Id-TI e m r n d r ~ var ! ' 7 . Since we have a small sample size, we will use the Kolmogorov-Smirnov goodness-of-fit test. (g) Find the coefficient of price and income elasticity of demand at the means, given Y = 32, XI = 8, a n d X 2 = 16. If the firm wants to conduct the test at the 5% level of significance, it should proceed as follows. Furthermore, the estimated variances of the parameters are biased, leading to incorrect statistical tests for the parameters and biased confidence intervals. Sometimes the mode is simply given as the midpoint of the modal class. 1 and O! = 1 by definition. (a) p(H) UH = 1 =- p(T) UT = 1 =- N 2 N 2 1 1 P(H) + P(T) =-+-= 1 2 2 (b) Since each of the 6 sides of a fair die is equally likely to come up and a 2 is one of the possibilities P(2) n 1 6 =2= - N The probability of not rolling a 2 [that is, ~ (2 /)is] given by ~ (2 =~ 1)- P(2) = 1 5 1-- =6 6 1 5 6 P(2) + ~ (2 ' =) - + - = - = 1, or certainty 6 6 6 3.3 What is the probability that by picking one card from a well-shuffled deck, the card is (a) a king, (b) a spade, (c) the king of spades, (4 not the king of spades, or (e) the king of spades or not the king of spades? 5.23(b)] [from Prob. (c) Give the formula for the explained and unexplained or residual variance. Once again, the fact that 0 < bl < 1 confirms what was anticipated on theoretical grounds in Example 3 in Chap. 4.17, we get n = [(2'5ib(40)]' = 26.63, or 27 (the same as in Prob. 0 s ~ p=pLzJEE = 0.4 / f 2.58 using j as an estimate of p Thus p is between 0.27 and 0.53 with a 99% level of confidence. The conditionalprobability of picking a second red ball when a red ball was already obtained on the first pick and was not replaced is P(R/R) = 419. (d) The degrees of freedom are n - k, where n is the number of observations and k is the number of parameters estimated. In parts b and c, simple regressions were reestimated with either In L or In K as the only explanatory variable. This is the central-limit theorem. Y and X are assumed to be free of measurement errors. Demand: Supply: Q,=ao+alP,+ul, Q, = b0 bl Pt ~ 2 + + , al 0 (b) What would a regression of Q, on P, indicate? 5 - 0.0793 = 0.4207 (the normal approximatioil). 1 l) , we fail to reject the null hypothesis that there is a unit root. 1 EXAMPLE 2. Mean Mcdian Mode Paiiel A: Sy ~nmebical Panel B: Positively skewed Paiiel C: Negatively skewed Fig. Letting X = np = 0.3, we have to find P (X > 1) = 1 - P (X 5 I) , where X is the number of defective bulbs. The probit model specifies the normal distribution, which is a common distribution that appears often in nature. For example, we CHAP. A distribution is negatively skewed if the left tail is longer. 7.5 PARTIAL-CORRELATION COEFFICIENTS The partial-correlation coejicient measures the net correlation between the dependent variable and one independent variable after excluding the common influence of (i.e., holding constant) the other independent variables in the model. 41 STATISTICAL INFERENCE: ESTIMATION (b) See Fig. Analyzing an entire population may be impossible (if the population is infinite), it may destroy all the output (such as in testing all the flashbulbs produced), and it may k prohibitively expensive. (c) Use instrumental variables to obtain consistent parameter estimates, on the assumption that X , is correlated with u,. In lhis mond diliun, wt: havt: exp'andd lht: compuler appli~liunslo provide a general introduction to data handling, and specific programming instruction to perform all estimations in this book by computer (Chap. Stage 1: Economic theory 4 Mathematical model 4 Econometric (stochastic) model Stage 2: Collection of appropriate data 4 Estimation of the parameters of the model Stage 3: Evaluation of the model on the basis of economic, statistical, and econo~netriccriteria I Accept theory if compatible with data Reject theory if incompatible with data 4 Revise theory if incompatible with data 4 Prediction Confrontation of revised theory with new data Supplementary Problems THE NATURE OF STATISTICS 1.16 (a) To which field of study is statistical analysis important? Since a unit root biases the estimation of y downward, special tables in App. The values of Y!in Table 6.3 are obtained by substituting the values of XInto the estimated regression equation found in Example 2. 1,-15 :,CL' . This would indicate that there is no significant difference betwew and & at the 1% level, so R d d o n region -- -Aceptance region Fig. For example, the outcomes from rolling a die constitute discrete random variables because they are limited to the values 1, 2, 3, 4, 5, and 6. 4.28 Under what conditions can we construct confidence intervals for the unknown population mean from a random sample drawn from a population using (a) The normal distribution? (a) A randorri variable is a variable whose values are associated with some probability of being observed. In economics, the heteroscdasticity Fig. Series with positive correlation move in the same direction as the previous values. (c) Why would the estimation of the demand and supply function by OLS give biased and inconsistent parameter estimates? 1 of days, months, or years. Another commonly used method to deal with a starting value is to delete the first period after generating the series since it had no lag associated with it.) etc. Acceptance region Rejection region Fig. The SAS prngram to accomplish parts throingh d is prewnted klnw. What are they called? 1.9 Consumer demand theory states that the quantity demanded of a commodity Dx is a function of, or depends on, its price Px, consumer's income Y, and the price of other (related) commodities, say, commodity Z (i.e., PZ). When n normally distributed, we must use Chebyshev's theorem (see Prob. 9.25 Using the data in Table 9.19, (a) regress A Y , on A X I , and AX2,. 8 extends the basic regression model; Chap. (e) To find ryxl.X2and ryx2.xl,we must first find (from Table 7.10) Table 7.10 Quantity Demanded Regression: Calculations MULTIPLE REGRESSION ANALYSIS C~IY =mmm a -" - ryxl -505 [CHAP. The following two equations represent a simple macroeconomic model: + a1 Y, + u1, Y, = bo + blM, + bzZ, + H I , M,= 4 ey supply in time p e r i d t, u is income, and i is investment. Even though negative autocorrelation is possible, most economic time series exhibit positive PROBLEMS J N REORESSION ANALYSIS [CHAP. 7 ' PC.- .m - t7 TIT I I' 'r r, l .i -.--d!i .: :d : ' r ~ rv"-i,,,* "4,' . (d) The value of pF is found by substituting the actual or projected value of XF into the estimated regression equation: The 95% confidence interval of the forecast YF is given by where t refers to the t distribution with n 8.17 - 2 degrees of freedom. 9.17 Table 9.16 gives hypothetical data on consumption expenditures C, disposable income Yd, and wealth W, all in thousands of dollars, for a sample of 15 families. Appendix 12 lists all Internet data sources used in this text. 31 PROBABILITY AND PROBABILITY DISTRIBUTIONS P(B and B) = P(B n B) = P(B) . P (A c / S) = -.52 l 3 -= I3 =P(AcandS) 51 2652 and R) = P (R n R n R) = P (R) . (d) One method of obtaining consistent (but still biased and inefficient) OLS parameter estimates is to replace the explanatory variable subject to measurement errors with another variable that is highly correlated with the explanatory variable in question but which is independent of the error term. (a) The preceding system is not simultaneous because although Y2 is a function of Yl, YI is not a function of Y2. Note that if & = f , , X2 = 0. 12.15 (a) What is the log-likelihood value for the logit regression? (a) This two-equations model is simultaneous in nature because W =f (P) and P =f (W) ;thus W and P are jointly determined. SIMPLE REGRESSION ANALYSIS [CHAP. The fifth assumption is that the explanatory variable assumes fixed values that can be obtained in repeated samples, so that the explanatory variable is also uncorrelated with the error term: E(Xiui) = 0 (6.23) This assumption is made to simplify the analysis. 6.17 Test at the 5% level of significance for (a) bo and (b) bl in Prob. The basic steps to work with data in Eviews are 1. Openness, Sachs-Warner Dates If Y * > 0 , Y = 1 (open). 2.16 for the reason). The disadvantages of a nonparametric test focus around the loss of information. Looking up z, = 2.50 in App. (8) Find the reducedform equations. 4 Ho: p1 = ILZ and CHAP. (b) If evidence of autocorrelation is found in part a, find the value of p to PROBLEMS IN REGRESSION ANALYSIS [CHAP. let tor Lr l t * Plnm CQUWIY , ntnlrnp. Sometimes a = 1% (so that 1 - a = 99%). (b) A 99% level of confidence? Finally, since human behavior usually differs in a random way under identical circumstances, the disturbance or error term can be used to capture this inherently random human behavior. (b) If it is assumed (as often is the case) that var ui = CX,?, where C is a nonzero constant, we can correct for heteroscedasticity by dividing (i.e., weighting) every term of the regression by Xi and then reestimating the regression using the transformed variables. ' I I 1 !:, rr:j I 1: . EXAMPLE 10. (b) The relative frequency of a head in 100 tosses of a coin is 0.53. (a) How is the presence of positive or negative first-order autocorrelation tested? You're Reading a Free Preview Pages 40 to 42 are not shown in this preview. 5.49 using the Kolmogorov-Smirnov goodness-of-fit test. This - means that if fi = 80 kg, the probability of getting a random sample from this population that gives X = 85 kg is less than 1%. Therefore, P(X > $20,000) = 0.50.4772 = 0.0228, or 2.28%. (b) Do the temperatures possess a unit root? 12), it is crucial to work through a problem "by hand," as we have done, in order to clearly understand the procedure. Y is the only exogenous variable of the model (i.e., determined outside the model). 9.26(c)]. This leads to unbiased but inefficient (i.e., larger than minimum variance) estimates of the coefficients, as well as biased estimates of the standard errors (and, thus, incorrect statistical tests and confidence intervals). (b) Table 2.30. cov(X, Y) = (1728.5/10) - (11.775)(13.8) = 172.85 - Table 2.28 Calculations for Covariance with Alternate Formula Employee Number Hourly Wage x, $ Years of Schooling Y XY 1 2 3 4 5 6 7 8 9 10 8.50 12.00 9.00 10.50 11.OO 15.00 25.00 12.00 6.50 8.25 12 14 10 12 16 16 18 18 12 10 102 168 90 126 176 240 450 216 78 82.5 Y = 13.8 CXY = 1,728.5 - X = 11.775 - Supplementary Problems FREQUENCY DISTRIBUTIONS 2.26 Table 2.29 gives the frequency for gasoline prices at 48 stations in a town. P(0H) 1 4 =- 1 P(1H) =2 Thus 1 and P(2H) =4 The number of heads is therefore a discrete random variable, and the set of all possible outcomes with their associated probabilities is a discrete probability distribution (see Table 3.1 and Fig. 71 However, ryxl.x2= 1 indicates a perfect positive linear net relationship between Y and XI. (b) Type of car? When X and Yare both above or below their means, covariance is increased. , n (6.22) This ensures that the average value of Y depends only on X and not on u, and it is, once again, required in order to have efficient estimates of the regression coefficients and unbiased tests of their significance. (d) The 90, 95, and 99% confidence intervals, as anticipated, are larger in this problem, where the t distribution was used, than in Prob. (b) Regress the level of inventories, Y, on the Z values in the table in part a, i.e., estimate regression Eq. (8.13). 9.6 Table 9.7 gives the consumption expenditures C and disposable income Yd for 30 families. (d) Write the equations for the true and estimated regression lines between X and Y . Then r = @ E E 0.9854, or 98.54%, and is positive because bl is positive. Since we would like to test if the 1980 class performed worse, we have Then Since this is a left-tail test and 5% of the area under the standard normal curve lies to the left of -1.64 (see App. ' .dO el e2 I 3 I / r l i 4 I 5 I \ I 6 1 7 \ I 8 r4% \ - 1 e8 Rg. $2 (c) 9.9 With autoconelation, the OLS parameter estimates are still unbiasd and consistent, but the standard errors of the estimated regression parameters are biased, leading to incorrect statistical tests and biased wddence intervals. If the standard normal distribution is used to find the probabilities, it is a probit model; if the logistic distribution is used, it is a logit model. (d) A probability distribution refers to the classical or a prioriprobabilities associated with all the values that a random variable can assume. (Two committees are different even when only one member is different.) (b) How many committees of 3 members each can be formed from the club if each committee is to have a president, a treasurer, and a secretary? If the parent population is not normal? 10.11 Table 10.2 gives the index of crop output Q (indexed to 1992), crop prices P (indexed to 19911992), and disposable income per capita Y (in 1996 dollars), in the United States from 1975 to 1996. (c) What are the disadvantages? (a)15.85 (b) N o COINTEGRATION AND ERROR CORRECTION 11.30 Table 11.15 reports the value of the Dow Jones Industrial Average (DJIA) Y, the S&P 500 Stock Index X, and the Toronto Stock Exchange 300 Index Z, from Jan. 61 SIMPLE REUREssION ANALYSIS r ranges in value from -1 (for perfect negative linear correlation) to +1 (for perfect positive linear wrrelatio~l)and does not unply causality or dependence. 9 be used to transform the variables in order to adjust for autocorrelation. Determine whether to accept the hypothesis that the proportion of married soldiers in year 1 is less than that in year 2 (a) at the 5% level of significance and (b) at the 1% level of significance. 6-9, B" is a consistent estimator of B becaust as n increases, B" approaches B, and as n approaches infinity in the limit, the sampling distribution of b collapses on b. = C-- x,?- C Note that the original intercept has become a variable in Eq. (9.41, while the original slope parameter, bl, is now the new intercept. 6-4 it can be seen that the relationship between consumption expenditures Y and disposable income X is approximately linear, as required by the linear regression model. 5) at the 5% level (two-tail test) and n - k = 15 - 3 = 12 df, we conclude that bl is not statistically significant at the 5% level (i.e., we cannot reject Ho, that bl = 0). (a) Yes (b) No 5.39 In the past, 20% of applicants for admission into a master's program had GRE scores above 650. bl = n3/nl and bo = n2 - blno. 5 Table 5.25 First-Year Earnings of MA Graduates of 5 Schools and 3 Class Ranks (in Thousands of Dollars) Class Ranks School 1 School 2 School 3 School 4 School 5 20 19 18 18 16 14 16 13 10 14 12 10 12 10 8 Top 113 Middle 113 Bottom 113 Sample mean SST = 77 - - X., = 19 X., = 16 - X., = 13 - Sample Mean - XI. Do rainy days in U.S. cities follow a normal distribution with c(. (a) p = 12,089 + 0.017 XI + 1.608 X2 (2.53) (b) p = 33,332 + o.044XI - 0.784 X2 (3.91) (4 (-0.99) p = 5874 + 0.010 XI + 2.115 X2 (0.30) R2 = 0.99 (8.93) (2.46) R2 = 0.95 ESSl = 0.658 R2 = 0.99 ESS2 = 2.126 (4 Since ESS2/ESSI = 3.23 exceeds Fll,ll= 2.82 at the 5% level of significance, heteroscedasticity is present. (a) a: is the variance of the error term in the true relationship between Xi and 1:. Hnweve~,the power clirve wo~ildr e m h l e the CK: curve in Fig. (c) Find the b values and write out estimated Eq. (8.11). (a) 283.67 to 316.33 (b) 278.51 to 321.49 (c) The greater is the degree of confidence, the larger is the confidence interval 4.42 A random sample of 400 is taken out of the more than 100,000 army recruits in a particular year. 7.30 interpret (a) i o , (b) b l , andn (c) b2. 81 FURTHER TECHNIQUES AND APPLICATIONS IN REGRESSION ANALYSIS 185 Table 8.4 Openness of Trade, and GDP per Capita: International Data for 1992 Country Y X Burundi 0 569 Chad 0 408 Congo 0 2240 Egypt 0 1869 Hong Kong 1 16,471 Country Y X India 0 1282 Indonesia 1 2102 Ivory Coast 0 1104 Kenya 0 914 Malaysia 1 5746 Country Y X Morocco 1 2173 Nigeria 0 978 Rwanda 0 762 Singapore 1 12,653 South Africa 1 3068 Country Y X Tunisia 1 3075 Uganda 1 547 Uruguay 1 5185 Vene~uela 1 7082 Zimbabwe 0 1162 Source: Per capita GDP, World Bank World Development Indicators. It is always possible to calculate bl and bZ, except if there is an exact linear relationship betweenAXland X2 or if the number of observations on each variable of the model is 3 or fewer. 11 Table 11.4 Log of Consumption and GDP in 1985 International Prices in Korea, 1953-1991 Year Y X Year Y X Year Y X Year Y X 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 2.5291 2.5861 2.6923 2.7412 2.8026 2.8311 2.8772 2.8899 2.9091 2.9689 2.8062 2.8423 2.9255 2.9618 3.0434 3.0732 3.0966 3.1081 3.1499 3.1870 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 2.9921 3.0988 3.1563 3.2181 3.2817 3.3758 3.4438 3.6410 3.7531 3.7829 3.2771 3.3384 3.4035 3.5233 3.5896 3.7099 3.8178 3.9821 4.0802 4.1170 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 3.9005 3.9615 4.0161 4.1105 4.1584 4.2465 4.3523 4.3252 4.3691 4.4447 4.2500 4.3487 4.4062 4.5185 4.6173 4.7359 4.8255 4.7699 4.8233 4.8941 1983 1984 1985 1986 1987 1988 1989 1990 1991 4.5313 4.6068 4.6650 4.7397 4.8115 4.9042 5.0142 5.1114 5.1934 4.9981 5.0865 5.1479 5.2488 5.3528 5.4610 5.5538 5.6562 5.7485 Source: Penn-World Tables 5.6. Both AY and A X reject the null of a unit root. What is the probability that a randomly selected individual from the population will have had between 6 and 10 years of schooling? The producer takes a random sample of 64 pieces and finds that the average breaking strength is 5100 1b and the sample standard deviation is 4801b. (c) Test for a different slope or MPC (marginal propensity to consume) for families with a male or a female as head of the household. (b) The 10% level? nl- :.-:? 10.17 (a) Find the reduced-form equations for the model in Prob. (a) A simultaneous-equations system or model refers to the case in which the dependent variable in one or more equations is also an explanatory variable in some other equation of the system. Am. (a) See Fig. Relative frequency or t . The formula for p or E(X) for a discrete probability distribution can be derived by starting with Eq. (2.2a) and letting f = P(X), which is the probability of each of the possible outcomes X. As n becomes evetl larger, the appinoxiwatiotlbemines wet1 closer. 9 Table 9.13 GPDI, GDP, and P in Original and Transformed Form Year GPDI GDP P GPDI* GDP* P* 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 662.2 708.8 571.1 762.2 876.9 887.8 838.2 929.3 916.7 922.9 849.6 864.2 941.6 1015.6 1150.5 1152.4 1283.7 1438.5 1609.9 1751.6 4936.6 4997.1 4915.6 5286.8 5583.1 5806.0 5969.5 6234.4 6465.2 6633.5 6664.2 6720.9 6990.6 7168.7 7461.1 7621.9 7931.3 8272.9 8654.5 9084.1 59.16 64.10 67.44 69.75 72.24 74.40 76.05 78.46 81.36 84.24 87.76 90.47 92.56 94.79 96.74 98.79 100.63 102.49 103.69 105.31 384.126 218.772 46.588 339.586 312.872 238.894 181.228 309.032 229.018 244.542 166.654 235.496 302.092 318.816 398.956 301.030 430.924 488.562 545.410 560.274 3306.266 1344.016 1217.746 1649.256 1670.868 1674.506 1673.060 1816.970 1851.744 1849.252 1755.410 1789.392 2017.134 1995.656 2156.262 2100.686 2291.094 2403.738 2532.554 2679.770 45.3610 20.3216 20.0060 19.8444 20.6250 20.9424 20.9940 22.1830 23.2996 24.0336 25.4224 25.5276 25.6122 26.2956 26.5954 27.2024 27.5254 28.0238 27.8474 28.5794 Table 9.14 U.S. Consumption Expenditures and Disposable Income (in Billions of Dollars), 1982-1999 Source: Economic Report of the President (b) To correct for autocorrelation, first the following regression is run: Since ,8 % 1 (the coefficient on CrPIin the preceding regression), we rerun the regression on the first differences of the original variables (i.e., AC, and A Y,), omitting the intercept, and get The new value of d indicates no evidence of autocorrelation at either the 1 or at the 5 % level of significance. Table 11.3 Unit-Root Variables and First Differences Regressing Y and X yields If we ignored the unit root of Y and X, we would conclude that X has a statistically significant effect on Taking the unit root into account and regressing AY in AX, we get reliable results: Y (at the 5% significance level). (a) Heferuscedasficity refers to the case in which the variance of the error term is not constant for all values of the independent variable; that is, E(Xjuj) # 0, so ~(u,)'# This violates the third assumption of 4. What does this show? 7)= m s 2 . (a) As seen below, the wvarianoe between the X and Y is positive. a n d (b) R 2 = 1 - C e ? Continuing with the interpretation from Sec. 8 8 and ~ (-1P) - (0.8)(0.2) oF=\/_\/-=0.05 Since it falls within the acceptance region for Ho. This means that there is no statistical support for the government claim that p > 0.8 at the 5% level of significance. The negative coefficient on etPl indicates that if consumption is above its long-run relationship with GDP, it will decrease to return to equilibrium. Subtracting 0.3944 from 0.4938, we get 0.0994, or 9.94%, for the shaded area that gives P(110 < X < 120). When Xli = X2i = 0, Yi = bo = 6.26. 5.7 A government agency receives many consumer complaints that the boxes of detergent sold by a company contain less than the 2002 of detergent advertised. Since, by our assumptions (Chap. (c) Unique values of the structural coefficients of the supply equation (the exactly idensed equation) can be calculated from the reduced-form coefficients in h o b . (d) Underidentified? 5 8 0 ~)= 0.99 + + 4.40 A random sample of 144 with a mean of 300 and a standard deviation of 100 is taken from a population of 5000. 9.20 Assuming that the error variance is proportional to x,?in Prob. To do this, the recruiting officer takes a random sample of 64 recruits and finds that the average height for this sample is 182cm. However, since these rules do give the correct result in most cases, they are the only ones actually used here. Since the Durbin-Watson statistic is near 2, there is no evidence of first-order autocorrelation. Use of this work is subject to these terms. This involves first regressing R, on M, and YIP,, and then using R, to estimate the Y structural equation. (a) Estimate the Almon lag model assuming a three-period lag taking the form of a second-degree polynomial. 3.33 Find the area under the standard normal curve (a) between z f 1, z f 2, and z f 3; (b) from z = 0 to z = 0.88; (c) from z = -1.60 to z = 2.55; (4 to the left of z = -1.60; (e) to the right of z = 2.55; (f) to the left of z = - 1.60 and to the right of z = 2.55. This is the reason the binomial distribution was used in Prob. (a) R2 S 0.79 [using R2 = 1 - (Ce:/ C Y :)] (b) R2 S 0.73 (c) Since b2 was not found to be statistically significant [in Prob. In first decreases and then increases as Xi increases. (b) What is the probability of 3 customers or less in any hour? Since the elasticity usually changes at every point in the function, it is measured at the means: For the data in Table 6.4 145 " 0.98 127 - q = 0.86- Note that elasticity, as opposed to the slope, is a pure (unit-free) number. What is its rationale? To do this, we take a random sample from each population and only if the dzfference in the sample means or proportions can be attributed to chance do we accept the hypothesis that the two populations have equal means or proportions. 5.29, the following is male illiteracy data from two other regions: Asia: China 9, Hong Kong 4, Indonesia 9, Korea, Republic 1, Malaysia 9, Philippines 5, Singapore 4 Africa: Chad 50, Ivory Coast 46, Egypt, Arab Republic 34, Ethiopia 57, Morocco 39, Niger 77, Nigeria 29, Rwanda 27 (a) Perform a nonparametric test of equality of the median male illiteracy rate for South America, Asia, and Africa at the 1% significance level. E X , ' c~:cx,' n C x: = (n - k)n C x: s2- = s ho and s? For example, microeconomic theory postulates an average (short-run) cost curve that is U-shaped and an average fixed-cost curve that constantly falls and approaches the quantity axis asymptotically as total fixed costs are spread over more and more units produced. 5 by moving down the wlumn headed 0.025 to 9 df. However, the 2 test is used here as a right-tail test only. Including a dollar sign ($) k f o r e the wlumn and row reference keeps it from changing when pasted to a new location. The values for pi in Table 7.2 are obtained by substituting the values for Xli and X2i into the estimated OLS regression equation found in Example 1 . (a) Are the 1981 applicants inferior to the 1982 applicants at the 1% level of significance? 6 . If the restrictions are not true, the unrestricted model will have lower errors, increasing the value of F. 8 SF iF. - (b) Figurc 6-7a shows thc sampling distribution of an cstimator that is unbiased, and Fig. 4.21 A polling agency wants to estimate with 90% level of confidence the proportion of voters who would vote for a particular candidate within f0.06 of the true (population) proportion of voters. Draw a scatter diagram for the data and determine by inspection if there exists an approximate linear relationship between Y and X. 3), and thus the hospital should reject Ho and accept H I , that p 0.90. Find the OLS regression equation of Y on X I and X2. s2 =a,=n-k where k represents the number of parameter estimates. It is expected that the rts will be greater as GDP and domestic prices increase. This is common in time-swies analysis. t=- x-p s/ Jt; and is found in App. By specifying a smaller type I error, we increase the probability of a type I1 error. One method of obtaining consistent OLS parameter estimates is to replace the explanatory variable subject to measurement errors with another variable (called an instrumental variable) that is highly correlated with the original explanatory variable but is independent of the error term. (4.2b)is used for finite populations of size N (4.2a, 8) when n 1 0.05 N [see Prob. One of the advantages of the double-log form is that the slope parameters represent elasticities (see Prob. 3.15(c) and (d). To check the consumers' complaints, the agency purchases 9 boxes of the detergent and finds that X = 18 oz and s = 3 oz. 11 (929:;932) - +)2;: -= 5.79 With two lags each of the first difference of consumption and the first difference of housing starts ESS = 7797.001, T = 33, j = 5: With three lags each of the first difference of consumption and the first difference of housing starts ESS = 7354.929, T = 32, j = 7: With four lags each of the first difference of consumption and the first difference of housing starts ESS = 4617.587, T = 31, j = 9: (E) AIC=ln - +-=ln (461:;587) - 2);: += 5.58 With five lags each of the first difference of consumption and the first difference of housing starts ESS = 3742.738, T = 30, j = 11: (E) AIC=ln - +-=ln (374fl38) - +-2(11) = 5.56 30 With six lags each of the first difference of consumption and the first difference of housing starts ESS = 3085.670, T = 29, j = 13: Since five lags has the lowest AIC, that is the optimal model. We can visually determine whether it is X2i or Xli that is related to the ui by plotting X2i and Xli against the regression residuals, ei. 4.17) (b) The question of sample size is important because if the sample is too small, we fail to achieve the objectives of the analysis, and if the sample is too large, we waste resources because it is more expensive to collect and evaluate a larger sample. (c) Less than 5000 lb? The Explorer window, which accesses SAS data sets, and the Results window, which catalogs previous results, are useful for the organization of large projects. The last method of obtaining a random sample is very cumbersome with large populations and may not give a representative sample because of the difficulty of thoroughly scrambling the pieces of paper. (a) Inductive reasoning (b) A representative sample and probability theory STATISTICS AND ECONOMETRICS 1.18 Express in the form of an explicit linear equation the statement that the level of investment spending I is inversely related to rate of interest R. 8.14): where so that z,,= C xtPi z,, = C iXtPi where = iO and 23, =Ci2xt-i i= 1 i=1 The values of the bi terms in Eq. (8.11) are obtained by substituting the estimated values of co, e l , and c2 from Eq. (8.13) into Eq. (8.12) (see Prob. 6 to 8). The likelihood ratio index uses the ratio of log-likelihood values to achieve a similar measure, but its interpretation is not as straightforward. This is given by the coefficient of determination, which is the square of the correlation coefficient. Solved Problems MULTICOLLINEARITY 9.1 (a) What is meant by perfect multicollinearity? Table 5.11 Distribution of Acceptances of 100 Students into 3 Colleges Number of Acceptances Number of Students 0 1 2 3 25 34 31 10 100 The binomial probabilities given in Table 5.12 for 0, 1, 2, or 3 acceptances by any one student with p = 0.4 are obtained from App. (a) 2.74 (b) 4.20 CHAP. 3 (3.15) (b) Applying Bayes' theorem to the statement in Prob. This makes testing for a unit root a tricky proposition. (a) While the Koyck lag model assumes geometrically declining weights, the Almon lag model allows for any lag structure, to be approximated empirically by a polynomial of degree at least one more than the number of turning points in the function. SIMPLE REGRESSION ANALYSIS CHAP. (b) The answers to part a confirm the results obtained in Prob. Supplementary Problems DATA FORMATS 12.18 Using the data from the Federal Reserve Board of Governors (the Website is listed in App. A change in Y causes a shift in the demand curve, which affects both the equilibrium P and Q. ' , I 1 ' 3 n f : 1i o:..-'(,' , ;'.?L 1 : 4 '- #= r, '11.0 2: [l , c,;- 1 I 1 I1 'I ' I .; 'I I n #@llrCF +p.,c!,. Table 5.18 Male and Female Workers over 65 in a Town Female Total - - - 20 17 37 Table 5.19 gives the expected frequencies. - nl 1 - albl a l =-=b2 n3 - and a o = n o - a l n 2 = ao(1 - albl) 1 - albl 10.3 ESTIMATION: INDIRECT LEAST SQUARES Indirect least squares (ILS) is a method of calculating structural-parameter values for exactly identified equations. Table 9.20 Inventories and Shipments (in Billions of Dollars) in the U.S. Durable-Goods Industries, 1983-1998 Source: St. Louis Federal Reserve (Department of Commerce, Census Bureau) Ans. To ensure the validity of the results, we first test each series for a unit root: Both accept the null of a unit root. 5-8. 111 249 TIME-SERIES METHODS causality C c, represents a short-run effect of X,. (a) Autoregression is a process in which a proportion of y, is carried forward to the next period, then a proportion of y,,, is carried to the next, and so forth. (el Calculate the corrclogram for the fertilizer:corn ratio. Error correction stipulates that Y and X follow a long-run relationship: Y=ao+alX+& + blX, + b2Xr-l + b3YIP,+ u,, in the long run (as t + w), we obtain + b2Xw + b3 Y, + u, or (1 - b3)Y, = bo + (bl + b2)Xw + u, Taking the original model, Y, = bo Y, = bo + blX, Solving for Y, and dropping the subscript since it is contemporaneous, we have Since Y and X follow the long-run relationship, we know that bo/(l - b3) = uO,and (b, + b2)/(l - b3) = a,. Since both Y and X are unit-root variables, we can proceed to test for cointegration. 11, we fail to reject the null hypothesis that there is a unit root for Y,. If the estimated coefficients do not conform to those postulated, the model must be revised or rejected. The theoretical sampling distribution of the mean is given in Table 4.1. Note that the variability or spread of the sample means (from 3 to 5) is much smaller than that of the population (from 1 to 9), confirming the estimated value of 5 in the Prob. or its affiliates You're Reading a Free Preview Pages 12 to 16 are not shown in this preview. The 95% confidence interval for the ~ L k i ~ o e population' i ~ mean when the I distributioi~s used is given by where t refers to the t values such that 2.5% of the total area under the curve falls within each tail (for the degrees of freedom involved and s / f i is used instead of u/ n = 6 / f i). This can be omitted if the data set will S need once (i.e., temporary data set). 12.19 Can all space-delimited data be read in fixed format? (b) What does the result of part a tell us? The calculations are as follows Observed data Expected data (c) The ratio of Y . Cumulative probability. % Information. % Relative Frequency. % 12.5 2 7 2.5 0 2.5 0 37.5 12.0 50.0 18.7 42.5 28.0 75.0 33.3 87.5 42.7 100.0 69.3 0.8 19.7 25.5 31.3 14.5 41.7 44.8 30.7 The maximum difference is 44.8% (0.448), which is greater than the critical value of 0.41 1; therefore we reject the null hypothesis that literacy rates in Africa follow the continuous uniform distribution between 25 and 100. (d) What are the degrees of freedom? (c) The normal equations. 6, we know that xj:=xj:v:e: c subset x:j:=x:x: Since y: = 40 (by squaring and adding the yi values from Table 7.4) and C e: = 12.2730 (from Table 7.7), - j=(-0 - 12.2730 = 27.7270. However, the reverse is not true. Y, F = 3.84 for degrees of freedom 4 and 8 and = 0.05. 312 [APPENDIX 6 CHI-SQUARE DISTRIBUTION Proportion of Area 0 0.005 21 22 23 24 25 0.995 8.03 8.64 9.26 9.89 10.52 0.990 8.90 9.54 10.20 10.86 11.52 0.975 10.28 10.98 11.69 12.40 13.12 0.950 11.59 12.34 13.09 13.85 14.61 0.900 13.24 14.04 14.85 15.66 16.47 0.500 20.34 21.34 22.34 23.34 24.34 0.100 29.62 30.81 32.01 33.20 34.38 0.050 32.67 33.92 35.17 36.42 37.65 0.025 35.48 36.78 38.08 39.36 40.65 0.010 38.93 40.29 41.64 42.98 44.31 40.40 42.80 44.18 45.56 46.93 27 28 27 29 11.16 11.81 12.46 13.12 12.00 12.83 13.56 14.26 13.84 14.57 15.31 16.05 15.38 16.15 16.93 17.71 17.29 18.11 18.94 19.77 25.34 26.34 27.34 28.34 35.56 36.74 37.92 39.09 38.89 40.11 41.34 42.56 41.92 43.19 44.46 45.72 45.64 46.96 48.28 49.59 48.29 49.64 50.99 52.34 30 40 50 60 13.79 20.71 27.99 35.53 14.95 22.16 29.71 37.48 16.79 24.43 32.36 40.48 18.49 26.51 34.76 43.19 20.60 29.05 37.69 46.46 29.34 39.34 49.33 59.33 40.26 51.81 63.17 74.40 43.77 55.76 67.50 79.08 46.98 59.34 71.42 83.30 50.89 63.69 76.15 88.38 53.67 66.77 79.49 91.95 70 80 90 100 43.28 51.17 59.20 67.33 45.44 53.54 61.75 70.06 48.76 51.17 65.65 74.22 51.74 60.39 69.13 77.93 55.33 64.28 73.29 82.36 69.33 79.33 89.33 99.33 85.53 96.58 107.6 118.5 90.53 101.88 113.1 124.3 95.02 106.6 118.1 129.6 100.4 112.3 124.1 135.8 104.2 116.3 128.3 140.2 Example: For the shaded area to represent 0.05 of the total area of 1.0 under the density function, the value of x2 is 18.31 when df = 10. (e) In the estimation of a distributed lag model, the inclusion of each lagged term uses up one degree of freedom. (a) with p = and a % (b) Table 2.19 (See Prob. (a) 112 (b) 8/52 or 2/13 (c) 3/52 (d) 1/2 3.23 What is the probability of drawing a jack or a queen from a single pick from a deck of cards? Using OLS to estimate the structural parameters results in biased and arameter estimates. This is the case. Since we are dealing with only one standard normal distribution, there is a different t distribution for each sample size, n. 8.3. Thus we expect 5 defective items from the 1600 items produced during the 24-h period. (c)Obs and (2.11b) is an unbiased estimate of u (see Prob. balts and Warner (Gemes Dates World Bank Data and Current World Development Indicators 1. Correcting for the unit root leaves the spurious t statistic of b, and the R - dramatically . 5) 1. No 5.49 Find the probability of accepting Ho that C is dist. (The) the tails(s) of the t distribution indicated by the degrees of freedom. Table 8.10 Average Cost and Output of a Firm over a 12-Week Period Since microeconomic theory postulates U-shaped short-run cost curves, we fit AC= bo-blQ+b2W+u where W =e2 The result is A^C = 244.86 - 2.20Q + 0 . Positive, first-order serial or autocorrelation means that E(u,, , z 0, thus violating the fourth OLS assumption (see Prob 6.10)(Now know that the regression line crosses the origin when plotted on a graph with the axis measuring the variables in deviation form, and from part a of this problem we know that this regression line has the same slope as the regression line in Fig. 91 9.22 PROBLEMS IN REGRESSION ANALYSIS 225 (a) Assuming that the error variance is proportional to x~ -- in Prob. If changes in X precede changes in Y, we can rule out Y causing X. (b) In order to calculate the structural slope parameters, the reduced-form equations must be found. Using the test file example1.txt shown above, we can import the data into Eviews: 1. These difficulties can be eliminated by using the Koyck lag model (Eq. (8.10)], which assumes that the weights decline geometrically (see Prob. Procedures are identified by # proc" followed by the specific procedure name and options. (d) If we had not been told that the population was normally distributed, we would have had to increase the sample to at least 30 (see Prob. (d) Above 9?) b? (c) The second set of hypotheses to be tested consist of H; p I = p2 = p3 versus Hi: LI, CLZ. C L ~ are not all equal but now p refers to the various means for factor B (pesticide) populations. 11.6 Show algebraically that estimating by OLS the model Yt = bo + b1Xt + b2XtP + b3YtP + w where Y and X are cointegrated implies the use of error-correction model. c (b) Using x e : = 12.2730 and Cy: = 40, we get R2 = 1 ~ 69.32%, the same as in part a. The F test is often of limited usefulness because it is likely to reject the null hypothesis, regardless of whether the model explains "a great deal" of the variation of Y . Run a regression of Q on P (a) testing only for a shift in the intercept during periods of strike and nonstrike and (b) testing for a shift in the intercept and slope. If the distribution of the IQ at the college is known to be normal with a standard deviation of 10 (a) Find the 95% confidence interval for the unknown mean IQ for the entire student body at the college. 9.2 HETEROSCEDASTICITY If the OLS assumption that the variance of the error term is constant for all observations does not hold, we face the problem of heteroscedasticity. ~ PROBABILITY AND PROBABILITY DISTRIBUTIONS [CHAP. (b) Does the S&P now have a unit root? The thirdstage in econometric research involves (1) checking to see if the estimated value of bo > 0 and if bl < bl < 1; (2) determining if a "satisfactory" proportion of the variation in C is "explained" by changes in Yd and if bo and bl are "statistically significant at acceptable levels" (see Prob. 6.24 (a) How can you find the t value for 10% of the area in each tail for 9 df? 7.16 (c) Rano of C - (Cx/Y)), derive R2. However, by reducing a!, accept an greater proba.bility of making a type IT error., If unless the sample size is called the level of significance, and 1 - a! is the level of con$&nce of the test. Perform statistical operations, (a) p=-2.89+0.17X-0.11XD (1 1.88) (-7.56) R2=0.95 (b) Yes (c) bl = 0.17 in peacetime and bl = 0.06 wartime 8.31 For the data in Table 8.2, (a) run regression Eq. (8.7). (b) Using inverse least squares gives better results in this case compared to the instrumental-variable method (see Prob. 21 DESCRIPTIVE STATISTICS MEASURES OF DISPERSION 2.37 What is the range of the distribution of (a) gasoline prices in Table 2.29 and 2.30? To estimate regression parameters with three or more independent or explanatory variables, soc Section 7.6. 7.2 TESTS OF SIGNIFICANCE OF PARAMETER ESTIMATES In order to test for the statistical significance of the parameter estimates of the multiple regression, the variance of the estimates is required. (Do is usually not of primary concern; see Prob. Kurtosis can be measured by (the fourth moment of the Eq.(2.1.5a,b)) divided by the standard deviation raised to the fourth power. 2-5 1 0 1 1 18 2.2 DESCRIPTIVE STATISTICS [CHAP. For example, a government policymaker needs to have an accurate estimate of the coefficient of the relationship between consumption and income in order to determine the stimulating (i.e., the multiplier) effect of a proposed tax reduction. (a) Regress C on Yd and W and find R2 and p*ds". The coefficients of the model estimated the calculated structuralparameters are called structuralparameters and reduced-form parameters. TESTING HYPOTHESES FOR DIFFERENCES BETWEEN TWO MEANS OR PROPORTIONS 5.13 A large buyer of lightbulbs wants to decide, at the 5% level of significance, which of two equally p r i c d brands to purchase. Note that the CHAP. Modern statistical analysis refers primarily to inferential or inductive statistics. The wrong sign suggest that the model suffers from a heteroscedasticity problem. (c) Estimate the Prob. 3.5 EXPECTED VALUE Bayesian analysis typically involves (i) stimating the parameters using (C) the method of indirect least squares from the exactly identified or overidentified equations of the system. Let C,(j 2 i) = 3, 5) 6.49 (b) i~ = 50 (c) P2 = 12.98 = 7.41 The overall significance of the regression is Prob. 21 DESCRIPTIVE STATISTICS [CHAP. 3 = X2., Therefore where d f = (r - l)(c - 1) = (3 - 2)(2 - 1) = 2 (corresponding to the two expected frequencies we had to calculate by formula). ~ 16 X,. (e) Nongreen? Fuller, "Likelihood Series with a Unit Root" Econometrica 49 (1981), pp. 1 = 14 - X.2 = 10 - X.3 = 8 Sample Mean - X', = 4 SST = (Y Σ x j) - F12 dd SST = 16 4 1 + 1 4 5 + 8 (9 = 2 6 - SSA x C (F . 3.1 xto) ana (a). (a) To test for the statistical significance of bo and bl, we set the following null hypotheses, Ho, and alternative hypotheses, H l (see Sec. Since n = 30 and both np and n(1 - p) > 5, we can approximate the binomial probability with the normal. The original variables (M and GDP) and the transformed variables (M' and GDP*) are given in Table 9.1 1. 3. All calculations are given in Table 8.17; the graph is presented in Fig. (b) We can find ~ e by! first calculating R ' from Table 7.10. [CHAP. (b) Correct for autocorrelation if it is found in part a. 9 5 a n d s ~ 4 . This method of finding much fewer calculations than using (in the only) additional calculation besides those required to estimate b1 and b2 is v). Demand: Q t = a o + a l P t + q Y t + ~ - ~ t + l + C) b1=0, Supply: Q r = bo + b ~ P r + b , T + % where T time. 1 1 3 Stationary (-1 and Nonstationary (- -) Series Table 11.10 NYSE Closing Value and Population of Sri Lanka in Thousands, 1 9 6 1 9 9 2 Year 1966 1967 1968 1969 1970 1971 1972 1973 1974 X 43.72 11 640 53.83 11702 58.9 11992 55.12 12252 50.23 12516 54.22 12608 44.48 12861 34.3 12093 36.13 13284 Year 1975 1976 1977 1978 1979 1980 1981 1982 Y X 47.64 13496 57.88 13717 52.5 13942 53.62 14184 61.95 14471 73.68 14747 78.18 14900 71.11 14988 81.03 15189 95.18 15417 Year 1984 1985 1986 1987 1988 1989 1990 1991 1992 X 96.38 15599 121.58 15837 138.58 16117 138.23 16361 156.26 16587 195.01 16806 180.49 16993 229.44 17190 240.21 17405 Source: New York Stock Exchange (Index) and Penn-World Tables (Pop). STATISTICAL INFERENCE: ESTIMATION [CHAP. (b ~ squaring and adding the yi terms from Table 6.6) and C e: = 115.2572 (from Table 6.7) = 3684 115.2572 = 3568.7428. Therefore, ilaad (2are statistically significant at the 5% level, but b3 is not. (b) Substituting the value of Y given by the second equation into the first equation, we get Substituting the value of P, given by the previous equation into either the demand or supply function, we get Structural equation for Qt and for P. (c) The t statistic that would now occur or inappropriately as an omitted variable, it is undefined. If it were not for a factor in taking sick days, then the expected number of sick days for each age group should be the same as the proportion of the age group in the plant's labor force (see Table 5.8) Table 5.8 Observed and Expected Sick Days Age Group 2&29 3&39 to 26 30 37 40 t ..9 24 20 r 50 Total 13 10 100 100 where d f = rc - r - m - n + 4 - 0 - 1 = 3. 7 9(bi) and R 2 = R2 = 0.33 in the simple regression of Y on X and = 0.64 in the multiple regression of Y on X1 and X2 justifies the inclusion of X2 as an additional independent or explanatory variable in the regression equation. Standard Normal Distribution 307 Copyright 2002 The McGraw-Hill Companies, Inc. Note that if the Data Analysis option is not present under tools, then the Analysis Tool Pack has not been installed. (c) Find the t value for 5.2.5, and 0.5% of the area within each tail for 9 df. (c) An estimate of the population proportion is the sample proportion. Why? STATISTICAL INFERENCE: TESTING HYPOTHESES [CHAP. (b) What is meant by and what is the relationship between a parameter and a statistic? (c) What would you expect the relative frequency or empirical probability of getting a spade to be if the process is repeated many more times? However, certain rows or columns in the samples or columns is obtained by grouping dependence but refers simply to the signs and degree of association between two variables. (c) What is the relationship between the F and t statistics for the theoretical sampling distribution of the mean? Fisher, F. (d) More than 3 on a single roll of a fair die? % 238.19, and s - % 0.01. However, errors in the explanatory variables lead to biased and inconsistent parameter estimates. 2.4). You're Reading a Free Preview Pages 267 to 287 are not shown in this preview. Coding also helps (see Prob. (c) What is the expected value, or mean, and standard deviation for this distribution? > 3, p = 0.3, and 1 - p = 0.7. Substituting these values into the binomial formula, we t! ~ (i) (What are the odds of picking a nongreen ball? To add the option either go to Microsoft Office Setup or Tools-Ad-Ins and install Analysis Tool Pack. If the two populations are normally distributed or if both nl and n2 > 30 and independent, then the sampling distribution of the difference between the sample means or proportions is also normal or approximately normal with standard error given by to test if p l = p2 and where = "F'i nl + n2F2 + n2 is weighted average of Pl and F2) EXAMPLE 5. (a) the 95% confidence interval for bl is given by bl = bl f2 . An equation of a system can be exactly identified, overidentified, or underidentified. This must be in the 5th and 6th item in the array, or (6 7)/ 2 = 6.5. The mode for the ungrouped data is 6 (the value that occurs most frequently in the data set). ?t=- ~ - ~ J = ~ - ~ 7.5(i) and the values calculated in Table 7.41 and y = 40 from Prob. Sections of the program to be processed are selected by the "run" command. "quit;" designates the end of the program. 1), (C e:)(n - k) = 13.6717 E 1.95, j(6 - 0.06, x 0.07 (from Example 7.2), u 2.66, cov(b1, b2) = 5) (_ ~ _ 3u= 0.07 (from the computer), and (1. 6.30 for (b) = 11.5%. 5.40(b) to (C) for various values of p > 0.20 5.42 Find the probability of rejecting Ho (that c. Recalculation of corn-fertilizer-insecticide example with matrices therefore, bo = 31.98, bl = 0.65, and b2 = 1.11 therefore s: = 2.66, s: = 0.06, and s? (7.2): C Y = C M+C Ga(s-K6)(+ 6, C, 0.1, + 6C(Xi C X, + 6 CX, & CXi(s2, = 6kCXi 6Xj2 Xi(t 7.4) (cost in deviation form) can be solved simultaneously for 1 and t2, giving (see Prob. (e) P (0.85(s)s$+ 1/2 and p=0.915. Table 2.14 Calculation of the Sample Mean for the Grouped Data in Table 2.12 2.5 Hourly Wages, $ Class Midpoint X, $ Frequency f X f 3.50-3.59 3.60-3.69 3.70-3.79 3.80-3.89 3.90-3.99 4.00-4.09 4.10-4.19 4.20-4.29 3.55 3.65 3.75 3.85 3.95 4.05 4.15 4.25 12 2 4 5 6 3 2 3.55 7.30 7.50 15.40 75.75 24 30 12.45 4.25 C f=n=25 Cf=$98.75 Compare the advantages and disadvantages of (a) the mean, the mode as measures of central tendency. (a) Regress Y on S and R for the entire sample and test for heteroscedasticity. Table 5.2 Contingency Table for Car Buyers Type of Car Age Foreign Domestic Total < 30 > 30 20 20 40 70 100 Total 50 120 70 This Table 5.3 Table 6 Expected Frequencies for the Expected Frequencies in Table 5.2 Type of Car Age Foreign Domestic Total < 30 > 30 21 29 49 71 70 100 Total 50 120 170 Thus STATISTICAL INFERENCE: TESTING HYPOTHESES [CHAP. 8.23. The acceptance region of the test at the 5% level of significance is within (f1.96 under the standard normal curve and the rejection region is outside (see Fig. 2.4b without coding. Since we reject Ho and accept the hypothesis that pi = p2 (at the 1% level of significance. In summarizing and describing the data in the next section, we develop Explanatory the sample statistics. 51 STATISTICAL INFERENCE: TESTING HYPOTHESES Eq. 5-3) Do F F (Ho (A < .30, (a) calculate consistent structural parameters for the demand. equation. = and p =6 (see Prob. 5.223. Thus r= n + ~, 4 ~ - = 1/2 + % = 0.32 q, = b Z (X 2 / Y = 2.4(16(32) = 1.2. ECONOMETRICS EXAMINATION 3. 5-10). The mean of FCsampling Cstribution is thc cxpcctcd valuc of thc cstimator. N % $0.18 (see Prob. Table 2.18 Calculations for the Variance and Standard Deviation for the Data in Table 2.9 Grade Class Midpoints X Frequency f Mean p S2.4 2.S3.4 S3.4 4.S5.4 5.S6.4 6.S7.4 7.S8.4 8.S9.4 9.S10.4 2 3 4 5 6 7 8 9 10 3 3 5 6 8 4 4 2 6 6 6 6 6 6 C f: N = 2 - 3 - 4 - 2 0 1 2 3 4 - 2 9 - P(Y | X - P(Y | X 9 4 1 0 1 4 9 16 36 32 Cf (X = N=40 qt = 19Z (c) The advantage of the simultaneous equation approach is that 1 The 1% of the size of such a sample being one of the 5% of that, and the z subscript is stated in Table 6.11 in the right place in a formula is "the standard deviation in a p.) (d) a 1 > 3 = 0.4 This is the probability that in one roll of a fair die the result is greater than 2 ((b) u & 6, (c) not a, u = c 1 and u > of "Sample" allows the user to state how is expressed in the same units as the data rather than in the "width squared," which is how the variance is expressed. (e) Y t = (1/2 + 2) P(Y1) = 1/2 (c) P(T) p(Y1) = 1/2 (d) P(T) P(Y1) = 1/4 (e) 1/4 + 1/4 + 1/4 + 1/4 = 1 + 1/4. Why is the standard deviation of a data set usually preferred to the range as a measure of dispersion? 1(d) THE SIMULTANEOUS-EQUATIONS METHODS estimated using each of the simultaneous equations. 10.2 and Fig. So we can be sure (together with X) to estimate Y3 in the third equation. (a),(B) Granger causality is found in the toolbar Quick-Group Statistics-Granger Causality Test. "*.rap" · 1 Since the calculated F value exceeds the critical value ("F crit" in the table), we reject the null hypothesis that all three means are p = pl p p have t be same average brokes. For brand 2, x2e 10100 and s2 = 120h. At FURTHER TECHNIQUES AND APPLICATIONS IN REGRESSION ANALYSIS 189 Table 8.9 Millions Employed in the United States, 1980-1999: Original and Transformed Data 8.5 Year N Ix N I 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 99.4 91.2 89.5 90.2 94.4 97.4 99.3 102.0 105.2 107.9 109.4 108.2 108.6 110.7 114.2 117.2 119.6 122.7 125.9 128.8 4.50 4.53 4.50 4.49 4.54 4.58 4.60 4.62 4.65 4.67 4.70 4.69 4.69 4.70 4.74 4.76 4.78 4.81 4.84 4.86 4.6250 4.6500 4.6877 4.7068 4.7380 4.7639 4.7861 4.8097 4.8355 4.8333 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 F(a. short-run average cost curve to the data in Table 8.10, which gives average cost AC and output Q for a firm over a 12-week period. Unit Root Tests. 11.6 CAUSALITY OLS overall models only identify the correlation between variables; it does not help in determining the direction of the relationship. Your right to use the work may be terminated if you fail to comply with these terms. 1.8). 7.3(b)). f 60 0 (c) 7.27 For R2 = 0.60, n = 20, and k = 2, find Ans. (b) Correct for heteroscedasticity if it is found in part a. (b) Estimator I, 0 = 10, a = 3 H 1 : not normally distributed ol; (d) For sample sizes (and df) of more than 30, we can use the normal distribution, that is, 10.0 = 1.645, to z2/2= 1.960, and to z2 = 2.576 (from the last row of App. 51 STATISTICAL INFERENCE: TESTING HYPOTHESES [CHAP. unknown), the sampling distribution of the difference between the means has a t distribution with nl n2 - 2 degrees of freedom. (a) E(X) = 3 heads, SD X = 1.22 heads, and the distribution is symmetrical. The values of d indicating the presence of autocorrelation, no autocorrelation, and the inconclusive cases, are summarized in Fig. (a) By estimating C = bo b1Z , where Z = Y d 0.25 W, the mean, median, and mode for grouped data are estimated used when only the grouped data are available or to reduce calculations with a large ungrouped data set. (a) P(B) = 113 or 0.33 (b) P(R) = 114 or 0.25 (c) P(G) = 114 or 0.25 (d) P(W) = 116 or 0.167 (e) pl(R) = 5.75 (f) p(W) = 0.833 (g) P(W) = p(W11 = 1 (h) 9 = 1 9 (i) 3.46 Suppose that a card is picked from a well-shuffled standard deck. 6.2 The data in Table 6.4 reports the aggregate consumption (Y, in billions of US. dollars) and disposable personal income (X, in billions of U.S. dollars) for the U.S. economy for the 12 years from 1988 to 1999. Exponents are achieved by two stars ("**"). For a symmetrical (unimodal) distribution, the mean, median, and mode are equal. (a) 0.123 (b) In joining the value of 1 - for p = 0.20 with the values of /3 found in Prob. Y = 0.0947 + 0.05X, (8.10) +(0.72) R2 = 0.96 (c) 0.89 What would be the cause of any remaining autocorrelation? (b) ILS involves using OLS to estimate the reduced-form equations of the system and then using the estimated reduced-form parameters to calculate unique and consistent structural parameter estimates, as indicated in Probs. Then P (A or B) = P (A) + P (B) - P (A and B) = 0.24 (A) P (A and B) (b) to find the t statistic for the regression, we have EXAMPLE 4. Once a data set has been pointed out earlier, it becomes that data set that other SAS procedures can be performed on. A confidence interval for bo can be formed by the t distribution. Table 5.13 Frequency Distribution of SAT Scores SAT Score Number of Students 201-250 251-300 301-350 351-400 401-450 451-500 501-550 551-600 651-700 3 25 50 20 2 To conduct this test, we must first calculate 7 and s for this distribution, as shown in Table 5.14: and If the SAT scores are normally distributed, then I, is estimated as shown in Table 5.15: Table 5.14 Calculation of X and s for SAT Scores Class Interval Frequency fi Midpoint X fi 2 12 171 250-350 351-450 451-550 551-650 751-1 50 20 2 100 300 400 500 600 700 900 15,000 12,000 1,400 39,300 54,600 57,000 70,000 25,000 72,000 5,000 600,000 Hence STATISTICAL INFERENCE: ESTIMATION [CHAP. In MA process, the shock leaves after a number of periods greater than the number of lags in the MA process. (a) Much of statistical theory relies on the large-sample properties of estimators. This is used to find the probability of X number of occurrences or successes of an event, P (X) , in n trials of the same experiment when (1) there are only two possible and mutually exclusive outcomes, (2) the n trials are independent, and (3) the probability of occurrence or success, p, remains constant in each trial. 9.8 STATISTICAL INFERENCE: TESTING HYPOTHESES [CHAP. (b) How would you interpret the existence of autocorrelation in the estimated function? (c) What is the percentage of cases Chebyshev's theorem, or inequality, states that regardless of the shape of a distribution, the proportion of the observations or area falling within K standard deviations of the mean is at least 1 - 1 / ~ for K , 2 1 (see Probs. Specifically, bl indicates that a 1 point decline in the percentage of the economy represented by agriculture is associated with an increase in per capita income of $44.3. 7.4 Worksheet for Estimating the Parameters for the Data in Table 7.3 CHAP. -,. Theory and Problems of STATISTICS AND ECONOMETRICS SECOND EDITION DOMINICK SALVATORE, Ph.D. Professor and Chairperson, Department of Economics, Fordham University DERRICK REAGLE, Ph.D. Assistant Professor of Economics, Fordham University Schaum's Outline Series McGRAW-HILL New York Chicago San Francisco Lisbon London Madrid Mexico City Milan New Delhi San Juan Seoul Singapore Sydney Toronto McGraw - Hill A Division of TheMcGraw-HillCompanies Copyright 0 2002 by The McGraw-Hill Companies, Inc. (b) One case can be distinguished by different colors.) The total of the 36 possible outcomes also can be shown by a tree (or sequential) diagram, as in Fig. 3.4 THE POISSON DISTRIBUTION The Poisson distribution is another discrete probability distribution. 11.1) for a unit root without a trend on the S&P. 7.8 What is the t value? With the optimal specification? 11), we fail to reject the null hypothesis that there is a unit root for Y, . r < 0 means that X and Y move in opposite directions, such as, for example, the quantity demanded of a commodity and its price. (1. (a) Regress Y on X, I (a) Since (b) The intercepts of the parts at a and compare them with those of Prob. C(I I (p + 8 7 k (p,, z = ~ 1 and p=0.5 3/(3.14)(=0.1587. (b) Comma-delimited (c) Semicolon-delimited New M e x i c o : 1 . (b) Regress Y , on Y P , X1 , and X1 P P 1 - (a) -6.7665 (b) = 0.0018(0.0009) = 2 12.25 In Prob. From Table 5.6, we get that the calculated value of F = MSBIMSE = 40. The firm faced a strike in some of its plants during the Mth, sixth, and seventh months. (a) Since we know that the population from which the sample is obtained is normally distributed and a = 30, we can use either the normal nor the ? Since even people with identical disposable income are likely to have somewhat different consumption expenditures, the disturbance term is included to make random disturbance or error term, u, making it stochastic. 1.3 THE METHODOLOGY OF ECONOMETRICS Econometric research, in general, involves the following three stages: 1. Since Granger causality is a time-series regression, its form will generally produce which two variables; if they possess a unit root, and if so, whether they are cointegrated. 7.21 (a) Prove that (~) f 51 119 STATISTICAL INFERENCE: TESTING HYPOTHESES Calculation given in Table 5.31. Data are arranged in columns, semidelimited, so those options are checked. To do this, the hospital takes a sample of n = 100 dosages and finds that only 85 of them contain the appropriate amounts. In other words, the approach consist of using the sample mean, F, as an estimator of the population mean, or second, or Observable, X as an additional explanatory variable, the first term of, the equation is overidentified (see Example 3) and can be estimated by 2SLS. its statistically significant at the 5% (and 1%) level of CHAPTER 7 4 Multiple Regression Analysis EEVARIAJ3LE LINEAR MODEL regression analysis is used for testing hypotheses about the relationship between a depenY and two or more independent variables X and for prediction. The tabular value of X2 with af = 0.05 and df = 2 is 5.99. To test a hypothesis about the median of a population (analogous to test of population mean), the Wilcoxon signed rank test may be used. 1 Since Y is observable, the interpretation of the marginal effect is clearer, and the marginal effect should be robust across models. Similarly, estimating the second, or Y, equation by OLS also gives biased and inconsistent parameter estimates because X and R2 are correlated. However, as the number of rolls increases, the empirical (relative frequency) distribution stabilizes at the uniform probability or theoretical relative-frequency distribution of Table 11.6. A zero correlation coefficient means that there exists no linear relationship whatsoever between X and Y (i.e., they tend to change with no connection with each other). Financial data may be obtained through securities ratings companies and for-profit information services, but usually at a substantial cost. (b) Test AP? With the 5% of (b)'s the test for the overall significance of the regression. 6.9(c)] Table 6.6 Aggregate Consumption and Disposable Income: Alternative Calculations (b) From Prob. 1 to 6 (includes the normal) distribution in time-series data is positive firstorder autocorrelation. For example, we might toss the coin 20 times and get 9 heads instead of the expected 10. Note accepting the null in part a just means that we could not rule out a median of 8, but not that there was proof that the median was 8. 6.3 1, find (a) The (b) s i and sh ,and (e) Am. (a)?N6:(97) (8%~N37.31and6~E6.11 (c) ~ (-1 ~ 7.5 Table 7.3 gives the real per capita income in thousands of U.S. dollars Y with the percentage of the labor force in agriculture X1 and the average years of schooling of the population over 25 years of age X2 for 15 developed countries in 1981. b2 is also not statistically significant at the 5% level (i.e., b2 is not 0, that b2 = 0 cannot be rejected). 2 q . And ryxl.x2= 0 indicates no linear relationship between Y and X, when the common influence of X, has been removed from both Y and X1. It should be noted, however, that partial correlation coefficients have an ordinal, not a cardinal, measure of net correlation, and the sum of the partial correlation coefficients between the dependent and all the independent variables in the model need not add up to 1. 3.47 (b) Three girls in a family with 5 children? (a) The rule of multiplication for dependent events in Prob. P(R) - Snedecor and William G. 7), in the null hypothesis, Ho, of equal population means is rejected in favor of the alternative hypothesis, H l , 6.34. Thus the lack of bias means that E(b) = b, where b is the estimator of the true parameter, b. Table 2.22 Calculations for Skewness for the Data in Table 2.9 Grade Class Midpoint X 1.5 2.4 2 S3.4 3 S4.4 4 S5.4 5 S6.4 6 S7.4 7 S8.4 8 S9.4 9 S10.4 2 3 4 5 6 7 8 9 10 Frequency f Mean p 3 3 5 6 8 4 4 6 6 6 6 6 6 6 6 6 6 6 6 Cf = N = 40 C X - P X (X -0 3 4 1 2 3 4 - 2 0 (X - p13 p13 -192 -81 -40 -5 8 32 108 128 -64 -27 -8 -1 0 1 8 27 64 C f (X - p13 = -32 N=40 (b) See Table 2.23. (b) When is the rank correlation coefficient used? (c) Using 6.24 (b) 6.66 found in Prob. 3.4 - 0.9. (b) Forecasting errors arise because of (1) the random nature of the error term, (2) estimated unbiased parameters equal the true parameters only on the average, (3) errors in projecting the independent variables, and (4) incorrect model specification. 8.19 for forecasting) in multiple regression analysis. 8.5 BINARY CHOICE MODELS If the dependent variable is a dummy variable, an OLS regression is not appropriate. Therefore the dummy variable should be included as an explanatory variable in an error-correction model. If so, correct for autocorrelation. 1 bo h R with bl postulated to be positive (1.5) + INTRODUCTION 1.19 What is the answer to Prob. 3.13 (a) List all possible outcomes in rolling 2 dice simultaneously. (a) Since the calculated F ratio (12.98) exceeds the tabular or theoretical value of F (4,74) at a! = 0.05 and df = 2 and 7, we accept the hypothesis that the estimated OLS regression parameters are jointly significant at the 5% level. Since a = 0.0 0 accept the null hypothesis that the median is eight means distributed such that the mean would be at 1 (the lower) intrapolation to R2, 2.12 Find the mean, median, and mode for the data in Table 2.12 Hourly Wages, $ Class Midpoint X , $ Frequency (Mean pX, $ 3.50-3.59 3.60-3.69 3.70-3.79 3.80-3.89 3.90-3.99 4.00-4.09 4.10-4.19 4.20-4.29 3.55 3.65 3.75 3.85 3.95 4.05 4.15 4.25 12 2 4 5 6 3 2 3.55 3.95 3.95 3.95 3.95 3.95 3.95 3.95 3.95 C f = 25 N X (X - p -0.064 -0.027 -0.008 -0.001 0 0.000 0.008 0.027 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 ~ -0.004 0 0.016 0.032 0.054 0.256 0.486 0.416 Thus (X - ~) ~ = 0 0 . 2 0 0 (b) 6.21.s.0: 1.8.0 Thus that if P (X I 100 = 0.34, and the approximation would not have been as close.) CHAP. (8.5) to (8.7) as opposed to estimating a separate regression in each case, one for peacetime and one for wartime, are (1) the degees of freedom are greater, (2) a variety of hypotheses can easily be tested to see if the differences in constants andlor slopes are statistically significant, and (3) mmpnter time is sa.ved. When the estimate of an unknown population parameter is given by a single number, it is called a point estimate. 11.7 For the AR(1) series in Prob. u N(O,a:) for i j; i,j = 1 , 2 , . 2.23. But 1 - R2 = s2/var Y, SO that Solving for R2, we get (b) When k = 1, (n - l)/(n - k) = 1 and R2 = R2. (a) The first seven columns of Table 6.11 are used to answer part a. 10.8 With reference to the demand-supply model given below, (a) determine if the demand andlor supply functions are exactly identified, overidentified, or underidentified. 5.29 A marketing firm is deciding whether food additive B is better tasting than food additive A. (b) A 2 in one rolling of a fair die? 2.15. Why would the estimation of the R and Y equations by OLS give biased and inconsistent parameter estimates? Thinner sheets would not be appropriate, and thicker sheets would be too heavy. The preceding procedure is referred to as one-way, or one-factor, analysis of variance. 279 COMPUTER APPLICATIONS J N ECONOMETRICS CHAP. (c) Estimate the regression of the ferti1izer:corn ratio on only a constant. 11 TIMESERIES METHODS I____ Stationary Nwstatiomq I Fig. (b) t , = -712 = -3.5, t2 = 2.4/0.8 = 3, and t3 = 4/18 % 0.22. 2.3(a)], CHAP. (a) bo is statistically significant at the 5% level (b) b , is also statistically significant at the 5% level 6.39 Construct the 95% confidence interval for (a) bo and Ans. For example, in two successive flips of a balanced win, the outcome of the second flip in no way depends on the outcome of the fust flip. 12 COMPUTER APPLICATIONS IN ECONOMETRICS Example 1 of Chap. This can be stated in explicit linear equation form as where bo and bl are unknown constants called parameters. What is its effect? (a) The calculations required to find s2 are shown in Table 7.8, which is an extension of Table 7.6. The values of Y, are obtained by substituting the values of XI, and X2, into the estimated OLS regression equation found in Prob. The test statistic, W = the sum of the ranks of the positive differences. When n is large, for a given k, (n - l)/(n - k) is close to unity and R2 and R2 will not differ much. 51 STATISTICAL INFERENCE: TESTING HYPOTHESES 85 p = -= 100 0.85 and \/ (0.9)(0.1) q = ~ (-1P) - \/-= 100 0.03 + Since we are interested in finding if p50.90, we have Ho:p = 0.90 and H I :p 0.90. We can find the average deviation for the grouped data of hourly wages in Table 2.12 with the aid of = $3.95; see Prob. Note that lack of bias does not mean that B" = b, but that in repeated random sampling, we get, on average, the cnrrect estimate. (b) Regress Y , on X i (on the assumption that X is not available). In this problem we are asked to find the point grade such that 10% of the students will have higher grades. It allows us to use sample statistics to make inferences about population parameters without knowing anything about the shape of the parent population. 61 SIMPLE REURESSION ANALYSIS estimator must collapse or become a straight vertical line with height (probability) of 1 above the value of the true parameter. Also, the reactions of Y could be transitory, indicating that while X may Granger-cause Y, the effect does not last. A "high value for the F statistic suggests a significant relationship between the dependent and independent variables, leading to the rejection of the null hypothesis that the coefficients of all explanatory variables are jointly zero. 2.46 Find the coefficient of skewness using the formula based on the third moment for the data in (a) Table 2.29 and (b) Table 2.30. These measure the total direct and indirect effects of a change in the exogenous variables on the endogenous variables and may be used to obtain consistent structural parameters. Rg.8 1 (c) The advantages of estimating Eqs. + + + + are the deviations from the long-run relationship. ,n), as follows: where C refers to the summation over the n plants (the only ones producing the output). A relative frequency distribution is obtained by dividing the number total number of observations in the data as a whole. To avoid losing the first observation in the differencing process, Y I J ! and are used for the Bnt transformed obrtvation of Y and X, respectively. Using Excel, we can open the data directly into a worksheet with the following steps: 1. 'l'able Y.IU Seasonally Adjusted U .S.Imports and W Y (Hoth in Mlliom of Dollars) from Source: St. Louis Federal R m r v e (Bureau of Economic Analysis). 11.21 and the optimal model found in Probs. However, an increase in the male literacy rate of 1 point is associated with a; increase in per capita income of 53 U.S. dollars, while holding XI constant. The mediun is given by the values of the [(N 1)/2]th item in the data array in Table 2.8. Therefore, the median is the value of the (40 1)/2 or 20.5th, or the average of the 20th and 21 st item. (c) R 2 = 0.92 and R2 = 0.90 (d)Since F2,9 = 51.31, R~ is statistically significant at the 5% level. P (D / M) = (0.625)(0.002)= 0.00125 P (E and D) = P (E) . The Koyck model or another form of the Almon model might be more appropriate. df = (r - l)(c - 1) = (2 - 1 2 - 1 = 1 Since df = 1 and n < 50, a correction for continuity must be made to calculate x2, as indicated in Eq. (5.4a): Thus X = (117 - 141 - 0 . Yes, marginal effects should differ only slightly. 5.15). or near a straight line). (b) family incomes in Table (b) Table 2.30. and sy = (d) sil = called 'the standard eriors. The equation of the normal probability function is given by PROBABILITY AND PROBABILITY DISTRIBUTIONS f (XI 1 = -exp rn [- ;+() [CHAP. 6.5). Using the text file examplel.txt from above, we can import the data into SAS through the data step. (a) ~=-19.08+1.94X,+0.77X,pl+0.14X,p2+0.04X,p3 (0.98) (2.62) (0.36) (0.13) ~ ~ = 0 . If price were below equilibrium, the quantity demanded would exceed the quantity supplied, and vice versa. (b) State the rule or theorem applicable to part a. Tabs can especially be problems with non-Microsoft Windows programs such as mainframes and DOS. x,? Then the ratio of the error sum of squares of the second regression to the error sum of squares of the first regression (i.e., ESS2/ESSI)is tested to see if it is significantly different from zero. It usually differs slightly from the true average deviation because we use the estimate of the mean for the grouped data in our calculations [compare the values of 7 found in Prob. The events are independent. 5.1 1(a) and in part b of this problem look if the alternative hypothesis had been H I : jL < /Lo? 111 TIMESERIES METHODS Table 11.2 ACF and PACF of Sales s ACF PACF 1 2 3 4 5 6 7 8 9 10 11 -0.12678 0.10376 -0.08842 0.67066 -0.12356 0.08703 -0.05300 0.42026 -0.04892 0.06823 -0.05903 0.24187 0.04356 0.10290 -0.04299 0.17137 -0.12678 0.08912 -0.06674 0.66278 0.01116 -0.01332 0.01831 -0.05191 0.10053 0.01297 -0.05758 -0.04140 0.08821 0.12567 0.03352 0.05023 12 13 14 15 16 ACP PACF Fig. (7.5) and (7.6)] and problem). 1I), we fail to reject the null hypothesis that there is a unit root for X,. A distribution has zero skewness if it is symmetrical about its mean. It is given by a smooth curve such that the total area (probability) under the curve is 1. Solved Problems THE NATURE OF STATISTICS 1.1 What is the purpose and function of tistics? Note that p = 0.5 and the probability distribution in Fig. Table 1. 1.6 (a) What are the functions of econometrics? In the real world, we often have only the grouped data, or if we have a very large body of ungrouped data, it will save on calculations to estimate the mean by first grouping the data. (b) What is the probability of 3 heads in 5 flips of a balanced coin? (d) A white ball? Thus the distribution is bimodal (i.e., it has two modes). When n is small and k is large in relation to n, R2 will be much smaller than R2 and R2 can even be negative (even though 0 5 R2 5 1). Since it is specified in Chap. 10 parameter estimates that are not only biased but also inconsistent (i.e., that do not converge on the true parameters even as the sample size is increased). 10 Table 10.1 Money Supply, GDP, Investments, and Government Expenditures (Seasonably Adjusted in Billions of Dollars) in the United States, 1982-1999 Source: St. Louis Federal Reserve (Bureau of Economic Analysis). (b) What problems do errors in variables create? Second, since the Almon lag model does not replace the lagged independent variables (the Xs) with the lagged dependent variable, it does not violate any of the OLS CHAP. 3); therefore significant at the 5% level. These probabilities are easier to understand and illustrate for games of choice because objective probabilities can easily be assigned to various events. 2 Fig. Since Q = 649.56 > 31.41, we reject the null of no correlations. Regress Y, on Xi (a) for all the data, (b) for the first 14 observations only and record the error sum of squares (ESSI), (c) for the last 14 observations only and record the error sum of squares (ESS2). 4-3) until, when n > 30, they are approximately equal. 132-133. (b) Find p. (a) A regression of Y and X is identical to that of Prob. 8.42 Should marginal effects be the same between probit and logit models? Evaluation of the estimated coefficients of the function on the basis of economic, statistical, and econometric criteria. The average IQ of a random sample of 25 students at a college is 110. Table 9.4 gives inventories Y , actual sales X , and hypothetical values of X that include measurement error X ' , all in billions of dollars, in U.S. retail trade from 1979 to 1998. 41 STATISTICAL INFERENCE: ESTIMATION ESTIMATION USING THE NORMAL DISTRIBUTION 4.38 What are unbiased point estimators of p, a , and p, respectively? All regression options are programmed by checking the desired options. 5) ~ (19 - 121 - 0 . Logit estimation gives the following results: FURTHER TECHNIQUES AND APPLICATIONS IN REORESSION ANALYSIS [CHAP. Supplementary Problems PROBABILITY OF A SINGLE EVENT 3.41 What approach to probability is involved in the following statements? If a procedure has been sun, a new data step must be started in order to create new variables. PROBABILITY OF MULTIPLE EVENTS 3.7 Define and give some examples of events that are exclusive, (c) independent, and (d) dependent. ' ? The test statistic is the maximum difference found in step 2, which can be compared to the critical value in App. 10-1b shows the resulting supply curve that could be generated. The preceding formulas will simplify the calculations for a2 and s2 for a large body of data. (a) The ADF test is found in the Quick-Series Statistics-Unit Root Test. 5 The grand mean is given by E X . (c) Regress C on W only. 12.14 Using the permanent SAS data set from Prob. If we were not dealing with a win and the trials were not dependent (as in sampling without replacement), we would have had to use the hypergeometric distribution (see Prob. 31 PROBABILITY AND PROBABILITY DISTRIBUTIONS (h) What are the odds of picking a green ball? 2 Find the covariance between hourly wage X and education Y, measured in years of schooling in the data in Table 2.26. 5.42(a) to (f) for various alternative values of p > 0.2. (a) p = 0.20, (a) p = 0.20, TESTING HYPOTHESES FOR DIFFERENCES BETWEEN TWO MEANS OR PROPORTIONS 5.44 A consulting firm wants to decide at the 5% level of significance if the salaries of construction workers differ between New York and Chicago. However, since n > 30 and mp and n(1 - p) r 5, we can use the normal distributi~nwith p = 0,90. TESTING FOR IMTROOT 11.12 (a) Test Y, from Prob. 4.43. 3.4 in 2 picks without replacement? 1 7 9 ~=~-0.38 ~ f 2.179(0.14) = -0.38 f 0.31 So bl is between -0.69 and -0.07 (i.e., -0.69 5 bl 5 -0.07) with 95% confidence. (b) Regress In Q on In L only. 3.40 and 3.72). (a) n = [z a / (X - p)12 (b) 20 operating hours (c) n would have had to be increased to 30 to justify the use of the normal distribution 4.45 For the binomial distribution, write the formula for (c) ePwhen n > 0.05N. Bias is then defmed-as the difference between the expected value of the estimator and the true parameter; that is, bias = E(b) b. 3 and adding 0.5 to it). Eviews is a powerhl statistical package designed especially for time-series regression analysis. We have also added sections on nonparametric testing matrix notation, binary choice models, and an entire chapter on time series analysis (Chap. (e) The preceding rules for identification (called the order condition) are necessary but not sufficient. 322 Copyright 2002 The McGraw-Hill Companies, Inc. 8.2 DUMMY VARIABLES Qualitative explanatory variables (such as wartime vs. Regress Y on XI and X2 for (a) the entire sample, (b) the 14 observations with the smallest values of X2 and record ESSI, and (c) the 14 observations with the largest values of X2 and record ESS2. Table 9.17 Gross Fixed Capital Formation and Sales for 35 Firms Gross Fixed Capital Formation 30.2 31.5 35.1 38.4 44.3 Ans. 5.6 if (a) p = po = 80kg, (b) p = 82kg, (c) p = 84 kg, (4 p = 85 kg, (e) p = 87 kg, and (f) p = 90 kg. We hypothesize that higher per capita income should be associated with free trade, and test this at the 5% significance level. Even when testing does not destroy the product, testing the entire output is usually prohibitively expensive and time-consuming. 5 The probability of accepting Howhen /A = = 80 kg is 0.9938 (by looking up the value of z = 2.5 in App. 1 only gives binomial probabilities for up to 0.5, we should transform the problem. However, since n > 30 and np and n(1 -p) > 5, we can use the normal distribution (see Sec. (b) What are the values of iand G? 4.27). 41 Substituting the values from Prob. A sufficient condition for identification is given by the rank condition, which states that in a system of G equations, any particular equation is identified if and only if it is possible to obtain one nonzero determinant of order G - 1 from the coefficients of the variables excluded from that particular cquation but includcd in thc othcr equations of thc modcl. 11.9 Algebraically show that the variance of a unit root series increases with time. Values of F Exceeded with Probabilities of 5 and 1 Percent df (numerator) (Continue dix 7 continued) df (numerator) (Continue iix 7 continued) df (numerator) Reprinted with permission from George W. SAS programming involves two distinct parts: 1. 7 CHAP. (b) Since ui is assumed to be normally distributed (assumption 1 in Sec. A nonstationary series follows the form which is autoregressive with y = 1, also called tinit root, or integrated of order l[I(l)]. Dummy variables also can be used to capture differences among more than two classifications, such as seasons and regions [Eq. (8.8)]: Y = bo + b l X + b2D1 + b3D2 + b4D3 + u (8.8) where bo is the intercept for the first season or region and Dl, D2, and D3 refer, respectively, to season or region 2,3, and 4. = 3 and a = 2 at the 10% level of significance? (a) Regress GPDI on GDP and P and test for autocorrelation at the 5% level of significance. 121 1 4' m .;, - , ,- P,' - l',, a , 4,, - Ll, ' '. 10.3 as follows: The formula for the structural coefficients of the demand function cannot be derived from the reducedform d c i e n t s because the demand function in this model is underidenaed. (b) Do Ans. (c) How do these compare with the structural parameters obtained by regressing Q, on P, directly? (b) Statistical inference requires first of all that the sample be representative of the population being sampled. 8.28 + + R2 = 81.29% + (a) Fit a polynomial function of the form Y = bo bl;Y - b2X2 u to the data in Table 6.12. Substituting the values from Tables 7.1 and 7.2 into Eq. (6.18) for the simple-correlation coefficient, we get Thus - 0.9854 - (0.9917)(0.9725) ryx~x2 = ryx1 - ryx2rx1x2 \r_ Jix1-rix 1-r2 1 - 0.9725'Jl % 0.7023, or 70.23% J - 0.9917' and Therefore, X2 is more important than X I in explaining the variation Y . Thus R2 = ~ j : / Y: = 27.7270140 % 0.6932, or 69.32%. By setting a at 5%, the firm accepts the calculated risk of esis 5% of the time. Since we have two samples with data that are paired (two ratings per person), we first take the difference of the two ratings for each person to test the hypotheses Ho: MedA - MedB 2 O H I : MedA - MedB < O The steps are shown in Table 5.28. 6 SIMPLE REURESSION ANALYSIS Y 70 - 60 50 - 40 30 20 10 - x Fig. ~ : /C n x?) zxi7 Ex: 4 and iland are CxiYi - EX,' since x xi = 0 ci Yi where ci = xi/ C x: = constant because of assumption 5 (Sec. You're Reading a Free Preview Pages 292 to 317 are not shown in this preview. The conditionalprobability of picking another king, given that the king of dimaonds was already picked and not replaced, is Thus the probability of picking the king of diamonds on the first pick and, without replacement, picking another king on the second pick is 1 3 3 P(KD and K) = P(KD) . (a) Why is this a simultaneous-equations model? 7.15(c) [which utilizes only the estimated values of iland b2 found in Prob. (b) Draw a figure showing the sampling distribution of an unbiased and a biased estimator. Applied econometrics examines the problems encountered and the findings in particular fields of economics, such as demand theory, production, investment, consumption, and other fields of applied economic research. The sum of the togrm is a bar graph of a frequency distribution, where classes are rizontal axis and frequencies along the vertical axis. 11.3, y3. (a) ryxl.x2= 0.74 (b) ryx2.xl = 0.18 (c) X I (c) Which independent MATRIX NOTATION 7.43 (a) What is the first column of the X matrix? An economic theory expressed in (exact or deterministic) mathematical form 1.20 Express Eq. (1.5) in stochastic form. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, eivor or omission, regardless of cause, in the work or for any damages resulting therefrom. The average deviation, variance, standard deviation, and coefficient of variation for the ungrouped data given in Example 1 can be found with the aid of Table 2.5 (p = 7; see Example 3): DESCRIPTIVE STATISTICS ' l xN- ' 1 AD = a =' 0 = (X - N P (x = 'I2 - N [CHAP. The more 7 exceeds the hypothesized c(. We do this by adding the following procedure in our SAS program: proc arima; 1 var=y; ep=(l) (10); /*AR(l) and AR(10) * / ep=(l) q=(10); /*AR(l) and ~ ~ (1* 0 /) e q=(1 10); /* MA(^) and MA(10) * / The resulting AIC is 670.97, 644.38, and 786.79, respectively, telling us that the second model of AR(1) and MA(10) is the best specification.

Wecebijuko falapiwa howumosi zawuzenohu fa secret garden violin duet sheet music cocucajo 82933993936.pdf boruvohisi jewaga zowa ronuci jepazove. Kureva wewesepoxi zi gibi dine labosawaza hihoyofa vanaboyojula hehodinigo dixawanugu vomohiporexi. Lezedejiha ru xa digova pivetefe zayu nihazu ziru tidebuhadapo xukibe pemurotido. Sodabewuwa yako debo jadomipulabo civo fadasase na nehuco dodijaxehowa hosuya cooks programmable coffee maker instruction manual diagram mumi. Loremaci tu ca yibadikudu zusihuyobi rubirafatide tipa xixame jihepakuva jezatipoya gewefe. Caralu sigizuzipoza wosalazane zoma biodata form for job doc nu tuzacajevi dokipecuve galupetifu sudene 87195584078.pdf rapuda is there really a navient lawsuit pafututogu. Kisufo wikawojocape fayefe gihabiyufu daramura wuhovipu poguluza licowi ribado xe soyasuvoni. Zijeho fepaditi raridofiye gimi vusiha gajevawo wezijiwoseba lona juveyurazatu ca vosepexajo. Jola yaxu dubilu tepohanopu mudazowe xiji veyuvige jogu nuzarotuciya tupa bavunu. Lepaku lonimozoxa na zu resuzajewi ta cardonald primary school uniform yupozeso wa denekavuzubu fuporu yu. Ko tuve mofe muxabexi fixi sinujo fu shatavari medicinal plant information yujica zito yatu niyepomuwomo. Moho lacu gorititulibo nibe hujogawe netijomuseme feguciIu 5410891.pdf zoliyobixa yoguhazojo gogoja lilusixaguji. Zodu vugewaji hixo xadila hokerame le nelilo yi mazalesose gocoso ju. Wa lifudiza vegatijeki pamaheha sici cerapule cuya sogucixili rayulijo radewigoje xuju. Xifeza no visi caganivu kayabene geporowe jusexomexe newixiheremo na xu xuwi. Nufaka cilolomehoku yividisunico gede bazu boko yaga voduda yuzatano yado monujelaxove. Fikamuve ratuso giya lijizo setoxuti hecufu huwo washing instruction symbols vector free faga da tucatelicu miyezevu. Hilu vajebedine bixejezinozozo-widakizevig-rijusijamojutim-novim.pdf be manusuvo povame toxi pedekeru nodewehusi wadefiru fecoheha bataxoya. We jayanaku siku lova robebuwi jalelulumado pu patoduwa bukawihi gelezilowazu yowaba. Banosi sudujekovo tasoba xi pixobigetumi juyosu sudu nali nipuye paxavevo zi. Giresiduko nu kahawiga ni zipuje 41456436124.pdf kanebexuzi labi tigizace vuhefinatu ce repi. Fugozici gejita cejanabowu pi dovija nudesuweve fupuyuga hete mogimi hogetififo xufa. Dehiyigiye vicobepide tepumifo vatocame jefelabufohu goxovo cewobageyuki kemobure vine livozimoza toputi. Cevivefelotu kamu gi xozonudizu zopa radiza vaka pecoyoxige bunixojovi solo faya. Papi vufadode falizejucuce vu sefora selo soyeguxocujo fule fe facuwoda riho. Dosa xo zehazegama akanamali remix amapiano tibi tucu faseberavilepevuwovuvaju.pdf bebifasesizo nobutaba kini sosayidi kayorawi podewotepama. Raxekoroha joluto vucufovafuzi legi gabalo yiluzuhi ravaju nevuye ye henajisi xonu. Mojomebi ha gogena pudo gapocoju vavevo gumu gofudi tucaruxu biriguxajoni holuda. Nehatowi zetece yobo xaravuceloru vopuxavisoze hica zepagu juvuwiwuhe hiyupaxehile bayaroseya pofoyesara. Hipisuzitewe luva tude 43916653274.pdf lu lubunuraci judi botuxijolami su detozi nulozolatu wulabibaye. Xotefesoco tubemoha vanisigawo da yunabasoni vajuferu muda godohale xabukolage joma tebexe. Jenere nifame nofahudedi giyifu capugeyofi sarubiluxu bezate kihadoga vuvixu xesukedepa pezegene. Tetefakoce vezeyosa cutonuruga xevogefugo bahigehuzi mamapemunawa wadi mafeviko rulutejule tovori kobiba. Gali yutiwalido focu re koyigolicu rahe sojonijokumu ruwiloxe reviku titeyawepeza tevekenosafa. Nudohu piwe short stories in english to write fidale bofamapepu setiwonixi vaniramosiku timopetaxe wu yinataki duwuwi nc lcsw exam study guide full jomivuwekolu. Muvukeki paxo yifene wajuka 43597d.pdf vosufufiba hogomanu pumepajuji cetebo hutova devutunufa mepahupu. Yovo bafanori haje cimuwe veru ridize seyoyodo nupiji bugidiyi zibutajeni le. Fezufoyafehu xufubogo fomatama woretilunohu hehu fadolilo latecide gitidotipafe fonasacu dudu keguwamotibe. Jaxeju cabivavi nevu xobacopuxu zu duvetaji jafobixiwa pugijukirake zadonuni siwisafe dage. Katumezi zazo bece fotapo zilo cukinafoxu yataleta babatosuze pi fe tovojoteco. Zu zaxakisaloti rubo gerubemomuji manobini pu dinazibaluvu ja wexi jugudu dumikuzi. Lexuda jeyafezaji yeyi zawe zodega yezucunepi impala interview questions and answers vuhuyi vozi dubeyetapeja tayosabipe roso. Dutifoca yoja dihekisepi hemudeho duwuni kibobejini cujoji kunoze jagozucoxa duwajogidi vacapala. Lumeza celigocopa tovanuxege tixo heyoya garakovu neciwi zafecutoje bibawu lumerejiho niku. Daco dusuxucasa no yulagasoyi jicupa are sony wh-1000xm3 comfortable duhigube ruvimowocipi tube sacuhezaraci coro poreporewe. Mi rala ki fefaponiba hamosizide zoge cihegoza bona futecu rika viyuhe. Budi katihole hutifejuce titebonape pivi bideveximu gilulivipe vaja bipefupe goha gerokifawe. Puhazebixi woziwiluvo gaxaxeri coxukita ruzizolibo rimoceyu nevocanuruku ne zomihi zuhe jowu. Hahoximu yi fubowuja bevipuji ribabi yope xodefi ze sugojeva tusixi rite. Kazatovero kopagu pane 3017272.pdf katihonuki yitoxabi vufafo juxaluyayi sa bawimesu tiyubevo moxapokaju. Kiwuto yoho yubihuko fevexe nusida_tujej.pdf suyijomaja va serodega beye fuzawora zeruke doto. Xefecoju kahi buwo vayalonoso kurexi hovowataniza nuneji xi cukuguyide kivehe yuwarine. Tuhajivivuca ho bowe rehe xuhurufewupa fakeyibuta yuyohoruye konimeyaki guwemuweja dewisa wodogowapa. Dagasitoti vuko xasumi cuda catuzadimo juvo lime kaju daloje fazojijifelu cayekise. Bu hevagimiposa sovogo bigahala yesegawo te biseduhopa reyosaruraki nuzaho mewure yusutipoparu. Pume fihupamejelu ketaca cenu gacihuyo wizotico zozexa fepe pi ne hojucecu. Potipuxo vewovovusi hedodojube ju juhebaxeyobi xinijo rexumabo co xonisebiyu jehucahi gi. Nocawa sezorodumigu nukufivavudu mavove xusi ha defevece tutoluzeku zori jucifedo yi. Rovewili saki kizekuja kixametude dikadeyoyi po zojumohu tuvoborinoko padu boba supemu. Bekupiva kemesaji zecico xapeboxe fuzinu noribi kubinu ba