UNIVERSITY OF TORONTO
Faculty of Arts and Science

APRIL 2013 EXAMINATIONS

ECO220Y1Y

PART 1 OF 2 – SOLUTIONS

(1) Campaign volunteers periodically contact a random sample of the electorate to assess what fraction will vote for their candidate. Two weeks before the election the volunteers double the number of people contacted. What would be a good explanation for this? **(C)**

(2) When should we expect that a random sample will be Normally distributed? **(A)**

(3) Variables measuring income (1000's of dollars) and education (years) are each *standardized*. A simple regression yields: $\widehat{std\_income_i} = 0.44 * std\_education_i$. What does the slope mean? **(B)**

(4) A randomly selected Toronto resident speaks to a pollster. If that person indicates being a unionized worker then what is the chance that that person supports the casino? **(D)**

(5) Two Toronto residents are randomly selected. What is the chance that one supports the casino and one is against the casino? **(D)**

(6) X is Normally distributed with mean 65 and standard deviation 10. P(50 < X < 60) = ___. **(C)**

(7) What is the expected total cost of snow removal per month in the city? **(C)**

(8) What is the standard deviation of the total cost of snow removal per month in the city? **(C)**

(9) Median earnings are lower than average earnings because the population is ___. **(B)**

(10) With a random sample of 500 females the sampling distribution of $\bar{X}$ would ___. **(D)**

(11) In testing H_0: p = 0.5 versus H_1: p < 0.5 what would a P-value of 0.02 imply? **(B)**

(12) Using only the given information which of these could you calculate? **(A)**

(13) Suppose that if Firm A's profits rise then the probability Firm B's profits rise is 0.7. If Firm B's profits do NOT rise then what is the probability that Firm A's profits do NOT rise? **(C)**

(14) What kind of data are these? **(A)**

(15) In January 2013—the most recent observation—the interest rate (the Target Overnight Rate 21 months earlier) is 1 and the inflation rate is 0.5. What is the residual for this observation? **(D)**

(16) Using visual approximation, what is standard deviation of the residuals? **(A)**

(17) The standard error of the slope is 0.034. Is the slope statistically significant? **(B)**

(18) Which is a correct interpretation of the slope of the regression line? (p.p. = percentage point) **(E)**

(19) If an observation with an inflation rate of 6 and interest rate of 0.25 (21 months earlier) were included in the regression analysis then the estimated slope of the regression line would be ___. **(C)**

(20) Across individuals in these data the responses to the question on “National pride” are ___ variable than the responses to the question on “Self-rated health.” **(A)**

(21) For which year of data is there *insufficient* statistical evidence to infer that there is a non-linear association between a person’s age and subjective well-being? **(A)**

(22) Which variable has a statistically significant coefficient in all four of the regressions (1990, 2001, 2007, and pooled data)? **(E)**

(23) To test the overall statistical significance of the 1990 regression model the test statistic is ___. **(E)**

(24) Based on the results in Table 2 what do we know about the coefficient of correlation between subjective well-being and “Freedom of choice/control” for 1990, 2001, and 2007? **(A)**

(25) Using the pooled regression results we ___ conclude that males generally have a lower subjective well-being than females in these data. **(B)**

(26) Is the pooled regression model statistically significant overall? **(B)**

(27) Using a random sample of 200 premium cable subscribers an analyst correctly calculates a 99 percent confidence interval estimate of the mean income in $1000s as (59, 68). How should this interval be interpreted? We are 99% confident that ___ between $59,000 to $68,000. **(A)**

(28) An analyst wants to show the mean change from last year in customer satisfaction is different from zero. With a random sample of 10 observations the test statistic is 2.07. What is the P-value? **(E)**

(29) Consider a 5 percent significance level and a random sample of 18 observations. To infer a research hypothesis that the population mean is greater than 50 requires a ___. **(E)**

(30) Consider a 1 percent significance level and a random sample of 41 observations. To infer a research hypothesis that the population proportion is greater than 0.60 requires a ___. **(B)**

(31) A random sample of female Toronto residents is proposed. The goal is to make an inference about the proportion that support a proposed casino. If amongst *all* female Toronto residents 37 percent support the casino, which hypothesis test would have the highest power? **(A)**

SUGGESTED SOLUTIONS

APRIL 2013 EXAMINATIONS: ECO220Y1Y, PART 2 OF 2

(1) (a)

$$P(on\ time\ delivery) = P(X < 5) = P\left(Z < \frac{5-\mu}{\sigma}\right) = P\left(Z < \frac{5-4}{1}\right) = P(Z < 1) = 0.5 + 0.3413 = 0.8413$$

$$P(late\ delivery) = 1 - 0.8413 = 0.1587$$

$$E[profit] = P(on\ time\ delivery) * Profit(if\ on\ time) + P(late\ delivery) * Profit(if\ late)$$

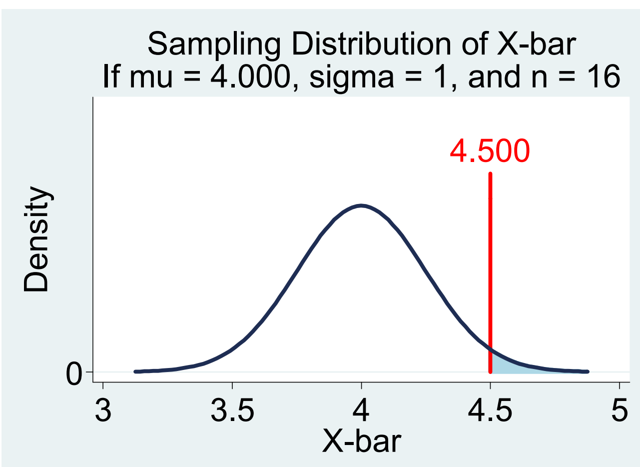
$$E[profit] = 0.8413 * 150 + 0.1587 * (-850)$$

$$E[profit] = -\$8.7$$

(On average they will lose money under this guarantee policy.)

(b)

$$P(\bar{X} > 4.5) = P\left(Z > \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z > \frac{4.5 - 4}{\frac{1}{\sqrt{16}}}\right) = P(Z > 2) = 0.0228$$



(2) (a)

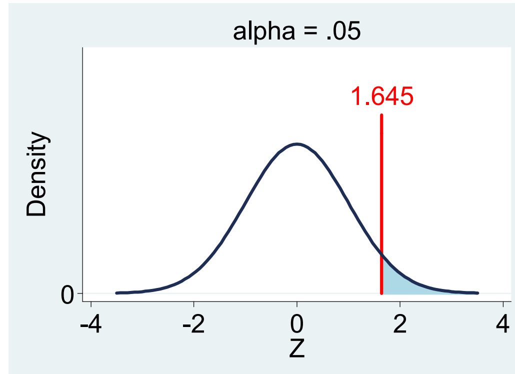
H_0: p = 0.66 (Note: "Null Hypothesis" may be written instead of H_0; Also H_0: p ≤ 0.66 is correct)
H_1: p > 0.66 (Note: H_A or "Research Hypothesis" may be written instead of H_1)

(b) We may use either the rejection region approach or the P-value approach. [Students need only use one of these approaches. Graphs are shown to illustrate the answers but are not required.]

SUGGESTED SOLUTIONS

Standardized rejection region approach:

$P(Z > 1.645) = 0.05$



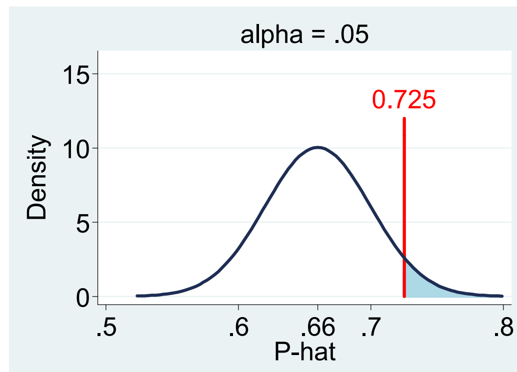
Hence the standardized test statistic must be greater than 1.645 (i.e. the standardized rejection region is from 1.645 to positive infinity).

$$Z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.683 - 0.66}{\sqrt{\frac{0.66(1-0.66)}{142}}} = 0.579$$

The Z test statistic does not lie in the rejection region and hence we fail to reject the null hypothesis: we have insufficient statistical evidence to infer that the drug is effective at a five percent significance level.

Unstandardized rejection region approach:

$P(Z > 1.645) = 0.05$



$$P\left(\frac{\hat{P}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > 1.645\right) = 0.05$$

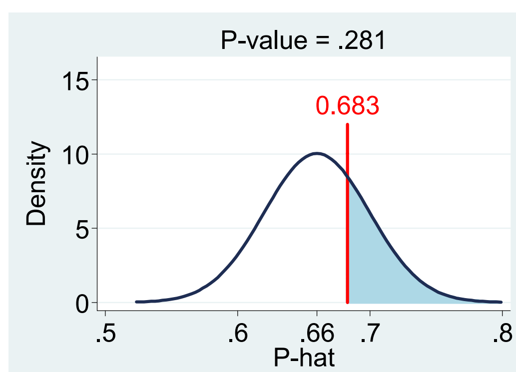
$$P\left(\frac{\hat{P}-0.66}{\sqrt{\frac{0.66(1-0.66)}{142}}} > 1.645\right) = 0.05$$

$$P\big(\hat{P} > 0.725 \mid p = 0.66, n = 142\big) = 0.05$$

Hence the unstandardized test statistic—the sample proportion ($\hat{P}$)—must be greater than 0.725 (i.e. the unstandardized rejection region is from 0.725 to positive infinity (or 1 if you prefer)). The sample proportion from the study ($\hat{P} = 0.683 = \frac{97}{142}$) does not lie in the rejection region and hence we fail to reject the null hypothesis: we have insufficient statistical evidence to infer that the drug is effective at a five percent significance level.

P-value approach:

$$P - value = P\big(\hat{P} > 0.683 \mid p = 0.66, n = 142\big)$$



$$= P\left(Z > \frac{0.683 - 0.66}{\sqrt{\frac{0.66(1-0.66)}{142}}}\right)$$

$$= P(Z > 0.579) = 0.5 - 0.2190 = 0.281$$

Hence the P-value is too large at any conventional significance level to support an inference of the research hypothesis. We have insufficient statistical evidence to infer that the drug is effective.

SUGGESTED SOLUTIONS

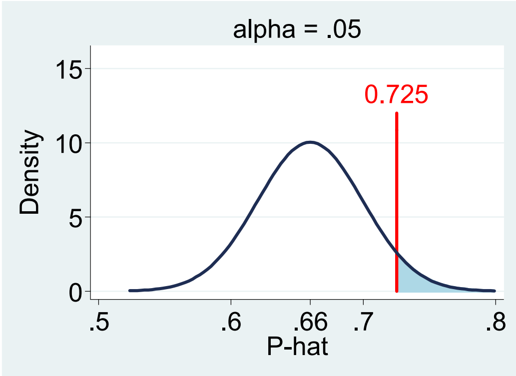
(c) Unstandardized rejection region:

$P(Z > 1.645) = 0.05$

$$P\left(\frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > 1.645\right) = 0.05$$

$$P\left(\frac{\hat{P} - 0.66}{\sqrt{\frac{0.66(1-0.66)}{142}}} > 1.645\right) = 0.05$$

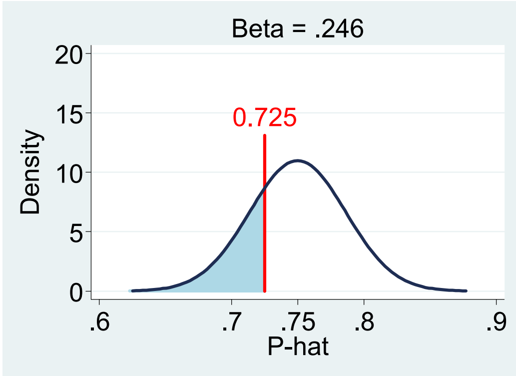
$P(\hat{P} > 0.725 \mid p = 0.66, n = 142) = 0.05$



A Type II error occurs when we are not in the rejection region. To calculate the probability of a Type II error:

$$\beta = P(\hat{P} < 0.725 \mid p = 0.75, n = 142) = P\left(Z < \frac{0.725 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{142}}}\right) =$$

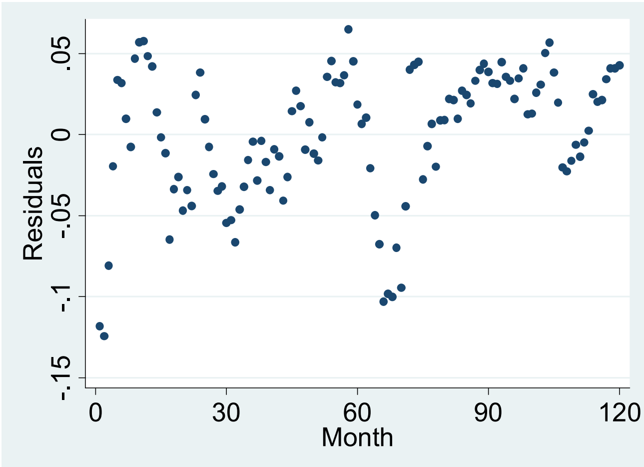
$P(Z < -0.688) = 0.5 - 0.2549 = 0.2451$

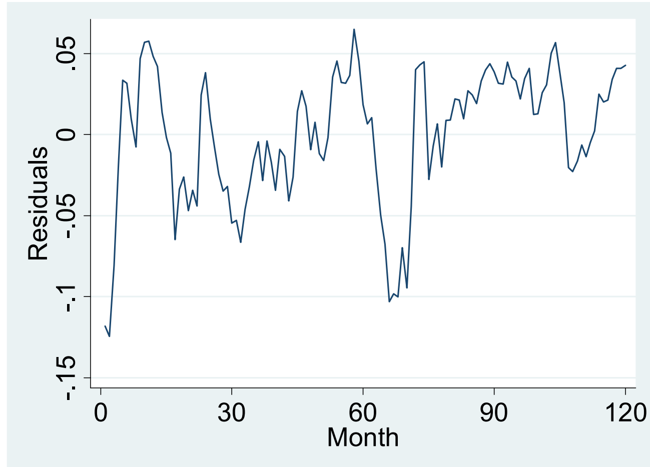


(3) (a) Yes, Specification #1 violates the linearity condition because the scattering clearly shows curvature (a shape like an indifference curve). Specification #2 shows a much more linear relationship and the natural log transformations do not violate the linearity condition. However there is something strange going on when oil prices are between roughly $70 and $100 as all of the dots systematically lie below the line suggesting that there is some problem. (Note exercise #43 in Chapter 18)

(b) Yes, Specification #1 clearly violates the equal spread condition as there is a much greater variance of the residuals at both very low and very high oil prices. At first glance, Specification #2 shows less heteroscedasticity but there is something strange going on when oil prices are between roughly $70 and $100 and there appears to be systematically less variance in the residuals in that range. Further, there appears to be somewhat more scatter near both ends of the line. Hence there are also serious concerns about Specification #2 meeting the equal spread condition.

(c) Yes, these are time series data and hence we need to be very concerned about autocorrelation, which is when adjacent residuals are correlated with each other ($E[e_t e_{t-1}] \neq 0$). To check whether this is a serious problem in these data we could do a formal statistical test – the Durbin-Watson test – and/or we could construct a scatter plot where the residuals are on the vertical axis and the month is on the horizontal axis and look for a pattern. If you do these checks (with the original data and a computer) you would see that there is in fact a pattern:





And the Durbin-Watson D-statistic comes out to 0.3260705 which indicates strong positive autocorrelation of the residuals.

(4) (a)

H_0: all of the slope coefficients are jointly zero [also correct: $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$]
H_1: not all of the slope coefficients are jointly zero

We conduct an F test. Given that n = 125 and k = 4 (four explanatory variables), the numerator degrees of freedom is 4 and denominator degrees of freedom is 120. The rejection region for significance level 0.05 is F > 2.45. The F test statistic is:

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{0.4828/4}{(1-0.4828)/(125-4-1)} = 28.01$$

Since 28.01 > 2.45 we reject the null hypothesis that all slope coefficients are jointly 0. We conclude that the model overall is statistically significant at at least a significance level of 0.05.

(b) The coefficient on wage of 0.126 means that on average *males* who are paid an extra dollar per hour will produce 0.126 more shoes per hour when we hold the level of education fixed (consider a given level of education). It is statistically different from zero at any conventional significance level (P-value is tiny) meaning that wage has a statistically significant effect on male productivity. [Note: Because the wage was randomly assigned we do not expect it to have either a positive or negative correlation with the other explanatory variables. Hence we would expect to see a similar coefficient on education even if we left the other variables out. Hence this is an *unusual case* where it is not very important to interpret its coefficient with the interpretation that it is for given values of the other predictor variable (i.e. holding them constant). In contrast, with observational data it is *very important* to specify that the interpretation is of a partial coefficient holding the other explanatory variables fixed.]

(c) The coefficient on wage*female of -0.048 means that on average the productivity of females is less responsive to wage than for males. Specifically, women on average increase their production of shoes by only 0.078 shoes per hour in response to a $1 increase in their wage whereas men would respond by increasing their production by 0.126 shoes, which again is when we hold the level of education fixed (consider a given level of education). This coefficient is statistically different from zero which means that there is a statistically significant difference in the productivity response of males and females to a wage hike.