ECON 395 Spring 2014

DEPARTMENT OF ECONOMICS - MIDTERM EXAMINATION

ECON 395 - SOLUTION

Spring 2014

June 9th, 2014 **Marian Miles**

INSTRUCTIONS:

The exam will last for 120 minutes. Please allocate your time wisely. You can obtain a total of **100 points**. You must always write the formula and some steps in your calculations to receive points. **Answer all questions with a sentence answer.** You may use non-programmable calculators during the exam.

1. The following table gives the joint probability distribution between employment status and college graduation from Canada.

Employment status and College Graduation in Canada

		Unemployed Y=0	**Employed Y=1**	**Total**
Education	**Non-college grads (X=0)**	0.045	**0.709**	**0.754**
	College grads (X=1)	**0.005**	0.241	**0.246**
	Total	0.050	**0.95**	**1**

a) Fill in the missing probabilities (both joint and marginal) **in the table**. (4 marks)

b) A randomly selected member of the population reports being unemployed. What is the probability that this person is a college graduate? (4 marks)

P(X=1|Y=0)= P(1,0)/P(Y=0) = 0.005/0.050 = 0.1

The probability that an unemployed person is a college graduate is 0.1 or 10%.

c) Are educational achievement and employment status independent? Prove your answer. (5 marks)

If independent then P(Y=0|X=1)=P(Y=0) but 0.0203≠0.050. **Education achievement and employment status are <u>not</u> independent.**
If independent then P(Y=0 and X=1)=P(Y=0) * P(X=1) but 0.005≠0.012 = 0.05*0.246.
Education achievement and employment status are <u>not</u> independent.

2. Let Y_1, Y_2, Y_3, and Y_4 be independent, identically distributed random variables from a population with a mean μ and a variance σ^2. Consider a different estimator of μ:

$$W = \frac{1}{8}Y_1 + \frac{1}{8}Y_2 + \frac{1}{4}Y_3 + \frac{1}{2}Y_4 .$$

This is an example of a weighted average of the Y_i.

a) Is W an unbiased estimator of μ? Show that it is – or it isn't (E(W) = ?). (4 marks)

$$E(W)=E(\frac{1}{8}Y_1 + \frac{1}{8}Y_2 + \frac{1}{4}Y_3 + \frac{1}{2}Y_4)= \frac{1}{8}E(Y_1)+ \frac{1}{8}E(Y_2)+ \frac{1}{4}E(Y_3)+ \frac{1}{2}E(Y_4)$$
$$= \frac{1}{8}\mu+ \frac{1}{8}\mu + \frac{1}{4}\mu + \frac{1}{2}\mu= \mu$$

<u>**W is an unbiased estimator of μ**</u>.

b) Find the variance of W. (4 marks)

$$var(W) = var(\frac{1}{8}Y_1 + \frac{1}{8}Y_2 + \frac{1}{4}Y_3 + \frac{1}{2}Y_4) = var(\frac{1}{8}Y_1) + var(\frac{1}{8}Y_2) + var(\frac{1}{4}Y_3) + var(\frac{1}{2}Y_4)$$
$$= (\frac{1}{8})^2 var(Y_1) + (\frac{1}{8})^2 var(Y_2) + (\frac{1}{4})^2 var(Y_3) + (\frac{1}{2})^2 var(Y_4) = \frac{22}{64}\sigma^2 = \mathbf{\frac{11}{32}\sigma^2}$$

The variance of W is $\frac{11}{32}\sigma^2$.

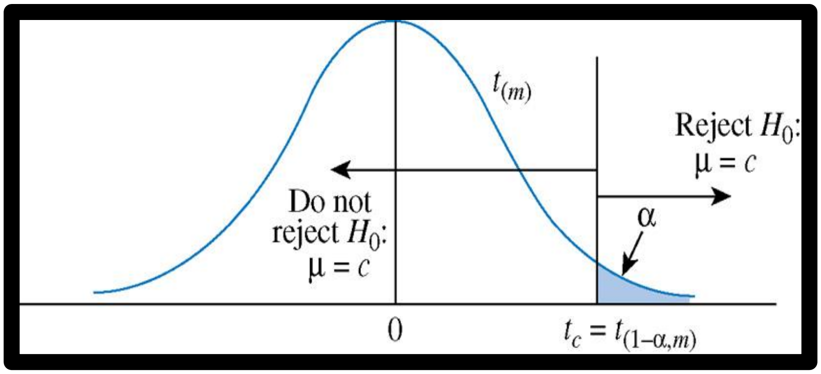
3. How much money do winners go home with from the television quiz show *Jeopardy?* To determine an answer, a random sample of winners was drawn and the amount of money each won was recorded. The mean of the sample of 25 winners was $26,050 and the standard deviation of the sample was $8,782.6. Test the hypothesis, at a 5% significance level, that the mean winnings are greater than $23,000.00

i. State the appropriate null and alternative hypothesis. (2 marks)

$$H_0: \mu = \$23{,}000$$
$$H_A: \mu > \$23{,}000$$

ii. State the sampling distribution, the significance level, the degrees of freedom, and draw the distribution with the rejection region. (5 marks)

$\alpha = .05$, Right Tail Test, df = 25 – 1 = 24, $t_c = t_{(.95,\ 25-1)} = 1.711$



iii. Calculate the test statistic. (4 marks)

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{26{,}050 - 23{,}000}{8782.6/\sqrt{25}} = 1.736$$

iv. State your decision rule, test the hypothesis. (3 marks)

Two Solutions:

1. Decision Rule: Reject H_0 if $t > t_c$, 1.711 < 1.736 Reject H_0
2. Decision Rule: Reject H_0 if *p-value*< $\alpha = .05$, 0.05 > $P(t_{24} > 1.736)$ > 0.025 Reject H_0.

v. State your conclusion. (2 marks)

There is sufficient evidence in this sample to reject the hypothesis that the mean winnings in the television quiz show *Jeopardy* are $23,000., at the 5% significance level. The evidence supports the alternative that the mean winnings are greater than $23,000.

4. **Rewrite each of the following statements to make the statement true:** (40 marks)

a) If two random variables have a covariance equal to “0”, then the two variables are statistically independent.

If two random variables are statistically independent, then the two variables have a covariance equal to “0”.

b) $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x^2 + n\bar{x}^2$ $\quad \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x^2 - n\bar{x}^2$

c) Let X and Y be discrete random variables with joint pdf f(x,y). Then E(aX + bY + c) would equal E(aX - bY + c) since we square the constants, a and b, when we apply the rules of expected values.

Let X and Y be discrete random variables with joint *pdf f(x,y)*. Then *var(aX + bY + c)* equals *var(aX - bY + c)* since we square the constants, *a* and *b*, when we apply the rules of variances.

d) If the Central Limit Theorem holds, that is, that the sampling distribution of $\overline{\mathrm{X}}$ is normally distributed, then when we test a hypothesis with sample information about a single mean we can always use a test statistic that is normally distributed.

If the Central Limit Theorem holds, that is, that the sampling distribution of $\bar{X}$ is normally distributed, then when we test a hypothesis about a single mean we can use a test statistic that is normally distributed <u>only</u> when we know the population variance, σ^2.
OR
If the Central Limit Theorem holds, that is, that the sampling distribution of $\bar{X}$ is normally distributed, then when we test a hypothesis with sample information about a single mean we can use a test statistic that is student t distributed.

e) A confidence interval can be used to test a hypothesis by checking if the alternative hypothesis value would lie between (inside) the confidence interval lower and upper bound.

A confidence interval can be used to test a two-tailed hypothesis by checking if the null hypothesis value would lie between (inside) the confidence interval lower and upper bound.

f) A Type II error is when we decide to reject the null hypothesis and the null hypothesis is in fact true.

A Type I error is when we decide to reject the null hypothesis and the null hypothesis is in fact true.

OR A Type II error is when we decide to not reject the null hypothesis and the null hypothesis is in fact false.

g) In regression analysis the dependent variable, Y, is a random variable in the model because the independent variable, X, is a random variable.

In regression analysis the dependent variable, *Y*, is a random variable in the model because the error term, e_i, is a random variable.

h) The Gauss-Markov Theorem tells us that the estimators b_1 and b_2 are BLUE, which proves the estimators are unbiased.

The Gauss-Markov Theorem tells us that the estimators b_1 and b_2 are BLUE, which proves the estimators have the smallest variance in the class of linear and unbiased estimators if SR1 – SR5 are true.

i) A large sample size, n, will make the sum of squared errors larger in regression analysis, and so the variance of the random error term, σ^2, will be larger.

A large sample size, n, will make the sum of squared errors larger in regression analysis only if there is more error, but since the denominator is n-k, the variance of the random error term, σ^2, will grow smaller as n gets larger.

j) When we interpret the estimate for the slope coefficient, b_2, we must always state if the interpretation can be used.

When we interpret the estimate for the intercept coefficient, b_1, we must always state if the interpretation can be used.

5. One factor that we expect to affect the price of a home is the number of square feet in the home. If the house is large we would expect to pay a higher price. While this is intuitively reasonable, let us ask the question "If the number of square feet in a house increases by 100, how much does the price of the house rise?" A simple specification for this model is given by

$$PRICE_i = \beta_1 + \beta_2 SQFT_i + \varepsilon_i, \qquad (1)$$

where $PRICE_i$ is in \$1000 of dollars and $SQFT_i$ is hundreds of square feet in living space. (Assume the error term is i.i.d. $N(0,\sigma^2)$.)

A sample of data of 15 house sales gives you the following information:

$$\sum_{i=1}^{15} PRICE_i = 1{,}718.840 \qquad \sum_{i=1}^{15} SQFT_i = 248.81$$

$$\sum_{i=1}^{15}(PRICE_i - \overline{PRICE})(SQFT_i - \overline{SQFT}) = 6{,}870.7511$$

$$\sum_{i=1}^{15}(PRICE_i - \overline{PRICE})^2 = 75{,}149.2113; \qquad \sum_{i=1}^{15}(SQFT_i - \overline{SQFT})^2 = 846.9803$$

$$\sum_{i=1}^{15}(PRICE_i - \widehat{PRICE}_i)^2 = 19{,}413.2979$$

a) Determine the ordinary least squares regression line. (5 marks)

$$b_2 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{6870.7511}{846.9803} = 8.1121$$

$$b_1 = \bar{Y} - b_2\bar{X} = \left(\frac{1718.84}{15}\right) - \left(8.1121 * \left(\frac{248.81}{15}\right)\right) = 114.589 - (134.5574) = -19.9680$$

The equation to the regression line is $\widehat{PRICE}_i = -19.968 + 8.112 SQFT_i$

b) Interpret the coefficients of the regression line. (5 marks)

The average sale price of land, without a house (the number of square feet is zero) is negative \$19,968.00. Since this obviously does not make sense, and we do not know the range of values for the data, this should not be used.

As the number of square feet in the house rises by one hundred the price of the house rises by \$8,112.10 on average. [OR An extra square foot is worth, on average, \$81.12 in the selling price of a house.]

c) Using the equation you calculated in part (a), what is the selling price of a house with 1650 square feet of living space? (3 marks)

$$P\hat{RICE}_i = -19.968 + 8.112 SQFT_i = -19.968 + (8.112 * 16.5) = 113.8816$$

The average selling price of a house with 1650 square feet of living space is $113,881.60.

Since you are curious about your results you get a much larger sample, and you use Stata to get the following output:

regress PRICE SQFT

Source	**SS**	**df**	**MS**
Model	1647887.086	1	1647887.09
Residual	**803913.4198**	878	915.6189
Total	2451800.506	879	2789.3066

Number of obs	=	880
F(1, 878)	=	1799.752
Prob > F	=	0.0000
R-squared	=	?????
Adj R-squared	=	0.6718
Root MSE	=	30.2592

PRICE	**Coef.**	**Std. Err.**	**t**	**P>t**	**[95% Conf.**	**Interval]**
SQFT	8.13889	0.19185	42.423	0.000	.7623	8.5154
_cons	-18.3857	3.2564	-5.646	0.000	-24.7769	-11.994

d) Write the equation in standard form. (4 marks)

$$\widehat{Price} = -18.386 + 8.139 SQFT \qquad OPTIONAL: R^2 = 0.6721$$

(s.e.) (3.256) (0.192)

e) What is the Sum of Squares for the error (SSE)? (2 marks)

The Sum of Squares for the error (SSE) equals 803,913.4198. (See the regression results above.)

f) Using the information from your Stata regression, answer the question, "If the number of square feet in the house rises by 100, how much does the price of the house rise?" (4 marks)

As the number of square feet in the house rises by one hundred the price of the house rises by an average of $8,138.89 (this is the value of b_2).