Econ 312: Midterm I

Thursday, March 19

Please do not turn this page over until instructed to do so.

Instructions (Please Read Carefully Before Starting)

- This test has a total of **100 points**. Unless otherwise instructed, you have 1h 50m to solve it, that is, 110 minutes. There are 15 multiple choice questions (each is worth 3 points) and 5 written questions 4, 5,14,24, and 8 points respectively).
- Show your work, unless you are explicitly told not to ! No credit will be given for correct answers if you do not justify your argument.
- Please be sure that your handwriting is **legible**!
- We will grade only what is written on your exam sheet. There should be plenty of space for all your answers. **Do not turn in anything aside from your exam sheet.**
- If time is running short, you should try to set up the problem without doing the final calculations.

Name :____

Signature:____

Multiple Choice (Just answer writing the letter corresponding to the statement you believe to be correct.)

Question	Answer
1	d
2	c
3	b
4	b
5	c
6	c
7	d
8	b
9	a
10	c
11	a
12	d
13	d
14	d
15	b
Score	

Part I. Multiple Choice (15 questions worth 3 points each).

1.Two random variables X and Y are independently distributed if all of the following conditions hold, with the exception of

a. $\Pr(Y = y|X = x) = \Pr(Y = y)$.
b. knowing the value of one of the variables provides no information about the other.
c. if the conditional distribution of Y given X equals the marginal distribution of Y.
d.$E(Y) = E(X|Y)$.

2. Assume that Y is normally distributed $N(\mu, \sigma^2)$. Moving from the mean (μ) 1.96 standard deviations to the left and 1.96 standard deviations to the right, then the area under the normal p.d.f. is

a. 0.67
b. 0.05
c. 0.95
d. 0.33

3. When there are ∞ degrees of freedom, the t distribution

a. can no longer be calculated.
b. equals the standard normal distribution.
c. has a bell shape similar to that of the normal distribution, but with "fatter" tails.
d. equals the χ_∞ distribution.

4. The sample average is a random variable and

a. is a single number and as a result cannot have a distribution.
b. has a probability distribution called its sampling distribution.
c. has a probability distribution called the standard normal distribution.
d. has a probability distribution that is the same as for the $Y_1 \ldots, Y_n$ i.i.d. variables.

5. The correlation between X and Y

a. cannot be negative since variances are always positive.
b. is the covariance squared.
c. can be calculated by dividing the covariance between X and Y by the product of the two standard deviations.
d. is given by $corr(X, Y) = \frac{Cov(X,Y)}{Var(X)Var(Y)}$.

6. The following types of statistical inference are used throughout econometrics, with the exception of

a. confidence intervals.
b. hypothesis testing.
c. calibration.
d. estimation.

7.The OLS estimator is derived by

a. connecting the Y_i corresponding to the lowest Xi observation with the Y_i corresponding to the highest X_i observation.
b. making sure that the standard error of the regression equals the standard error of the slope estimator.
c. minimizing the sum of absolute residuals.
d. minimizing the sum of squared residuals.

8. The following are all least squares assumptions with the exception of:

a. The conditional distribution of ϵ_i given X_i has a mean of zero.
b. The explanatory variable in regression model is normally distributed.
c. $(X_i, Y_i), i = 1, \ldots, n$ are independently and identically distributed.
d. Large outliers are unlikely.

9. The slope estimator, β_1, has a smaller variance, other things equal, if

a. there is more variation in the explanatory variable, X.
b. there is a large variance of the error term.
c. the sample size is smaller.
d. the intercept, β_0, is small.

10. The t-statistic is calculated by dividing

a. the OLS estimator by its standard error.
b. the slope by the standard deviation of the explanatory variable.
c. the estimator minus its hypothesized value by the standard error of the estimator.
d. the slope by 1.96.

11. The confidence interval for the sample regression function slope

a. can be used to conduct a test about a hypothesized population regression function slope.
b. can be used to compare the value of the slope relative to that of the intercept.
c. adds and subtracts 1.96 from the slope.
d. allows you to make statements about the economic importance of your estimate.

12. An estimator $\hat{\beta}_1$ of the population value β_1 is unbiased if

a. $\hat{\beta}_1 = \beta_1$
b. $\hat{\beta}_1$ has the smallest variance of all estimators.
c. $\hat{\beta}_1 \rightarrow^p \beta_1$.
d. $E(\hat{\beta}_1) = \beta_1$

13. An estimator $\hat{\beta}_1$ of the population value β_1 is consistent if

a. $\hat{\beta}_1 \to^p \beta_0$.
b. its mean square error is the smallest possible.
c. $\hat{\beta}_1$ is normally distributed.
d. $\hat{\beta}_1 \to^p \beta_1$.

14. An estimator $\hat{\beta}_1$ of the population value β_1 is more efficient when compared to another estimator $\tilde{\beta}_1$, if

a. $E(\hat{\beta}_1) < E(\tilde{\beta}_1)$.
b. it has a smaller variance.
c. its c.d.f. is flatter than that of the other estimator.
d. both estimators are unbiased, and $Var(\hat{\beta}_1) < Var(\tilde{\beta}_1)$.

15. A type I error is

a. always the same as (1-type II) error.
b. the error you make when rejecting the null hypothesis when it is true.
c. the error you make when rejecting the alternative hypothesis when it is true.
d. always 5%.

Part II. Written Questions.

Question 1.(4pts) Clearly state the Central Limit Theorem.

If $(Y_1, \ldots, Y_n)$ are i.i.d and $\sigma_Y^2 < \infty$, then when n is large the distribution of $\bar{Y}$ is well approximated by a normal distribution.

- $\bar{Y}$ is approximately distribution $N(\mu_Y, \frac{\sigma_Y^2}{n})$ ("normal distribution with mean μ_Y and variance σ_Y^2/n")
- $\sqrt{n}(\bar{Y} - \mu_Y)/\sigma_Y$ is approximately distributed $N(0,1)$ (standard normal)
- That is , "standardized" $\bar{Y} = \frac{\bar{Y}-E(\bar{Y})}{\sqrt{Var(\bar{Y})}} = \frac{\bar{Y}-\mu_Y}{\sigma_Y/\sqrt{n}}$ is approximately distributed as $N(0,1)$.
- The large is n, the better is the approximation.

Question 2.(5pts) Explain the difference between $\hat{\beta}_1$ and β_1; between the residual $\hat{\epsilon}_i$ and the regression error ϵ_i.

β_1 is the value of the slope in the population regression. The is value is unknown. $\hat{\beta}_1$ (an estimator) gives a formula for estimating the unknown value of β_1 from a sample. Similarly, ϵ_i is the value of the regression error for the i-th observation;ϵ_i is the difference between Y_i and the population regression line $\beta_0 + \beta_1 X_i$. Because the value of β_0 and β_1 are unknown, the value of ϵ_i is unknown. In contrast, $\hat{\epsilon}_i$ is the difference between Y_i and $\hat{\beta}_0 + \hat{\beta}_1 X_i$; thus , $\hat{\epsilon}_i$ is an estimator of ϵ_i.

Question.3 (14pts) In a survey of 1000 likely voters, 550 responded that they would vote for the incumbent and 450 responded that they would vote for the challenger. Let p denote the fraction of all likely voters who preferred the incumbent at the time of the survey, and let $\hat{p}$ be the fraction of survey respondents who preferred the incumbent.

a.(2pts) Use the survey result to estimate p.

$\hat{p} = \frac{550}{1000} = 0.55$

---

b.(4pts) Use the estimator of the variance of $\hat{p}$, $\frac{\hat{p}(1-\hat{p})}{n}$, to calculate the standard error of your estimator.

$\widehat{\text{Var}}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n} = \frac{0.55\times(1-0.55)}{1000} = 0.0002475.$
Thus, the standard error is $SE(\hat{p}) = \sqrt{\widehat{\text{Var}}(\hat{p})} = 0.0157.$

---

c.(4pts) What is the p-value for the test $H_0 : p = 0.5$vs. $H_1 : p > 0.5$?

The computed t-statistic is $t = \frac{\hat{p}-p_0}{SE(\hat{p})} = \frac{0.55-0.5}{0.0157} = 3.185$. Because of the large sample size ($n = 1000$), the p-value for the test $H_0 : p = 0.5$vs. $H_1 : p > 0.5$:

$$\text{p-value} = \Pr(Z > 3.18|H_0 : p = 0.5) = 1 - \Phi(3.185) = 0.0007.$$

---

d.(4pts) Did the survey contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey? Explain.

For the test $H_0 : p = 0.5$vs. $H_1 : p > 0.5$, we can reject the null hypothesis at the 5% significant level. The p-values 0.0007 is less than 0.05. Equivalently the calculated t-statistic 3.185 is greater than the critical value 1.645 for a one sided test with a 5% significant level. The test suggests that the survey contained statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.

Question 4. (24 pts) A regression of average weekly earnings (AWE, measured in dollars) on age (measured in years) using a random sample of college-educated full-time workers aged 25-65 yield the following:

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, SER = 624.1$$

a. (4pts) Explain what the coefficient values 696.7 and 9.6 mean.

The coefficient 9.6 shows the marginal effect of Age on AWE; that is, AWE is expected to increase by $9.6 for each additional year of age. 696.7 is the intercept of the regression line. It determines the overall level of the line.

---

b. (4pts) The standard error of the regression is 624.1. What are the unit of measurement for the SER (dollars? Years? or is SER unit-free)? What is the interpretation of the SER here?

SER is in the same units as the dependent variable (Y, or AWE in this example). Thus SER is measures in dollars per week. The SER is a measure of the spread of the observations around the regression line. The magnitude of the typical deviation from the regression line or the typical regression error here is 624.1 dollars.

---

c. (4pts) The regression R^2 is 0.023. What are the unit of the measurement for the R^2 (dollars? Years? or is R^2 unit-free)? What is the meaning of the regression R^2 here?

R^2 is unit free. The regression $R^2 = 0.023$ indicates that 2.3 percent of the variation in AWE is explained by the model.

---

d. (4pts) What is the regression's predicted earnings for a 25-year-old worker?

$696.7 + 9.6 \times 25 = \$936.7$.

---

e.(4pts) Will the regression give reliable predictions for a 99-years-old worker? Why or why not?

No. The oldest worker in the sample is 65 years old. 99 years is far outside the range of the sample data.

---

f. (4pts) The average age in this sample is 41.6 years. What is the average value of AWE in the sample? (*Hint:* $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$)

The sample mean of AWE is $696.7 + 9.6 \times 41.6 = \$1,096.06$.

Question. 5 (8pts) Suppose that a researcher, using data on class size (CS) and average test scores from 1000 third-grad classes, estimates the OLS regression,

$$\widehat{TestScore} = 520.4 - 5.82 \times CS, \quad R^2 = 0.08, SER = 11.5$$

$$(20.4) \ \ (2.21)$$

a. (4pts)Construct a 95% confidence interval for β_1, the regression slope coefficient. The 95% confidence interval for β_1 is $\{-5.82 \pm 1.96 \times 2.21\}$ that is , [-10.152, -1.48].

---

b. (4pts)Calculate the p-value for the two side test of the null hypothesis $H_0 : \beta_1 = 0$. Do you reject the null hypothesis at the 5% level? At the 1% level?

The t-statistic:

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = -2.6335$$

The p-value for the test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$

$$= 2\Phi(-2.6335) = 2 \times 0.0042 = 0.0084$$

The p-value is not larger than 0.01 so we can reject the null hypothesis at the 5% significance level and also at the 1% significance level.
