Econ 312: Midterm I

Thursday, March 18

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 3, 9,16,16, and 11 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	a
2	d
3	d
4	b
5	d
6	c
7	b
8	a
9	b
10	b
11	c
12	d
13	d
14	d
15	c
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. $\Pr(Y = y|X = x) = \Pr(X = x)$.
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.$E(Y) = E(Y|X)$.

2. The skewness of the distribution of a random variable Y is defined as follows:

a. $\frac{E(Y^3 - \mu_Y)}{\sigma_Y^2}$
b. $E(Y - \mu_Y)^3$
c. $\frac{E(Y^3 - \mu_Y^3)}{\sigma_Y^3}$
d. $\frac{E(Y - \mu_Y)^3}{\sigma_Y^3}$

3. The Student t_m distribution is

a. the distribution of the sum of m squared independent standard normal random variables.
b. the distribution of a random variable with a chi-squared distribution with m degrees of freedom, divided by m.
c. always well approximated by the standard normal distribution.
d. the distribution of the ratio of a standard normal random variable, divided by the square root of an independently distributed chi-squared random variable with m degrees of freedom divided by m.

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 variances.
d. is given by $corr(X, Y) = \frac{E(X-\mu_X)(Y-\mu_Y)}{\sqrt{E(X-\mu_x)^2}\sqrt{E(Y-\mu_Y)^2}}$.

6. The critical value of a two-sided t-test computed from a large sample

a. is 1.64 if the significance level of the test is 5%.
b. cannot be calculated unless you know the degrees of freedom.
c. is 1.96 if the significance level of the test is 5%.

d. is the same as the p-value.

7. The t-statistic for $H_0 : \mu_Y = \mu_0$ is defined as follows:

a. $t = \frac{\bar{Y}-\mu_0}{\sigma_Y^2/\sqrt{n}}$ b. $t = \frac{\bar{Y}-\mu_0}{s_Y/\sqrt{n}}$ c. $t = \frac{(\bar{Y}-\mu_0)^2}{\sigma_Y^2/n}$ d. 1.96

8. The regression R^2 is defined as follows:

a. $\frac{ESS}{TSS}$
b. $\frac{RSS}{TSS}$
c. $\frac{TSS}{ESS}$
d. $\frac{RSS}{ESS}$

9. 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.

10. The sample average of the OLS residuals is

a. some positive number since OLS uses squares.
b. zero.
c. unobservable since the population regression function is unknown.
d. dependent on whether the explanatory variable is mostly positive or negative.

11. The OLS estimator is derived by

a. connecting the Y_i corresponding to the lowest X_i observation with the Y_i corresponding to the highest Xi observation.
b. making sure that the standard error of the regression equals the standard error of the slope estimator.
c. minimizing the sum of $\hat{\epsilon}_i^2$.
d. minimizing the sum of residuals .

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.$E(\epsilon_i|X_i) = 0$ says that

a. dividing the error by the explanatory variable results in a zero (on average).
b. the sample regression function residuals are unrelated to the explanatory variable.
c. the sample mean of the Xs is much larger than the sample mean of the errors.
d. the conditional distribution of the error given the explanatory variable has a zero mean.

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. To obtain the slope estimator using the least squares principle, you divide the

a. sample variance of X by the sample variance of Y.
b. sample covariance of X and Y by the sample variance of Y.
c. sample covariance of X and Y by the sample variance of X.
d. sample variance of X by the sample covariance of X and Y.

Part II. Written Questions.

Question 1.(3pts) 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 (9pts) Consider the following alternative estimator for the population mean: .

$$\tilde{Y} = \frac{1}{n}(\frac{1}{2}Y_1 + \frac{3}{2}Y_2 + \frac{1}{2}Y_3 + \frac{3}{2}Y_4 + \cdots + \frac{1}{2}Y_{n-1} + \frac{3}{2}Y_n)$$

Prove that $\tilde{Y}$ is unbiased and consistent, but not efficient when compared to $\bar{Y}$.

$E(\tilde{Y}) = \frac{1}{n}(\frac{1}{2}E(Y_1) + \frac{3}{2}E(Y_2) + \frac{1}{2}E(Y_3) + \frac{3}{2}E(Y_4) + \cdots + \frac{1}{2}E(Y_{n-1}) + \frac{3}{2}E(Y_n)) = \frac{1}{n}\mu_Y(1/2 + 3/2 + 1/2 + 3/2 + \cdots + 1/2 + 3/2) = \frac{n}{n}\mu_Y = \mu_Y$. Hence, $\tilde{Y}$ is unbiased.

$Var(\tilde{Y}) = \frac{1}{n^2}((\frac{1}{2})^2 Var(Y_1) + (\frac{3}{2})^2 Var(Y_2) + (\frac{1}{2})^2 Var(Y_3) + (\frac{3}{2})^2 Var(Y_4) + \cdots + (\frac{1}{2})^2 Var(Y_{n-1}) + (\frac{3}{2})^2 Var(Y_n)) = \frac{1}{n^2}\sigma_Y(1/4 + 9/4 + 1/4 + 9/4 + \cdots + 1/4 + 9/4) = \frac{1}{n^2}\frac{n}{2}(1/4 + 9/4)\sigma_Y = 1.25\frac{\sigma_Y}{n}$.

Since $Var(\tilde{Y}) \to 0$ as $n \to \infty$, $\tilde{Y}$ is consistent. $\tilde{Y}$ has a larger variance than $\bar{Y}$ and is therefore not as efficient.

Question.3 (16pts) Following Alfred Nobel's will, there are five Nobel Prizes awarded each year. These are for outstanding achievements in Chemistry, Physics, Physiology or Medicine, Literature, and Peace. In 1968, the Bank of Sweden added a prize in Economic Sciences in memory of Alfred Nobel. You think of the data as describing a population, rather than a sample from which you want to infer behavior of a larger population. The accompanying table lists the joint probability distribution between recipients in economics and the other five prizes, and the citizenship of the recipients, based on the 1969-2001 period.

Joint Distribution of Nobel Prize Winners in Economics and Non-Economics Disciplines, and Citizenship, 1969-2001

	U.S Citizen(Y=0)	Non-U.S Citizen (Y=1)	Total
Economics Novel Prize(X=0)	0.118	0.049	0.167
Non-Economics Disciplines(X=1)	0.345	0.488	0.833
Total	0.463	0.537	1.000

a. (3pts)Compute $E(Y)$ and interpret the resulting number.

$E(Y) = 0.537$. 53.7 percent of Nobel Prize winners were non-U.S. citizens.

b. (5 pts) Calculate and interpret $E(Y|X = 1)$ and $E(Y|X = 0)$.

$E(Y|X = 1) = 0.586$. 58.6 percent of Nobel Prize winners in non-economics disciplines were non-U.S. citizens. $E(Y|X = 0) = 0.293$. 29.3 percent of the Economics Nobel Prize winners were non-U.S. citizens.

c. (4pts) A randomly selected Nobel Prize winner reports that he is a non-U.S. citizen. What is the probability that this genius has won the Economics Nobel Prize? A Nobel Prize in the other five disciplines?

There is a 9.1 percent chance that he has won the Economics Nobel Prize, and a 90.9 percent chance that he has won a Nobel Prize in one of the other five disciplines.

---

d. (4pts) Are two variables independent? Explain.

They are not independent since $P(0,0)(=0.118) \neq P(x=0) \times P(y=0)(=0.167 \times 0.463)$.

Question 4. (16 pts) A regression of average hourly earnings (AHE, measured in dollars) on Age (measured in years) from Current Population Survey, estimate the OLS regression:

$$\widehat{AHE} = 3.324 + 0.452 \times Age, \quad R^2 = 0.08, SER = 8.7$$

a. (4pts) Explain what the coefficient values 3.324and 0.452 mean.

The coefficient 0.452 shows the marginal effect of Age on AHE; that is, AHE is expected to increase by \$0.452 for each additional year of age. 3.324 is the intercept of the regression line. It determines the overall level of the line.

---

b. (4pts) The standard error of the regression is 8.7. 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 AHW 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 \$8.7.

---

c. (4pts) The regression R^2 is 0.08. 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 model explains 8 percent of the variation in AHE.

d. (4pts) What is the regression's predicted earnings for a 26-year-old worker?

$\widehat{AHE} = 3.324 + 0.452 \times 26 = \$15.076$

Question. 5 (11pts)A manufacturer claims that a certain brand of VCR player has an average life expectancy of 5 years and 6 months with a standard deviation of 1 year. Assume that the life expectancy is normally distributed.

(a) (3pts)Selecting one VCR player from this brand at random, calculate the probability of its life expectancy exceeding 5 years.

Pr($Y > 5$)=Pr($Z > \frac{5-5.5}{1}$)=Pr($Z > -0.5$)=1-0.30853754=0.691

(b) (4pts) The Critical Consumer magazine decides to test fifty VCRs of this brand. The average life in this sample is 6 years and 6 month and the sample standard deviation is 2 years. Calculate a 95% confidence interval for the average life. Interpret the confidence interval.

$95\% CI = \bar{Y} \pm t_{0.025,49}\frac{2}{50} = 6.5 \pm 2.01 \times \frac{2}{\sqrt{50}} = [5.932, 7.068]$

We are 95% confident that the population mean lies between 5.932 and 7.068.

(c)(4pts) Calculate the t-value for the two side test of the null hypothesis $H_0 : \mu_Y = 6$. Do you reject the null hypothesis at the 5% level? At the 1% level?

$t = \frac{\bar{Y}-\mu_0}{s_Y/\sqrt{n}} = \frac{6.5-6}{2/\sqrt{50}} = 1.768$

Critical values are $t_{0.025,49} = 2.01$ (5% level) and $t_{0.005,49} = 2.68$ (1% level).

Since t is less than the critical values, we can not reject the null hypothesis at both 1% and 5% level.
