Econ 184b Professor Brainerd

Practice problems for the first midterm exam

Part I: Multiple Choice

1. Two random variables X and Y are independently distributed if all of the following conditions hold, with the exception of
   a. $\Pr(Y = y \mid X = x) = \Pr(Y = y)$.
   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. $\Pr(Y = y \mid X = x) = \Pr(X = x, Y = y)$
   e. None of the above (i.e., all of the above conditions hold).

2. An estimator $\hat{\mu}_Y$ of the population value μ is consistent if

   a. $\hat{\mu}_Y \xrightarrow{p} \mu_Y$

   b. its standard error is the smallest possible.

   c. Y is normally distributed.

   d. $\bar{Y} \xrightarrow{p} 0$

   e. $E(\hat{\mu}_Y) = \mu$

3. With i.i.d. sampling each of the following is true except

   a. $E(\bar{Y}) = \mu_Y$

   b. $\mathrm{Var}(\bar{Y}) = \sigma_Y^2/n$

   c. $E(\bar{Y}) > E(Y)$

   d. $\bar{Y}$ is a random variable.

   e. $\bar{Y}$ is normally distributed.

4. Which of the following statements is correct?
   a. $TSS = ESS + SSR$
   b. $ESS = SSR + TSS$
   c. $ESS > TSS$
   d. $R^2 = 1 - (ESS/TSS)$
   e. (a) and (d)

5. The OLS residuals, $\hat{u}_i$, are defined as follows:

a. $\hat{Y}_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$

b. $Y_i - \beta_0 - \beta_1 X_i$

c. $Y_i - \hat{Y}_i$

d. $(Y_i - \overline{Y})^2$

e. none of the above.

6. The slope estimator, β_1, has a smaller standard error, other things equal, if
   a. the residuals are homoscedastic.
   b. there is a large variance of the error term, u.
   c. the sample size is smaller.
   d. the intercept, β_0, is small.
   e. there is more variation in the explanatory variable, X.

7. Under the three least squares assumptions, the OLS estimator for the slope
   a. is BLUE.
   b. has a normal distribution even in small samples.
   c. is unbiased
   d. (a) and (c)
   e. (a), (b) and (c)

8. In the presence of heteroskedasticity, and assuming that the usual three least squares assumptions hold, the OLS estimator is
   a. efficient
   b. BLUE
   c. unbiased and consistent
   d. unbiased but not consistent
   e. none of the above.

<u>Part II: True/False/Uncertain</u> You must provide an explanation of your answer in order to receive any credit.

1. The sample average is a random variable and has a probability distribution that is the same as for the $Y_1,....,Y_n$ i.i.d. variables.

2. Assume that Y is normally distributed $N(\mu_Y, \sigma_Y^2)$. Moving from the mean (μ_Y) 1.64 standard deviations to the left and 1.64 standard deviations to the right, then the area under the normal p.d.f. is 0.05.

3. For a given sample, if we reject a null hypothesis at a 5% significance level, then we also reject the null hypothesis at a 10% significance level.

4. If the estimated coefficient on X_i is statistically significant, it means that X_i has an economically large impact on Y.

5. When the estimated slope coefficient ($\hat{\beta}_1$) in the simple regression model is zero, then $R^2 = 0$.

6. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X^2 + u_i$ is a linear regression.

Part III: Short answer problems

1. Suppose heights in a population are normally distributed with a mean of 68" and variance 25. What is the probability that a randomly selected individual will have a height between 58" and 78"?

2. Consider the joint probability distribution for discrete random variables X and Y:

	Y = 10	Y = 20	Y = 30
X = 10	0.2	0.1	0.1
X = 20	0.3	0.2	0.1

Calculate E(X | Y = 20).

3. Adult males are taller, on average, than adult females. Visiting two recent under-12 year old soccer matches on a Saturday, you do not observe an obvious difference in the height of boys and girls of that age. You suggest to your little sister that she collect data on height and gender of children in 4th to 6th grade as part of her science project. The accompanying table shows her findings.

Height of Young Boys and Girls, Grades 4-6, in inches

Boys			Girls		
$\overline{Y}_{Boys}$	s_{Boys}	n_{Boys}	$\overline{Y}_{Girls}$	s_{Girls}	n_{Girls}
57.8	3.9	55	58.4	4.2	57

a. Let your null hypothesis be that there is no difference in the height of females and males at this age level. Specify the alternative hypothesis.

b. Find the difference in height and the standard error of the difference.

c. Generate a 95% confidence interval for the difference in height.

d. Calculate the *t*-statistic for comparing the two means. Is the difference statistically significant at the 1% level? Which critical value did you use? Why would this number be smaller if you had assumed a one-sided alternative hypothesis? What is the intuition behind this?

4. A regression of income at age 40 on an individual's SAT scores yields the following 95% confidence interval for the estimated regression slope coefficient for the individual's SAT scores:

$$\beta_{SAT} = (0.432, 5.442)$$

a. Construct and carry out a 2-sided hypothesis test for the null hypothesis of no effect of SAT scores on income at age 40. Make sure to clearly state the null and alternative hypotheses. Can we reject the null?

b. From the information above calculate $\hat{\beta}_{SAT}$ and $SE(\hat{\beta}_{SAT})$.

c. Construct the confidence interval for the predicted effects on income at age 40 of increasing SAT scores by 2 points.

5. Sir Francis Galton, a cousin of James Darwin, examined the relationship between the height of children and their parents towards the end of the 19th century. It is from this study that the name "regression" originated. You decide to update his findings by collecting data from 110 college students, and estimate the following relationship:

$$\widehat{Studenth} = \underset{(7.2)}{19.6} + \underset{(0.10)}{0.73}\times Midparh,\ R^2 = 0.45,\ SER = 2.0$$

where *Studenth* is the height of students in inches, and *Midparh* is the average of the parental heights. Values in parentheses are heteroskedasticity robust standard errors. (Following Galton's methodology, both variables were adjusted so that the average female height was equal to the average male height.)

a. Interpret the estimated coefficients.

b. What is the prediction for the height of a child whose parents have an average height of 70.06 inches?

c. What is the interpretation of the SER here?

d. Given the positive intercept and the fact that the slope lies between zero and one, what can you say about the height of students who have quite tall parents? Who have quite short parents?

e. If children, on average, were expected to be of the same height as their parents, then this would imply two hypotheses, one for the slope and one for the intercept.

(i) What should the null hypothesis be for the intercept? Calculate the relevant *t*-statistic and carry out the hypothesis test at the 1% level.
(ii) What should the null hypothesis be for the slope? Calculate the relevant *t*-statistic and carry out the hypothesis test at the 5% level.

6. You recall from one of your earlier lectures in macroeconomics that the per capita income depends on the savings rate of the country: those who save more end up with a higher standard of living. To test this theory, you collect data from the Penn World Tables on GDP per worker relative to the United States (RelProd) in 1990 and the average investment share of GDP from 1980-1990 (sK), remembering that investment equals saving. The regression results in the following output:

$$\widehat{RelProd} = \underset{(0.04)}{0.08} + \underset{(0.38)}{2.44}\times s_K\,,\ R^2=0.46,\ SER = 0.21$$

a. Interpret the regression results carefully, including the R2.

b. Calculate the t-statistics to determine whether the two coefficients are significantly different from zero.

c. You accidentally forget to use the heteroskedasticity-robust standard errors option in your regression package and estimate the equation using homoskedasticity-only standard errors. This changes the results as follows:

$$\widehat{RelProd} = \underset{(0.04)}{0.08} + \underset{(0.26)}{2.44}\times s_K\,,\ R^2=0.46,\ SER = 0.21$$

You are delighted to find that the coefficients have not changed at all and that your results have become even more significant. Why haven't the coefficients changed? Are the results really more significant? Explain.

d. Upon reflection you think about the advantages of OLS with and without homoskedasticity-only standard errors. What are these advantages? Is it likely that the error terms would be heteroskedastic in this situation?