Econ 184b Professor Brainerd

Solutions to the first midterm practice problems

Part I: Multiple choice

1. d
2. a
3. c
4. a
5. c
6. e
7. c
8. c

Part II: True/False/Uncertain

1. False, the variance of the sample average is the variance of the i.i.d. variables divided by n, so the probability distributions are not the same.

2. False, Pr(-1.64 < Z < 1.64) =.90, so the area under the normal is .90.

3. True. We need more confidence that the null is false to reject the null at a 5% significance level. For a t-test, the critical value associated with a 5% significance level is higher than that associated with a 10% significance level. So, any time that we reject at the 5% level we would also reject at the 10% level.

4. False. Statistical significance merely indicates that a coefficient is statistically different from 0, in other words statistical significance indicates that the coefficient would be unlikely to arise due to sampling error. It does not indicate anything about the economic significance of the variable; economic significance indicates the coefficient is large enough to be economically important.

5. True. If the slope coefficient is 0, Y and X are uncorrelated. Therefore none of the variation in Y is "explained" by X, so ESS = 0. Since R^2 = ESS/TSS, R^2 will be 0.

Alternatively, one can show this mathematically:

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ but $\hat{\beta}_1 = 0$ so $\hat{Y}_i = \hat{\beta}_0$.

$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = \overline{Y} - 0 = \overline{Y}$ so $\hat{Y}_i = \overline{Y}$

Since $R^2 = \dfrac{\sum (\hat{Y}_i - \overline{Y})^2}{\sum (Y_i - \overline{Y})^2}$, the numerator will be 0.

6. True. Although it is not linear in the variables, the regression is linear in the parameters, which is the requirement for a linear regression.

Part III: Short answer problems

1. Let Y be the random variable denoting height. Then Y~N(68, 25), so we want to find Pr(58 < Z < 78). Standardizing so we can use the standard normal probabilities, we have Pr[(58-68)/5 < Z < (78-68)/5] which is Pr(-2 < Z < 2) = Pr(Z<2) - Pr(Z < -2) = 0.9772-0.0228 = 0.9544.

2. Find Pr(Y) and Pr(X):

	Y = 10	Y = 20	Y = 30	Pr(X)
X = 10	0.2	0.1	0.1	0.4
X = 20	0.3	0.2	0.1	0.6
Pr(Y)	0.5	0.3	0.2	1.0

To find the conditional expectation of X if Y is 20:

E(X | Y = 20) = (10)Pr(X=10 | Y=20) + (20)Pr(X = 20 | Y = 20)

$= (10)\frac{\Pr(X=10,Y=20)}{\Pr(Y=20)} + (20)\frac{\Pr(X=20\,,Y=20)}{\Pr(Y=20)} = (10)\frac{0.10}{0.30} + (20)\frac{0.20}{0.30}$

= 3.33 + 13.33 = 16.667

3. a. $H_0 : \mu_{Boys} - \mu_{Girls} = 0$ vs. $H_1 : \mu_{Boys} - \mu_{Girls} \neq 0$

b. $\overline{Y}_{Boys} - \overline{Y}_{Girls}$ = -0.6, SE($\overline{Y}_{Boys} - \overline{Y}_{Girls}$) = $\sqrt{\frac{3.9^2}{55} + \frac{4.2^2}{57}}$ = 0.77.

c. $-0.6 \pm 1.96 \times 0.77$ = (-2.11, 0.91).

d. t = -0.78, so $|t| < 2.58$, which is the critical value at the 1% level. Hence you cannot reject the null hypothesis. The critical value for the one-sided hypothesis would have been 2.33. Assuming a one-sided hypothesis implies that you have some information about the problem at hand, and, as a result, can be more easily convinced than if you had no prior expectation.

4. a. H_0: $\beta_{SAT} = 0$ and H_1: $\beta_{SAT} \neq 0$. The 95 percent confidence interval is the set of values that cannot be rejected using a two-sided hypothesis test with 5 percent significance level. The null is rejected since it does not contain zero.

b. $$5.442 = \hat{\beta}_{SAT} + 1.96 * SE(\hat{\beta}_{SAT}) \qquad (1)$$

$$0.432 = \hat{\beta}_{SAT} - 1.96 * SE(\hat{\beta}_{SAT}) \qquad (2)$$

Subtract (2) from (1) to get:

$$5.01 = 3.92\ SE(\hat{\beta}_{SAT})$$

$$SE(\hat{\beta}_{SAT}) = 1.278$$

$$5.442 = \hat{\beta}_{SAT} + 1.96 * 1.78$$

$$\hat{\beta}_{SAT} = 2.937$$

c. Increasing the SAT by 2 points is predicted to increase income by between (0.432*2, 5.442*2) = (0.864, 10.884) with 95 percent confidence.

5. a. The coefficient on Midparh means that, for every one-inch increase in the height of their parents, a student's height increases by .73 of an inch. The intercept has no reasonable interpretation, as it gives the height of a student whose parents have an average height of zero inches.

b. A child whose parents have an average height of 70.06 inches will have a predicted height of:

$$19.6 + 0.73 \times 70.06 = 70.74 \text{ inches}$$

c. 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 two inches.

d. Tall parents will have, on average, tall students, but they will not be as tall as their parents. Short parents will have short students, although on average, they will be somewhat taller than their parents.

e. (i) If students are expected to be the same height as their parents, this implies that the regression line is simply Y = X, i.e. the intercept is 0 and the slope coefficient is 1. The null hypothesis for the intercept is: H_0: $\beta_0 = 0$. The t-statistic is:

$$\frac{19.6}{7.2} = 2.72$$

For $H_1 : \beta_0 \neq 0$, the critical value for a two-sided alternative is 2.58. So we reject the null hypothesis in (i).

(ii) For the slope, the null hypothesis is H_0: $\beta_1 = 1$ and the alternative hypothesis is H_1: $\beta_1 \neq 1$. The t-statistic is:

$$\frac{0.73 - 1}{0.10} = -2.70$$

The critical value for a two-sided alternative is -1.96, so we reject the null hypothesis in (ii).

6. a. An increase in the saving rate of 0.1 results in an increase in relative GDP per worker of 0.244. There is no interpretation for the intercept. The regression explains 46 percent of the variation in GDP per worker relative to the United States.

b. The t-statistics are 2.00 and 6.42 for the intercept and slope, respectively. (You should use a two-sided test for the intercept, since there are no prior expectations on whether it should be positive or negative.) The intercept is statistically significant at the 5% level, but not at the 1% levels; the critical values suggest significance at any reasonable probability level of the size of the test for the slope. Since we expect a positive sign on the slope, one could conduct a one-sided test here (a two-sided test is fine too; just be clear on which one you are using).

c. Whether you use homoskedasticity-only or heteroskedasticity-robust standard errors does not affect the estimator, only the formula for the standard errors. If the assumption of homoskedasticity was valid, then the results would be more significant. However, given that homoskedastic errors are unusual in economic data, it is safer to conduct inference under the assumption of heteroskedasticity.

d. In the presence of homoskedasticity in addition to the usual three least squares assumptions, OLS is BLUE (the Gauss-Markov theorem). This means that OLS is the most efficient estimator of all unbiased linear estimators. Homoskedasticity-only standard errors are typically smaller than heteroskedasticity-robust standard errors, which results in t-statistics that are too large, and hence rejection of the null hypothesis too often, if the errors are in fact heteroskedastic. Economic theory suggests that the error terms would be heteroskedastic in this situation; there are virtually no examples where economic theory suggests that the errors are homoskedastic.