Disclaimer: this is not intended to be a comprehensive review – I would have to come up with like 100 practice problems to do so. Instead, think of this as a simply a quasi-random selection of problems that you can use to gauge your general level of readiness for the exam. I may or may not add questions over the course of the next several days, depending on what the lectures leading up to the midterm cover.

Problem 1. Provide the general definition of a Type I error.

Problem 2. Suppose y has sample variance 1, x has sample variance 4, and the sample covariance between x and y is 1. What is the sample correlation coefficient?

Problem 3. You regress the z-score for variable y on the z-score for variable x and obtain a slope estimate of 0.6. Provide a simple interpretation of this slope coefficient estimate.

Problem 4. Provide the four population assumptions used for the linear regression model. Which of these assumptions are required for OLS estimates to be unbiased?

Problem 5. The OLS estimator

(a) minimizes the sum of vertical deviations of actual data points from the regression line.

(b) minimizes the sum of horizontal deviations of actual data points from the regression line.

(c) minimizes the sum of squared vertical deviations of actual data points from the regression line.

(d) minimizes the sum of squared horizontal deviations of actual data points from the regression line.

(e) none of the above.

Problem 6. The standard error of the regression is a measure of

(a) the standard deviation of the slope coefficient.

(b) the standard deviation of the intercept coefficient.

(c) the standard deviation of the dependent variable.

(d) the standard deviation of the residual.

(e) none of the above.

Problem 7. Regression of y on x yields slope coefficient $b_2 = 0.50$ and correlation coefficient $r_{xy} = 0.40$. It follows that regression of x on y using the same data yields

(a) slope coefficient 2.0.

(b) correlation coefficient 0.40.

(c) both (a) and (b).

(d) neither (a) nor (b).

Problem 8

You are given the following information following regression of y on an intercept and x:

- explained sum of squares = 40
- total sum of squares = 160
- number of observations = 10

Part a. Give the R^2 for this regression.

Part b. Give the correlation coefficient between x and y.

Part c. Give the standard error of the residual for this regression.

Problem 9

This question uses data for hospitals in New York state in 2011. `meancost` is the average cost of knee replacements and `meancharge` is the average charge for knee replacements.

```
. summarize meancharge meancost
```

Variable	Obs	Mean	Std. Dev.	Min	Max
meancharge	169	47957.27	22248.97	14953	123131
meancost	169	21007.51	10376.75	7021	86730

```
. regress meancharge meancost
```

Source	SS	df	MS
Model	3.1277e+10	1	3.1277e+10
Residual	5.1886e+10	167	310694328
Total	8.3163e+10	168	495016535

```
Number of obs =     169
F(  1,   167) =  100.67
Prob > F      =  0.0000
R-squared     =  0.3761
Adj R-squared =  0.3724
Root MSE      =   17627
```

meancharge	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
meancost	1.314908	.1310541	10.03	0.000	1.056171	1.573644
_cons	20334.33	3068.893	6.63	0.000	14275.5	26393.15

```
.
. display "t_.005;167 = " invttail(167,.005) "   t_.01;167 = " invttail(167,.01) ///
>       _n "t_.025;167 = " invttail(167,.025) "   t_.05;167 = " invttail(167,.05) ///
>         "  t_.10;167 = " invttail(167,.10)
t_.005;167 = 2.6055891   t_.01;167 = 2.3488842
t_.025;167 = 1.974271   t_.05;167 = 1.6540291  t_.10;167 = 1.2866415
```

Part a. How does the mean charge change when the mean cost increases by one thousand dollars?

Part b. Give a 95 percent confidence interval for the population slope coefficient.

Part c. Give a 99 percent confidence interval for the population slope coefficient.

Part d. The claim is made that the mean charge is not associated with the mean cost. Test this claim at significance level 0.05. State clearly the null and alternative hypotheses and your conclusion.

Part e. The claim is made that mean charge increases with the mean cost. Test this claim at significance level 0.05. State clearly the null and alternative hypotheses and your conclusion.

Part f. Give the predicted mean hospital charge when the mean cost is $20,000.

Part g. Suppose we regressed `meancharge` on only an intercept. What value do you expect for the intercept coefficient?