

## Unit 1 Quiz

You have 1 hour and 15 minutes to complete this quiz. There are 47 total points.

1. (16 points) What are the five Gauss Markov assumptions for multiple linear regression? First, state each assumption. Then, for each assumption, briefly describe a real-world situation in which it would be violated. One assumption is filled in for you as an example.

#	Assumption [2 points each]	Violation [2 points each]
1	<i>The population model is linear in parameters</i>	<i>Estimating a non-linear function, like a Cobb-Douglas function (<math>Y = AL^\alpha K^{1-\alpha}</math>)</i>
2		
3		
4		
5		

2. [4 points] Assuming all five Gauss-Markov assumptions hold is sufficient to prove which of the following statements? Circle the appropriate answer(s), and note that multiple answers may be correct. No justification is needed.

- (a) OLS estimates are unbiased
- (b) OLS is BLUE
- (c)  $\hat{\sigma}^2 = \frac{SSR}{n-k-1}$  is an unbiased estimator of  $\sigma^2$
- (d)  $\frac{\hat{\beta}_j - \beta}{se(\hat{\beta}_j)} \sim t_{n-k-1}$

3. *[27 points]* Suppose you are interested in the effect of attendance (**attend**) on final exam scores (**final**, out of 40 points), controlling for students' past GPA (**priGPA**). A description of each variable is on the last page of the quiz. You use OLS to predict **final** and you get the following results:

```
. reg final attend priGPA
```

Source	SS	df	MS	Number of obs	=	680
Model	2021.72415	2	1010.86207	F(2, 677)	=	52.48
Residual	13040.2229	677	19.2617768	Prob > F	=	0.0000
				R-squared	=	0.1342
				Adj R-squared	=	0.1317
Total	15061.9471	679	22.1825435	Root MSE	=	4.3888

  

final	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
attend	-.0172012	.0341483	-0.50	0.615	-.0842504 .0498481
priGPA	3.237554	.3419779	9.47	0.000	2.56609 3.909019
_cons	17.96611	.9660608	18.60	0.000	16.06928 19.86295

```
.
.
```

- (a) List the dependent variable(s) and independent variable(s). *[2 points]*

- (b) What is the population model that this regression estimates? *[2 points]*

- (c) What is the estimating equation? *[2 points]*

- (d) Interpret  $\hat{\beta}_{attend}$ . That is, what does  $-0.0172$  mean? (Note that it is not sufficient to only say that there is a positive or negative relationship). *[2 points]*

- (e) Fill in the following table with the corresponding estimates: [3 points]

$\hat{\beta}_0$		$se(\hat{\beta}_{priGPA})$	
$R^2$		degrees of freedom	
$SSE$		$SSR$	

- (f) Suppose you want to test whether  $\beta_{attend}$  is statistically significant from zero at the 5% level. Conduct a hypothesis test by filling in the table below. You do not need to show your work. [3 points]

Null hypothesis:	
Alternative hypothesis:	
Test statistic(s):	
P-value:	
Decision:	

- (g) Suppose Rogetta decides that since she wants to know about the effect of missing classes, she will add **missed**, the number of classes missed, to the regression in addition to the previous independent variables. Explain why this is a bad idea. [3 points]

- (h) Suppose that you estimate the previous model, but now you control for student's ACT scores (ACT). You get the following results:

```
. reg final attend priGPA ACT
```

Source	SS	df	MS	Number of obs	=	680
Model	3032.09408	3	1010.69803	F(3, 676)	=	56.79
Residual	12029.853	676	17.7956405	Prob > F	=	0.0000
				R-squared	=	0.2013
Total	15061.9471	679	22.1825435	Adj R-squared	=	0.1978
				Root MSE	=	4.2185

  

final	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
attend	.0793386	.0352349	2.25	0.025	.0101556 .1485216
priGPA	1.915294	.372614	5.14	0.000	1.183674 2.646914
ACT	.4010639	.0532268	7.54	0.000	.2965542 .5055736
_cons	9.834203	1.423709	6.91	0.000	7.038779 12.62963

- i. Based on this and the previous result (and assuming no correlation between ACT score and prior GPA), what is the correlation between ACT score and attendance?

*[3 points]*

- ii. After including ACT score in the model, is it likely that omitted variable bias persists? If yes, provide one example of a variable that would be likely to contribute to this bias and explain why. If no, explain your reasoning.

*[3 points]*

- (i) Suppose you instead predict the *log* of final exam scores, `lfinal`. You now obtain the following estimates:

```
. reg lfinal attend priGPA ACT
```

Source	SS	df	MS	Number of obs	=	680
Model	4.44557174	3	1.48185725	F(3, 676)	=	50.10
Residual	19.9956331	676	.029579339	Prob > F	=	0.0000
				R-squared	=	0.1819
Total	24.4412048	679	.035995883	Adj R-squared	=	0.1783
				Root MSE	=	.17199

  

lfinal	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
attend	.0027021	.0014365	1.88	0.060	-.0001185	.0055226
priGPA	.0751518	.0151914	4.95	0.000	.0453239	.1049797
ACT	.0152682	.00217	7.04	0.000	.0110074	.0195291
_cons	2.627812	.0580442	45.27	0.000	2.513844	2.741781

- i. Interpret  $\hat{\beta}_{attend}$ . That is, what does 0.0027 mean? (Note that it is not sufficient to only say that there is a positive or negative relationship). [2 points]
- ii. Of the three models estimated in this quiz, which model best fits the data? Explain why you believe it is the best fit. [2 points]

```

.
.
. describe
Contains data from attend.dta
  obs:          680
  vars:          9                29 Jul 2005 21:00
  size:        16,320            (_dta has notes)

```

---

variable name	storage type	display format	value label	variable label
final	byte	%8.0g		final exam score (out of 40)
attend	byte	%8.0g		classes attended out of 32
missed	byte	%9.0g		number of classes missed
priGPA	float	%9.0g		cumulative GPA prior to term
priGPA_sq	float	%9.0g		cumulative GPA prior to term, squared
ACT	byte	%8.0g		ACT score
ACT_sq	float	%9.0g		ACT score, squared
lpriGPA	float	%9.0g		log of cumulative GPA prior to term
lfinal	float	%9.0g		log of final exam score

---

Sorted by:  
 Note: Dataset has changed since last saved.

```

.
.
. summarize

```

Variable	Obs	Mean	Std. Dev.	Min	Max
final	680	25.89118	4.709835	10	39
attend	680	26.14706	5.455037	2	32
missed	680	5.852941	5.455037	0	30
priGPA	680	2.586775	.5447141	.857	3.93
priGPA_sq	680	6.987682	2.892395	.734449	15.4449
ACT	680	22.51029	3.490768	13	32
ACT_sq	680	518.8809	158.8418	169	1024
lpriGPA	680	.9272564	.2192617	-.1543174	1.368639
lfinal	680	3.236557	.1897258	2.302585	3.663562