

Exercise 5.1

A survey of a random sample of students at the University of New Hampshire was conducted. We are interested in predictors of grade point average (GPA), which is measured on a 4-point scale.

We are interested in the relationship between GPA and alcohol consumption. Alcohol consumption is captured in the variable `drink`, where a higher number means more drinking. (Note: this drinking scale goes from 0=no drinking, to 33=max drinking score). We are curious as to whether the relationship between GPA and drinking is different for men and women. Thus, a regression model was fit using drinking score, sex (gender; 1=male, 0=female), and their interaction to predict GPA. Use the Stata output provided at the end of the problem to answer the questions below.

(a) What is the estimated intercept? Write a one-sentence interpretation of this quantity.

(b) What is the estimated coefficient for the main effect of `drink`? Write a one-sentence interpretation of this quantity.

(c) What is the estimated coefficient for the main effect of `gender`? Write a one-sentence interpretation of this quantity.

(d) What is the estimated slope of drinking score for females? For males?

(e) Among students who have a drinking score of 19 (the sample mean drinking score), what is the estimated difference in GPA between men and women? Indicate which group has a higher estimated GPA (men or women).

(f) Is there evidence that the relationship between GPA and drinking score is different for males and females? Cite specific evidence from the output. (Assume $\alpha = 0.05$.)

```
. generate drink_gender = drink*gender
```

```
. regress gpa drink gender drink_gender
```

Source		SS	df	MS	Number of obs	=	218
-----+-----					F(3, 214)	=	7.23
Model		4.20817373	3	1.40272458	Prob > F	=	0.0001
Residual		41.5435654	214	.19412881	R-squared	=	0.0920
-----+-----					Adj R-squared	=	0.0792
Total		45.7517391	217	.210837507	Root MSE	=	.4406

gpa		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
drink		-.020912	.0065085	-3.21	0.002	-.033741 -.008083
gender		-.243004	.1884557	-1.29	0.199	-.6144711 .1284631
drink_gender		.0078177	.0093203	0.84	0.403	-.0105538 .0261891
_cons		3.237566	.1175289	27.55	0.000	3.005904 3.469229

Exercise 5.2

Data on standardized tests were recorded for 200 students. We are interested in predictors of writing score (`write`; higher is better; range is 31 to 67 in the dataset). In particular, we want to know whether sex (`female`; 1=female, 0=male) and social studies score (`socst`; higher is better; range is 26 to 71 in the dataset) are associated with writing score. We also think that the effect of social studies score on writing score might depend on sex. Thus, a regression model was fit using sex, social studies score, and their interaction to predict writing score. Use the Stata output provided at the end of the problem to answer the questions below.

(a) For females, what is the estimated regression line for social studies score predicting writing score?

(a) For males, what is the estimated regression line for social studies score predicting writing score?

(c) Is there a significant difference in the effect of social studies score on writing score for females versus males? Cite specific evidence from the output. (Assume $\alpha = 0.05$.)

(d) Is there a significant effect of social studies score on writing score for males? Cite specific evidence from the output. (Assume $\alpha = 0.05$.)

(e) Is there a significant effect of social studies score on writing score for females? Cite specific evidence from the output. (Assume $\alpha = 0.05$.)

(f) What is the estimated difference in mean writing score for females versus males, among students with a social studies score of 52 (approximately the sample mean)? Indicate which group has a higher estimated GPA (females or males).

(g) How do you interpret the intercept estimate in this model?

```
. generate female_socst = female*socst
```

```
. regress write female socst female_socst
```

Source	SS	df	MS	Number of obs	=	200
-----+-----				F(3, 196)	=	49.26
Model	7685.43528	3	2561.81176	Prob > F	=	0.0000
Residual	10193.4397	196	52.0073455	R-squared	=	0.4299
-----+-----				Adj R-squared	=	0.4211
Total	17878.875	199	89.843593	Root MSE	=	7.2116

write	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
female	15.00001	5.09795	2.94	0.004	4.946132	25.05389
socst	.6247968	.0670709	9.32	0.000	.4925236	.7570701
female_socst	-.2047288	.0953726	-2.15	0.033	-.3928171	-.0166405
_cons	17.7619	3.554993	5.00	0.000	10.75095	24.77284

```
. lincom socst + female_socst
```

```
( 1) socst + female_socst = 0
```

write	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
(1)	.420068	.0678044	6.20	0.000	.2863482	.5537878

Exercise 5.3

A survey of a random sample of students at the University of New Hampshire was conducted. We are interested in predictors of grade point average (GPA), which is measured on a 4-point scale.

For this question, we are interested in whether a student has a job or not impacts GPA (`job`: 1=has job, 0=does not). We are also interested in whether being in a fraternity/sorority (`frat`: 1=in fraternity/sorority, 0=not) might modify this effect. Thus, a regression model was fit using having a job, being in a fraternity/sorority, and their interaction was used to predict GPA. Use the Stata output provided at the end of the problem to answer the questions below.

- (a) What is the estimated intercept? Write a one-sentence interpretation of this quantity.

- (b) What is the estimated coefficient for the main effect of `job`? Write a one-sentence interpretation of this quantity.

- (c) What is the estimated coefficient for the main effect of `frat`? Write a one-sentence interpretation of this quantity.

(d) Calculate the estimated mean GPA for each group of students (e.g., “no job, not in fraternity/sorority”, “has job, not in fraternity/sorority”, etc.).

(e) What is the estimated difference in GPA for students who work compared to students who do not work, for students not in a fraternity/sorority?

(f) What is the estimated difference in GPA for students who work compared to students who do not work, for students who are in a fraternity/sorority?

(g) Is there evidence that the effect of having a job on GPA is significantly different for students who are in a fraternity/sorority compared to students not in a fraternity/sorority? Cite specific evidence from the output. (Assume $\alpha = 0.05$.)


```
. generate job_frat = job*frat
```

```
. regress gpa job frat job_frat
```

Source		SS	df	MS	Number of obs	=	216
-----+-----					F(3, 212)	=	2.89
Model		1.78161624	3	.59387208	Prob > F	=	0.0363
Residual		43.5312053	212	.205335874	R-squared	=	0.0393
-----+-----					Adj R-squared	=	0.0257
Total		45.3128216	215	.21075731	Root MSE	=	.45314

gpa		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
job		.0579913	.0713873	0.81	0.418	-.0827285 .1987111
frat		-.2200616	.1066417	-2.06	0.040	-.4302755 -.0098477
job_frat		.2746088	.153546	1.79	0.075	-.0280637 .5772812
_cons		2.794462	.0562051	49.72	0.000	2.683669 2.905254

Exercise 5.4

Data on standardized tests were recorded for 200 students. Researchers are interested in the relationship between reading score (`read`) and both math score (`math`) and social studies score (`socst`). For these scores, higher is better (summary statistics shown in Stata output). A regression model was fit using math score, social studies score, and their interaction to predict reading score. Use the Stata output provided at the end of the problem to answer the questions below.

- (a) What is the estimated intercept? Write a one-sentence interpretation of this quantity.

- (b) What is the estimated coefficient for the main effect of `math`? Write a one-sentence interpretation of this quantity.

- (c) What is the estimated coefficient for the main effect of `socst`? Write a one-sentence interpretation of this quantity.

(d) Does the association between math score and reading score significantly depend on social studies score? Cite specific evidence from the output. (Assume $\alpha = 0.05$.)

(e) What is the estimated slope of math score for students whose social studies score is 52 (approximately the sample mean)?

(f) What is the estimated slope of math score for students whose social studies score is 62 (approximately 1 SD above the sample mean)?

(g) Explain how the effect of math score on reading score is impacted by social studies score. (Using your answers to parts (e) and (f) may be helpful.)

(h) What about the effect of social studies on reading score? How is it affected by math score?

```
. summarize read math socst
```

Variable	Obs	Mean	Std. Dev.	Min	Max
read	200	52.23	10.25294	28	76
math	200	52.645	9.368448	33	75
socst	200	52.405	10.73579	26	71

```
. generate math_socst = math*socst
```

```
. regress read math socst math_socst
```

Source	SS	df	MS	Number of obs	=	200
Model	11424.7622	3	3808.25406	F(3, 196)	=	78.61
Residual	9494.65783	196	48.4421318	Prob > F	=	0.0000
Total	20919.42	199	105.122714	R-squared	=	0.5461
				Adj R-squared	=	0.5392
				Root MSE	=	6.96

read	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
math	-.1105123	.2916338	-0.38	0.705	-.6856552	.4646307
socst	-.2200442	.2717539	-0.81	0.419	-.7559812	.3158928
math_socst	.0112807	.0052294	2.16	0.032	.0009677	.0215938
_cons	37.84271	14.54521	2.60	0.010	9.157506	66.52792

Exercise 5.5

The model from the previous exercise was rerun, centering the predictors by subtracting their approximate sample means. Use the Stata output provided at the end of the problem to answer the questions below.

- (a) What is the estimated intercept? Write a one-sentence interpretation of this quantity.

- (b) What is the estimated coefficient for the main effect of `math_52`? Write a one-sentence interpretation of this quantity.

- (c) What is the estimated coefficient for the main effect of `socst_52`? Write a one-sentence interpretation of this quantity.

```

. generate math_52 = math - 52
. generate socst_52 = socst - 52
. generate math_52_socst_52 = math_52*socst_52

. regress read math_52 socst_52 math_52_socst_52

```

Source		SS	df	MS	Number of obs	=	200
-----+-----					F(3, 196)	=	78.61
Model		11424.7622	3	3808.25406	Prob > F	=	0.0000
Residual		9494.65783	196	48.4421318	R-squared	=	0.5461
-----+-----					Adj R-squared	=	0.5392
Total		20919.42	199	105.122714	Root MSE	=	6.96

	read		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----							
	math_52		.4760853	.0640923	7.43	0.000	.3496862 .6024843
	socst_52		.3665534	.055092	6.65	0.000	.2579042 .4752025
math_52_socst_52		.0112807	.0052294	2.16	0.032	.0009677	.0215938
	_cons		51.15685	.5674107	90.16	0.000	50.03784 52.27587
