

Exercise 4.1

We wish to test the relative effects of three drugs (numbered 1, 2, 3) on the reduction of fever. We randomly assign 5 children with a fever to each of the three drugs. Reduction in fever (in degrees) is noted 4 hours after administration of drug. (Note: a positive value for the variable reduction means the fever went down.) Our interest is in predicting reduction in fever with drug type. Stata code is run and is shown below. The model fit is: $E(REDUCTION) = \beta_0 + \beta_1 DRUG2 + \beta_2 DRUG3$

```
. generate drug1 = (drug==1) if !missing(drug)
. generate drug2 = (drug==2) if !missing(drug)
. generate drug3 = (drug==3) if !missing(drug)

. regress reduction drug2 drug3
```

Source	SS	df	MS	Number of obs	=	15
Model	5.82933309	2	2.91466655	F(2, 12)	=	6.79
Residual	5.14800001	12	.429000001	Prob > F	=	0.0106
Total	10.9773331	14	.784095222	R-squared	=	0.5310
				Adj R-squared	=	0.4529
				Root MSE	=	.65498

reduction	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
drug2	-1.16	.4142463	-2.80	0.016	-2.062565 -.2574348
drug3	-1.44	.4142463	-3.48	0.005	-2.342565 -.5374348
_cons	1.52	.2929164	5.19	0.000	.88179 2.15821

- (a) What is the correct interpretation of the estimated intercept in this model?

- (b) Is the test of the intercept meaningful in this model? If yes, write a one sentence interpretation of the test result. If not, explain why it is not meaningful.
- (c) Write a one sentence interpretation of the estimated coefficient for drug2.
- (d) In terms of the estimated regression coefficients $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$, what is the estimated mean difference in fever reduction for Drug 3 minus Drug 2? Plugging in the values of these coefficients, what is this difference?
- (e) Calculate the estimated mean fever reduction for Drug 2 and for Drug 3 and use these means to confirm your answer to (c).

(f) Can we tell from the given Stata output whether the true mean fever reduction for Drug 2 is significantly different from the true mean fever reduction for Drug 3? If yes, are they significantly different? If not, what additional Stata commands would you have to run to get the answer?

(g) Do we need a partial F-test in order to test whether drug type is a significant predictor of fever reduction? If not, what information on the Stata output can we use for this test?

(h) If I rerun the model, but make Drug 3 the reference group, which of the following will change and which will stay the same: Overall F-test, R^2 , Adjusted R^2 , Intercept estimate, Coefficient for drug2?

(i) Suppose we also have the weight of the children (lbs), and add this to our regression model, so that the model we are fitting is:

$$E(REDUCTION) = \beta_0 + \beta_1 DRUG2 + \beta_2 DRUG3 + \beta_3 WEIGHT$$

Interpret each of the parameters in this new model.

Exercise 4.2

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.

One variable we are interested in is `religion`, which captures a student's religious preference (1=Protestant, 2=Catholic, 3=Other). We also have as predictors number of hours studied per week (`study`), age (years), and sex (`gender`; 1=male, 0=female).

A regression model was fit to predict GPA using these variables. Use the Stata output provided at the end of the problem (note: output spans 2 pages) to answer the questions below.

- (a) What category is the reference group for the `religion` variable?

- (b) Interpret the coefficient for the dummy variable `religion1`.

- (c) Interpret the coefficient for the dummy variable `religion2`.

- (d) Interpret the coefficient for `study`. Is there a significant effect of hours studied on GPA?

- (e) What is the estimated mean difference in GPA between Protestant and Catholic students? Be sure to indicate which group has the higher GPA.
- (f) Is there a significant effect of religion on GPA (controlling for hours studied, age, and gender)? Cite a p-value in your answer.
- (g) Is there a significant difference in mean GPA between Protestant students and students of “Other” religion? Cite a p-value in your answer.
- (h) Is there a significant difference in mean GPA between Protestant students and Catholic students? Cite a p-value in your answer.
- (i) Explain how the degrees of freedom for the partial F-test that was conducted are calculated.

```

. generate religion1 = (religion==1) if !missing(religion)
. generate religion2 = (religion==2) if !missing(religion)
. generate religion3 = (religion==3) if !missing(religion)

. regress gpa religion1 religion2 study age gender

      Source |          SS            df           MS       Number of obs   =
-----+----- F(5, 206)        =     8.91
      Model |  7.71521219          5   1.54304244  Prob > F      =  0.0000
      Residual | 35.6880173         206   .173242802 R-squared      =  0.1778
-----+----- Adj R-squared      =  0.1578
      Total | 43.4032295         211   .205702509 Root MSE       =  .41622

-----+
      gpa |      Coef.    Std. Err.      t    P>|t| [95% Conf. Interval]
-----+
  religion1 |  -.0659408   .0818991   -0.81   0.422   -.2274086   .0955271
  religion2 |  -.1738371   .0647677   -2.68   0.008   -.3015297  -.0461445
      study |   .0123059   .0033299    3.70   0.000    .0057409   .018871
        age |   .0375648   .0095926    3.92   0.000    .0186526   .0564771
     gender |  -.1366602   .0584907   -2.34   0.020   -.2519773  -.0213431
      _cons |    2.00595   .2163566    9.27   0.000    1.579393  2.432507
-----+
  . test religion1 religion2

  ( 1) religion1 = 0
  ( 2) religion2 = 0

  F(  2,    206) =     3.71
  Prob > F =  0.0260

```

```
. lincom religion1 - religion2  
  
( 1) religion1 - religion2 = 0  
  
-----  
gpa | Coef. Std. Err. t P>|t| [95% Conf. Interval]  
-----+  
(1) | .1078964 .0778378 1.39 0.167 -.0455645 .2613572  
-----
```

Exercise 4.3

In the University of New Hampshire student survey, information was collected on the amount of alcohol students consumed. The result was a 33-point drinking scale score, where a higher score means more alcohol consumption. We are interested in whether there are differences in alcohol consumption among the years in school (year: 1=Freshman, 2=Sophomore, 3=Junior, 4=Senior), and also differences by sex (gender; 1=male, 0=female).

A regression model was fit to predict drinking score (`drink`) using year in school and sex. Use the Stata output provided at the end of the problem (note: output spans 2 pages) to answer the questions below.

- (a) Is there a significant effect of year in school on drinking score, adjusting for sex? Cite a p-value in your answer.
- (b) Is there a significant effect of sex on drinking score, controlling for year in school?
- (c) Interpret the estimated coefficient for `year3` in this model
- (d) Interpret the estimated coefficient for `gender` in this model

(e) Complete the table below with the estimated differences in means and the p-values for the comparisons.
(Note: You may notice that this table contains a LOT of comparisons/p-values, with no adjustment for the fact that you are doing a lot of tests. Later in the course we will learn about the problem of “multiple comparisons” and ways to adjust for the fact that doing lots of tests increases the chance you find something “spurious”.)

Group 1	Group 2	Estimated Mean Difference, Group 2 – Group 1	95% CI for Difference	P-value
Freshman	Sophomore			
Freshman	Junior			
Freshman	Senior			
Sophomore	Junior			
Sophomore	Senior			
Junior	Senior			

(f) Rank the years in school by estimated mean drinking score (adjusted for sex), from largest to smallest (i.e., from drinks the most to the least). (Hint: use the estimated differences from the table.)

```

. generate year1 = (year==1) if !missing(year)
. generate year2 = (year==2) if !missing(year)
. generate year3 = (year==3) if !missing(year)
. generate year4 = (year==4) if !missing(year)

. regress drink year2 year3 year4 gender

```

Source	SS	df	MS	Number of obs	=	243
				F(4, 238)	=	9.59
Model	1517.94605	4	379.486512	Prob > F	=	0.0000
Residual	9417.27206	238	39.56837	R-squared	=	0.1388
				Adj R-squared	=	0.1243
Total	10935.2181	242	45.1868517	Root MSE	=	6.2903

drink	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
year2	1.845409	1.266336	1.46	0.146	-.6492491 4.340067
year3	.1760209	1.233302	0.14	0.887	-2.253561 2.605603
year4	-2.31221	1.271725	-1.82	0.070	-4.817485 .1930656
gender	3.778905	.8144889	4.64	0.000	2.174377 5.383433
_cons	17.46344	1.046591	16.69	0.000	15.40167 19.5252

```
. test year2 year3 year4
```

```
( 1) year2 = 0
( 2) year3 = 0
( 3) year4 = 0
```

```
F( 3, 238) = 4.69
Prob > F = 0.0033
```

```
. lincom year3 - year2
```

```
( 1) - year2 + year3 = 0
```

drink	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----					
(1)	-1.669388	1.066031	-1.57	0.119	-3.76945 .4306743

```
. lincom year4 - year2
```

```
( 1) - year2 + year4 = 0
```

drink	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----					
(1)	-4.157619	1.114837	-3.73	0.000	-6.353827 -1.96141

```
. lincom year4 - year3
```

```
( 1) - year3 + year4 = 0
```

drink	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----					
(1)	-2.488231	1.077143	-2.31	0.022	-4.610182 -.366279

Exercise 4.4

Information on 74 automobiles was collected (in 1978) to study the relationship between gas mileage (mpg) and various features of the cars. The predictors we would like to investigate are the weight of the car (pounds; lbs), price of the car (price; \$), and whether the car is made in the U.S. (foreign: 0=made in U.S.; 1=made outside U.S.).

We would like to know whether adding price and foreign-made status improve on a model that includes weight as the only predictor. Use the Stata output on the next 2 pages to answer this question. (Perform all steps of the appropriate hypothesis test.) Note that more output is provided than you need for the problem.

```
. regress mpg pounds
```

Source	SS	df	MS	Number of obs	=	74
Model	1591.9902	1	1591.9902	F(1, 72)	=	134.62
Residual	851.469256	72	11.8259619	Prob > F	=	0.0000
				R-squared	=	0.6515
				Adj R-squared	=	0.6467
Total	2443.45946	73	33.4720474	Root MSE	=	3.4389

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
pounds	-.0060087	.0005179	-11.60	0.000	-.0070411 -.0049763
_cons	39.44028	1.614003	24.44	0.000	36.22283 42.65774

```
. regress mpg pounds price
```

Source	SS	df	MS	Number of obs	=	74
Model	1595.93249	2	797.966246	F(2, 71)	=	66.85
Residual	847.526967	71	11.9369995	Prob > F	=	0.0000
				R-squared	=	0.6531
				Adj R-squared	=	0.6434
Total	2443.45946	73	33.4720474	Root MSE	=	3.455

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
pounds	-.0058175	.0006175	-9.42	0.000	-.0070489 -.0045862
price	-.0000935	.0001627	-0.57	0.567	-.000418 .0002309
_cons	39.43966	1.621563	24.32	0.000	36.20635 42.67296

```
. regress mpg pounds price foreign
```

Source	SS	df	MS	Number of obs	=	74
				F(3, 70)	=	45.93
Model	1620.30716	3	540.102388	Prob > F	=	0.0000
Residual	823.152295	70	11.7593185	R-squared	=	0.6631
				Adj R-squared	=	0.6487
Total	2443.45946	73	33.4720474	Root MSE	=	3.4292

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
pounds	-.0067758	.0009048	-7.49	0.000	-.0085805 -.0049712
price	.0000566	.0001922	0.29	0.769	-.0003268 .00044
foreign	-1.855891	1.289063	-1.44	0.154	-4.426846 .7150641
_cons	41.95948	2.377726	17.65	0.000	37.21725 46.7017