Homework 5 Solutions:
Chapter 6: #12a, 13 (Use Bonferroni method in both cases)
Chapter 7: #1, 22, 28b, 31

Assigned Mon, Oct 26:

6.12a. Management wants g = 5 simultaneous confidence intervals with overall 95% confidence. Therefore, each interval will be done using α/g = .05/5 = .01. The appropriate t-multiplier has $\frac{.01}{2}$ = .005 in the upper tail. The appropriate degrees of freedom = n – p = 52 – 4 = 48. Therefore, the t-multiplier is written as t(.995, 48). In Stata, the upper tail area is used and the df value comes first, so the command is: `generate tmult=invttail(48,.005)`. We also need to find $\hat{Y}_h$ and s{$\hat{Y}_h$} for each of the 5 new sets of X values. See the answer to 6.13 (below) for the Stata commands for doing that, with the exception that the command `predict sepred, stdf` is now `predict seYhat, stdp.` The results are as follows:

X_{h1}	X_{h2}	X_{h3}		
302,000	7.2	0:	$4292.79 \pm 2.6822(21.3567)$	$4235.507 \leq E\{Y_h\} \leq 4350.073$
245,000	7.4	0:	$4245.29 \pm 2.6822(29.7021)$	$4165.623 \leq E\{Y_h\} \leq 4324.957$
280,000	6.9	0:	$4279.42 \pm 2.6822(24.4444)$	$4213.855 \leq E\{Y_h\} \leq 4344.985$
350,000	7.0	0:	$4333.20 \pm 2.6822(28.9293)$	$4255.606 \leq E\{Y_h\} \leq 4410.794$
295,000	6.7	1:	$4917.42 \pm 2.6822(62.4998)$	$4749.783 \leq E\{Y_h\} \leq 5085.057$

6.13 Management wants g = 4 prediction intervals with a family confidence coefficient of 95%. Therefore, each interval will be done using α/g = .05/4 = .0125. The appropriate t-multiplier has .0125/2 = .00625 in the upper tail. The appropriate degrees of freedom = n – p = 52 – 4 = 48. Therefore, the t-multiplier is written as t(.99375, 48). In Stata, the upper tail area is used and the df value comes first, so the command is: `generate tmult=invttail(48,.00625)`. We also need to find the predicted values and s{pred} for each interval. Here are the Stata commands and results for the whole problem, where the Y variable is called "Labor" and the X variables, in order, are called "Shipped" "Indirect" and "Holiday":

```
. set obs 56
obs was 52, now 56
```
* Note that you can add the following data using the data browser instead of the "replace" command.
```
. replace Shipped = 230000 in 53
. replace Shipped = 250000 in 54
. replace Shipped = 280000 in 55
. replace Shipped = 340000 in 56
. replace Indirect = 7.5 in 53
. replace Indirect = 7.1 in 54
. replace Indirect = 7.3 in 54
. replace Indirect = 7.1 in 55
. replace Indirect = 6.9 in 56
. replace Holiday = 0 in 53
. replace Holiday = 0 in 54
. replace Holiday = 0 in 55
. replace Holiday = 0 in 56
```

```
. regress Labor Shipped Indirect Holiday in 1/52

      Source |       SS       df       MS              Number of obs =      52
-------------+------------------------------           F(  3,    48) =   35.34
       Model |  2176606.18     3  725535.393           Prob > F      =  0.0000
    Residual |  985529.745    48  20531.8697           R-squared     =  0.6883
-------------+------------------------------           Adj R-squared =  0.6689
       Total |  3162135.92    51  62002.6652           Root MSE      =  143.29

------------------------------------------------------------------------------
       Labor |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     Shipped |   .0007871   .0003646     2.16   0.036      .0000541    .0015201
    Indirect |  -13.16602   23.09173    -0.57   0.571     -59.59506    33.26302
     Holiday |   623.5545   62.64095     9.95   0.000      497.6064    749.5025
       _cons |   4149.887   195.5654    21.22   0.000      3756.677    4543.098
------------------------------------------------------------------------------

. predict newlabor
(option xb assumed; fitted values)
. generate tmult = invttail(48,.00625)
. predict sepred, stdf
. generate lower = newlabor - tmult*sepred
. generate upper = newlabor + tmult*sepred
. list newlabor lower upper in 53/56
     +--------------------------------+
     | newlabor      lower      upper |
     |--------------------------------|
 53. |  4232.17   3849.911    4614.43 |
 54. | 4250.545   3871.478   4629.613 |
 55. | 4276.791   3900.122    4653.46 |
 56. | 4326.649   3947.913   4705.385 |
     +--------------------------------+
```

Assigned Wed, October 28

7.1 In each case, df = the number of terms being added to the model. So, they are:
(1) 1 **(2)** 1 **(3)** 2 **(4)** 3

7.22 The analyst doesn't understand how to interpret the tests for individual regression terms when the predictor variables are correlated. Remember that each test of H_0: $\beta_k = 0$ investigates whether that predictor is needed in the model, *given* that all of the other predictor variables are in the model. Therefore, when there are only three predictor variables, the individual tests are more likely to be statistically significant than when there are seven predictor variables in the model. Adding those extra variables reduces the impact of each of the other variables. The results do indicate that the predictor variables have some form of correlation with each other. If they did not, then the tests for the individual coefficients would remain the same regardless of what other predictor variables were in the model.

7.28b For testing whether or not $\beta_5 = 0$ the extra sum of squares is $SSR(X_5 \mid X_1, X_2, X_3, X_4)$.
For testing whether or not $\beta_2 = \beta_4 = 0$ the extra sum of squares is $SSR(X_2, X_4 \mid X_1, X_3, X_5)$.

7.31 The reduced models are written as follows. Remember that for (3) and (4), you have to subtract the terms with the fixed βs from the Y_i on the left hand side of the equation. It would not be wrong to simply put terms such as $5(X_{i1} + X_{i2})$ into the right hand side of the equation to specify the reduced model, but you wouldn't be able to run the model on the computer that way.

$$(1)\quad Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

$$(2)\quad Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_4 \sqrt{X_{i3}} + \varepsilon_i$$

$$(3)\quad Y_i' = Y_i - 5(X_{i1} + X_{i2}) = \beta_0 + \beta_3 X_{i1} X_{i2} + \beta_4 \sqrt{X_{i3}} + \varepsilon_i$$

$$(4)\quad Y_i' = Y_i - 7\sqrt{X_{i3}} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$