Niranjan Poudel
STAT 5100
Spring 2019

Homework 3

Exercise 1

Here, the given estimated regression equation is E{Y} = 350.7- 0.18X and given two-sided P-value for the estimated slope is 0.91. Although, the claim seems to be legit, looking at the equation, but the P-value is greater than 0.05, we have insufficient evidence to accept that there is relationship between advertising expenditures and sales, so the member of the student team is rushing to make the predictions, on the basis of given results.

Exercise 2

a. 99% confidence interval for β_1 is obtained which can be illustrated in the following table;

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	99% Confidence Limits	
Intercept	1	2.11405	0.32089	6.59	<.0001	1.27390	2.95420
act	1	0.03883	0.01277	3.04	0.0029	0.00539	0.07227

From the above observation, the confidence interval for β_1 is 0.00539 to 0.07227. Furthermore, the P-value is less than 0.01 which suggest the value is not zero. From this confidence interval we can say that if we were to run new sample collection, we can be 99% confident of getting the value of β_1 in the given interval. The 99% confidence interval does not include zero. The director may be interested in whether the confidence interval includes zero because, if the value of β_1 would be zero, then there would be no relationship between gpa and act.

c. The P-value is less than 0.01 which suggest the value is not zero. From this confidence interval we can say that if we were to run new sample collection, we can be 99% confident of getting the value of β_1 in the given interval. So, the director may be interested whether there might be zero or not to strengthen the dependency of gpa in the act.

Exercise 3

a The humidity level in the greenhouse tomorrow when we set the temperature level at 31 degrees Celsius, the prediction interval is appropriate.

b Families whose disposable income is $23,500 spend, on the average, for meals away from home, the confidence interval for a mean response is appropriate.

c Number of kilowatt-hours of electricity consumed by industrial and commercial users is given appropriate by prediction intervals.

Exercise 4

a) The 95 percent interval estimate of the mean freshman GPA for students whose Act test score is 28 is given in the following table;

Niranjan Poudel
STAT 5100
Spring 2019

Homework 3

Obs	gpa	act	predict	uconf	lconf
3	3.778	28	3.20121	3.34103	3.06138
23	3.731	28	3.20121	3.34103	3.06138
25	3.556	28	3.20121	3.34103	3.06138
27	2.420	28	3.20121	3.34103	3.06138
33	2.929	28	3.20121	3.34103	3.06138
79	4.000	28	3.20121	3.34103	3.06138
108	3.621	28	3.20121	3.34103	3.06138
109	3.792	28	3.20121	3.34103	3.06138
118	3.914	28	3.20121	3.34103	3.06138
120	2.948	28	3.20121	3.34103	3.06138

From above table for ACT test score 28 the 95% mean freshman GPA is within the range 3.0614 to3.341. If a number of data were to be collected for act score of 28 the mean of the gpa for those act score would fall within the interval mentioned above.

b) The prediction interval for 95% interval is given in the following table;

Predicted intervals for act=28

Obs	gpa	act	predict	upred	lpred
3	3.778	28	3.20121	4.44306	1.95935
23	3.731	28	3.20121	4.44306	1.95935
25	3.556	28	3.20121	4.44306	1.95935
27	2.420	28	3.20121	4.44306	1.95935
33	2.929	28	3.20121	4.44306	1.95935
79	4.000	28	3.20121	4.44306	1.95935
108	3.621	28	3.20121	4.44306	1.95935
109	3.792	28	3.20121	4.44306	1.95935
118	3.914	28	3.20121	4.44306	1.95935
120	2.948	28	3.20121	4.44306	1.95935

From the above table if Mary Jones obtained a score of 28 on the entrance test her freshman GPA would be from the interval of 1.96 to 4.44. This means she has 95 percent chance of scoring from 1.96 to 4.4 based on her act entrance score.

c) Yes, the prediction interval in b) is wider than the confidence interval in a). Yes, the prediction interval should be wider than the confidence interval. The prediction interval is the prediction of actual value whereas the confidence interval is prediction for the mean of the values for certain act score. The average of the values will be somewhat inner that the actual value of prediction in the linear regression function, as the linear regression function is compacted around the mean predicted value.

d) Following table gives the boundary values of the 95 percent confidence band for regression line when X_h=28 using Working-Hoteling and Bonferroni intervals.

simulataneous 95% interval estimation of individual prediction at one X-level, using Working-Hotelling and Bonferroni

Obs	act	Yhat	seYhat	WH_lower	WH_upper	B_lower	B_upper
121	28	3.20121	0.070609	3.02616	3.37626	3.06138	3.34103

From the above table we get the confidence band from 3.026 to 3.376 which is wider than the confidence interval in the problem (a) which is 3.061 to 3.341. It should be wider than the confidence interval because *t<W.* So, the confidence band in Working-Hoteling gives wider predictions.

Exercise 5

If the analyst fitted regression model as $Y_i = \beta_0 + \beta_1 X + \epsilon_i$ and after conducting a f-test, finding the p-value 0.33 and convincing him/her to conclude $\beta_1 \neq 0$ then α level used by the analyst was greater than 0.33. If the p-value is less than α level than the null hypothesis $\beta_1=0$ does not have enough evidence to be supported. If the α level was 0.01 than the analyst would have no enough evidence to reject the null hypothesis listed, $\beta_1=0$.

Exercise 6.

a. Bonferroni joint confidence intervals for β_0 and β_1 is given by the following table;

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	6.33679	0.52128	12.16	<.0001	5.30384	7.36973
risk	1	0.76042	0.11444	6.64	<.0001	0.53364	0.98720

Here, the table shows 95% Confidence limits but the prediction is for two parameters so we use α/g which shows 95%. So, the joint confidence interval for β_0 is 5.304 to 7.370 and for β_1 is 0.534 to 0.987.

b. The expectations of the researcher suggest that the value of the β_0 should be approximately 7 and the value of β_1 should be approximately 1. From the above confidence limits the value could be 7 and approximately 1 (after round off). From the

Homework 3

estimate we cannot suggest that the value should be that, but we can say that the value could be that. There is a chance of the value being as said by the researchers, there is a probability but not certainty. Furthermore, the standard value chosen for this parameters are 6.337 and 0.760 from the parameter estimates.

c. Here we have the table of estimate using Working-Hoteling and Bonferroni which is as follow;

(95% interval estimation of mean rtesponse at 4 X levels using Working-Hoteling abd Bonferroni

Obs	risk	Yhat	seYhat	WH_upper	WH_lower	B_upper	B_lower
114	2	7.8576	0.30979	8.6263	7.08899	8.6442	7.07105
115	3	8.6180	0.21768	9.1581	8.07796	9.1707	8.06536
116	4	9.3785	0.15808	9.7707	8.98624	9.7799	8.97709
117	5	10.1389	0.16968	10.5599	9.71788	10.5697	9.70806

Here the narrower observation is given by Working-Hoteling approach so for the calculation of the estimate the expected hospital stay we use Working-Hoteling approach.

d. The family of intervals using Working-Hoteling is shown below for simultaneous risks level 2,3,4 and 5.

WH_upper	WH_lower
8.6263	7.08899
9.1581	8.07796
9.7707	8.98624
10.5599	9.71788

For the risk X=2 the confidence interval is 7.089 to 8.626, X=3 is 8.078 to 9.158, X=4 is 8.986 to 9.771 and for X=5 is 9.718 to 10.560. This means that if we were to collect data for the risk level at given intervals, we have 95% confidence that the mean of the length of stay in days for the given certain level falls between the given interval of days.

Exercise 7.

From the problem 6 we get the linear regression equation as E{Y}= 6.337 +0.760X.

If the average length of stay is to be reduced to 9 days than using the above equation.

We have, 9= 6.337 +0.760X

i.e., (9-6.337)/0.760 =X

i.e., X =3.504

Therefore, if the hospital wants to reduce the average length of stay to 9 days than the infection risk to achieve this level is 3.5 approximately.

Appendix: SAS Code

```
/* Starter SAS code for STAT 5100 Homework 3 */


/* Exercises 2 AND 4 */

data college;   input gpa act @@; cards;
 3.897 21 3.885 14 3.778 28 2.540 22 3.028 21 3.865 31 2.962 32 3.961
27 0.500 29 3.178 26
 3.310 24 3.538 30 3.083 24 3.013 24 3.245 33 2.963 27 3.522 25 3.013
31 2.947 25 2.118 20
 2.563 24 3.357 21 3.731 28 3.925 27 3.556 28 3.101 26 2.420 28 2.579
22 3.871 26 3.060 21
 3.927 25 2.375 16 2.929 28 3.375 26 2.857 22 3.072 24 3.381 21 3.290
30 3.549 27 3.646 26
 2.978 26 2.654 30 2.540 24 2.250 26 2.069 29 2.617 24 2.183 31 2.000
15 2.952 19 3.806 18
 2.871 27 3.352 16 3.305 27 2.952 26 3.547 24 3.691 30 3.160 21 2.194
20 3.323 30 3.936 29
 2.922 25 2.716 23 3.370 25 3.606 23 2.642 30 2.452 21 2.655 24 3.714
32 1.806 18 3.516 23
 3.039 20 2.966 23 2.482 18 2.700 18 3.920 29 2.834 20 3.222 23 3.084
26 4.000 28 3.511 34
 3.323 20 3.072 20 2.079 26 3.875 32 3.208 25 2.920 27 3.345 27 3.956
29 3.808 19 2.506 21
 3.886 24 2.183 27 3.429 25 3.024 18 3.750 29 3.833 24 3.113 27 2.875
21 2.747 19 2.311 18
 1.841 25 1.583 18 2.879 20 3.591 32 2.914 24 3.716 35 2.800 25 3.621
28 3.792 28 2.867 25
 3.419 22 3.600 30 2.394 20 2.286 20 1.486 31 3.885 20 3.800 29 3.914
28 1.860 16 2.948 28
;
run;

proc reg data=college;
       model gpa= act/ clb alpha = 0.01;
       title1'Regression with 99% interval';
run;

proc reg data = college;
       model gpa= act/ clb alpha= 0.05;
       output out= confidence predicted= predict
                             ucl=upred
                             lcl=lpred
```

```
                          uclm= uconf
                          lclm= lconf;
       title1'Regression with 95% interval';
run;

proc print data= confidence;
       where act=28;
       var gpa act predict uconf lconf;
       title1'Confidence intervals';
       title2'for act=28';
run;

proc print data= confidence;
       where act=28;
       var gpa act predict upred lpred;
       title1'Predicted intervals for act=28';
run;

data dummy; input act check; cards;
28  1
;
data trick; set college dummy;
proc reg data=trick noprint;
       model gpa=act;
       output out=out1 predicted=Yhat stdp=seYhat;
data out1; set out1;
       alpha=0.05;
       p=2;
       n=120;
       g=1;
       W = sqrt(p*finv(1-alpha,p,n-p)); /* WH crit. val. */
        t = tinv(1-alpha/(2*g),n-p); /* Bonf. crit. val. */
        WH_upper = Yhat + W*seYhat;
        WH_lower = Yhat - W*seYhat;
       B_upper = Yhat + t*seYhat;
       B_lower = Yhat - t*seYhat;

proc print data= out1;
       where check=1;
       var act Yhat seYhat WH_lower WH_upper B_lower B_Upper;
       title1 'simulataneous 95% interval estimation of individual
prediction';
       title2 'at one X-level, using Working-Hotelling and
Bonferroni';
run;
```

```
/* Exercise 6 */

data senic; input length risk @@; cards;
  7.13  4.1  8.82  1.6  8.34  2.7  8.95  5.6  11.2  5.7  9.76  5.1
  9.68  4.6  11.18  5.4  8.67  4.3  8.84  6.3  11.07  4.9  8.3  4.3
  12.78  7.7  7.58  3.7  9  4.2  11.08  5.5  8.28  4.5  11.62  6.4
  9.06  4.2  9.35  4.1  7.53  4.2  10.24  4.8  9.78  5  9.84  4.8
  9.2  4  8.28  3.9  9.31  4.5  8.19  3.2  11.65  4.4  9.89  4.9
  11.03  5  9.84  5.2  11.77  5.3  13.59  6.1  9.74  6.3  10.33  5
  9.97  2.8  7.84  4.6  10.47  4.1  8.16  1.3  8.48  3.7  10.72  4.7
  11.2  3  10.12  5.6  8.37  5.5  10.16  4.6  19.56  6.5  10.9  5.5
  7.67  1.8  8.88  4.2  11.48  5.6  9.23  4.3  11.41  7.6  12.07  7.8
  8.63  3.1  11.15  3.9  7.14  3.7  7.65  4.3  10.73  3.9    11.46
4.5
  10.42  3.4  11.18  5.7  7.93  5.4  9.66  4.4  7.78  5  9.42  4.3
  10.02  4.4  8.58  3.7  9.61  4.5  8.03  3.5  7.39  4.2  7.08  2
  9.53  5.2  10.05  4.5  8.45  3.4  6.7  4.5  8.9  2.9  10.23  4.9
  8.88  4.4  10.3  5.1  10.79  2.9  7.94  3.5  7.63  5.5  8.77  4.7
  8.09  1.7  9.05  4.1  7.91  2.9  10.39  4.3  9.36  4.8  11.41  5.8
  8.86  2.9  8.93  2  8.92  1.3  8.15  5.3  9.77  5.3  8.54  2.5
  8.66  3.8  12.01  4.8  7.95  2.3  10.15  6.2  9.76  2.6  9.89  4.3
  7.14  2.7  13.95  6.6  9.44  4.5  10.8  2.9  7.14  1.4  8.02  2.1
  11.8  5.7  9.5  5.8  7.7  4.4  17.94  5.9  9.41  3.1
;
run;

proc reg data=senic;
       model length=risk / clb alpha = 0.05; /* g=2 */
       title1 'regression with CI 90%';
run;

data dummy; input risk check @@; cards;
  2 1  3 1  4 1  5 1
;
data temp; set senic dummy;
proc reg data=temp noprint;
       model length=risk;
       output out=out2 predicted=Yhat stdp=seYhat;
       data out2; set out2;
              alpha=0.05;
              p = 2;
              n= 113;
              g= 4;
               W = sqrt(p*finv(1-alpha,p,n-p));
               t = tinv(1-alpha/(2*g),n-p);
               WH_upper = Yhat + W*seYhat;
```

```
              WH_lower = Yhat - W*seYhat;
              B_upper = Yhat + t*seYhat;
              B_lower = Yhat - t*seYhat;
      proc print data=out2;
             where check=1;
             var risk Yhat seYhat WH_upper WH_lower B_upper B_lower;
             title1 '(95% interval estimation of mean rtesponse at
                           4 X levels using Working-Hoteling abd
Bonferroni';
run;
```