Andrew Curtis

STAT 5100

Spring 2022

Homework #3

Exercise 1

It is an unexpected result to have a negative slope when comparing expenditure on advertising and its effect on sales. However, a few additional things may be considered. The slope is only slightly negative and there is a high initial intercept value. The two-sided p-value is 0.91, this is very large and shows little significance. Additionally, a small slope value indicates little to no correlation between the two variables. I would say that the student fails to have evidence for a significant correlation, negative or otherwise, between advertising expenditure and sales.

Exercise 2

a)

The 99% confidence interval for β_1 is (0.00539, 0.07227). This is to say, that we have 99% confidence that the true value for β_1 or the slope falls within these values. The values are small, so we would expect a very small positive correlation between ACT score and GPA. The interval does not include 0. If it did, that would indicate the possibility of no correlation between the variables. The admissions official might be interested to see if the no correlation is a possibility, because it would not be a good metric to choose students on, and it still might not be with the low β_1 value.

c)

The p-value is 0.0029, this is less than the stated significance level of 0.01. This indicates that there is evidence to reject the null hypothesis that there is no effect of ACT score on GPA at the end of the freshman year. This p-value also shows that there is some correlation between the two variables. This supports the results found in a), specifically that 0 does not appear in the confidence interval for β_1.

Exercise 3

a)

The dependent variable value is being asked for based on a new value of the independent variable, this calls for a prediction interval.

b)

This is asking for a response from a mean of families at a certain value. Since it is that mean response, it should be a confidence interval.

c)

Notice the question is for next month. It is asking for a prediction of the dependent variable given that the business activity independent value stays at the current index. Predicting the future value will use a prediction interval.

Exercise 4

a)

The 95% confidence interval for a prediction of GPA from an ACT score of 28 is (3.06138, 3.34103).

b)

For a score of 28 on the ACT, the prediction interval is (1.95935, 4.44306) for GPA. This is much wider than the confidence interval, which should be the case. There is greater uncertainty involved when predicting a single new value.

c)

This is much wider than the confidence interval, which should be the case. There is greater uncertainty involved when predicting a single new value, rather than regressing based on many values.

d)

Working-Hotelling confidence band for an ACT score of 28 is (3.02616, 3.37626). This is slightly wider than the confidence interval in part a. This makes sense because you are adding one new predicted value.

Exercise 5

The analyst rejected the null hypothesis in favor of the alternative hypothesis. His α value must have been greater than the p-value of 0.033. If the α value was 0.01, the appropriate conclusion would be to fail to reject the null hypothesis. Not significant enough evidence to reject.

Exercise 6

a)

Using a 90% family confidence limit:

For β_0, the confidence limit is (5.30384, 7.36973).

For β_1, the confidence limit is (0.53364, 0.98720).

b)

The researcher suggests that the true value for β_0 should be equal to 7. This does fall within the confidence limit. The researcher suggests the true value for β_1 should be equal to 1. This does not fall within the confidence limit. Meaning even with the 90% joint confidence limit this value does not appear. There may be some error, something wrong with the model, or the researcher has an incorrect prediction.

c)

The working-hotelling and Bonferroni confidence bands are as follows:

Simultaneous 95% interval estimation of mean response at four X-levels, using Working-Hotelling and Bonferroni

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

The working-hotelling confidence bands are tighter (more efficient) for each observation value.

Part d on next page.

d)

See above part c). Additionally, the higher the risk the value, the longer the length of stay is in general.

The working-hotelling procedure uses Yhat values and standard error of Yhat, along with the working-hotelling critical value that is calculated from the F distribution and the simultaneous confidence limit. Essentially, this is a confidence band that simultaneously considers β_0 and β_1.

Exercise 7

This involves an inverse prediction of X_h. Assuming the 90% simultaneous confidence interval for the model from part a,

The point estimate for the risk value from a stay of 9 days is 3.50229.

Appendix SAS code

```
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;
  output out=collegeout r=resid p=pred;
  title1 'Simple linear model';
run;

proc reg data=college;
  model gpa = act / clb alpha=.01;
                                          /* 1-alpha is level */
  output out=confidence p=Predict
                        ucl=uPred   /* upper and lower limits for */
                        lcl=lPred   /*      individual prediction */
                        uclm=uConf  /* upper and lower limits for */
                        lclm=lConf; /*      group mean confidence */
  title1 'Regression with 99% interval estimation';
run;

data dummy; input act; cards;
28
;
data trick; set college dummy;
run;

proc reg data=trick;
  model gpa = act / clb alpha=.05;
                                          /* 1-alpha is level */
  output out=confidence p=Predict
                        ucl=uPred   /* upper and lower limits for */
                        lcl=lPred   /*      individual prediction */
                        uclm=uConf  /* upper and lower limits for */
                        lclm=lConf; /*      group mean confidence */
  title1 'Regression with 95% interval estimation';
run;
```

```
proc print data=confidence;
  where act = 28;
       /* which observations to use in proc */
  var act gpa Predict lPred uPred lConf uConf;
  title1 'Predicted values and confidence and predicted
intervals';
  title2 'for act = 28; this is 95% interval';
run;

data dummy; input act check; cards;
  28  1
  ;
data temp; set college dummy;
proc reg data=temp noprint;
  model gpa = act;
  output out=out1 p=Yhat stdp=seYhat;
    /* KEY: stdp is SE of mean prediction */
data out1; set out1;
  alpha = 0.05;  /* 1-alpha is simult. conf. level */
  p = 2;         /* # of beta's (including intercept) */
  n = 120;        /* sample size */
  g = 1;         /* number of simultaneous intervals */
  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
   'Simultaneous 95% interval estimation of mean response';
  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=.05;
  title1 'Simultaneous 90% confidence intervals on betas';
run;

data dummy; input risk check @@; cards;
  2 1  3 1  4 1  5 1
;
data temp; set senic dummy;
run;

proc reg data=temp noprint;
  model length = risk;
  output out=out1 p=Yhat stdp=seYhat;
    /* KEY: stdp is SE of mean prediction */
data out1; set out1;
  alpha = 0.05;  /* 1-alpha is simult. conf. level */
  p = 2;         /* # of beta's (including intercept) */
  n = 113;        /* sample size */
  g = 4;         /* number of simultaneous intervals */
  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 risk Yhat seYhat WH_lower WH_upper
      B_lower B_upper;
  title1
   'Simultaneous 95% interval estimation of mean response';
  title2
   'at four X-levels, using Working-Hotelling and
    Bonferroni';
run;
```