STAT 350

Assignment 2: Solutions

1. From the text question 1.19, 1.23, 2.13 a and b and 2.23 a, b and c. In 2.23 c give a P-value and interpret this P-value.

   The following SAS code was used for all these parts:

```
data gpa;
 infile "CH01PR19.DAT";
 input gpa test;
proc glm;
  model gpa=test;
  estimate "fit_at_5" intercept 1 test 5;
  estimate "fit4.7" intercept 1 test 4.7;
  output out=gpaout r=resid p=fitted ;
run;
proc print data=gpaout;
  var test gpa fitted resid;
run;
proc means data=gpaout;
 var resid;
run;
proc rank normal=vw data=gpaout out=gpaout2;
  var resid;
  ranks normscr;
run;
proc gplot data=gpaout2;
 plot gpa*test fitted *test /overlay;
run;
proc gplot data=gpaout2;
 plot fitted*test;
 plot resid*normscr;
run;
```

   obtaining the following output:

```
Dependent Variable: GPA
                                    Sum of           Mean
Source                    DF       Squares         Square  F Value     Pr > F
Model                      1    6.43372807     6.43372807    34.00     0.0001
Error                     18    3.40627193     0.18923733
Corrected Total           19    9.84000000
```

R-Square	C.V.	Root MSE	GPA Mean
0.653834	17.40057	0.43501	2.50000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TEST	1	6.43372807	6.43372807	34.00	0.0001
Source	DF	Type III SS	Mean Square	F Value	Pr > F
TEST	1	6.43372807	6.43372807	34.00	0.0001

Dependent Variable: GPA

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
fit_at_5	2.50000000	25.70	0.0001	0.09727213
fit4.7	2.24802632	21.12	0.0001	0.10643937

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
INTERCEPT	-1.699561404	-2.34	0.0311	0.72677682
TEST	0.839912281	5.83	0.0001	0.14404759

OBS	TEST	GPA	FITTED	RESID
1	5.5	3.1	2.91996	0.18004
2	4.8	2.3	2.33202	-0.03202
3	4.7	3.0	2.24803	0.75197
4	3.9	1.9	1.57610	0.32390
5	4.5	2.5	2.08004	0.41996
6	6.2	3.7	3.50789	0.19211
7	6.0	3.4	3.33991	0.06009
8	5.2	2.6	2.66798	-0.06798
9	4.7	2.8	2.24803	0.55197
10	4.3	1.6	1.91206	-0.31206
11	4.9	2.0	2.41601	-0.41601
12	5.4	2.9	2.83596	0.06404
13	5.0	2.3	2.50000	-0.20000
14	6.3	3.2	3.59189	-0.39189
15	4.6	1.8	2.16404	-0.36404
16	4.3	1.4	1.91206	-0.51206
17	5.0	2.0	2.50000	-0.50000
18	5.9	3.8	3.25592	0.54408
19	4.1	2.2	1.74408	0.45592
20	4.7	1.5	2.24803	-0.74803

Analysis Variable : RESID

```
 N           Mean        Std Dev        Minimum        Maximum
---------------------------------------------------------------
20   -7.77156E-17      0.4234117     -0.7480263      0.7519737
---------------------------------------------------------------
```

Now to answer the questions:

1.19 a) The estimates are $\hat\beta_0 = -1.7$ and $\hat\beta_1 = 0.84$, rounded off sensibly.

b) It is not so easy to plot the data and regression line; the code above does so but the fitted line is plotted only at the available x values. However, nothing in the plot suggests any big problems.

c) The first `estimate` line produces the estimae 2.5 with a standard error of 0.097.

d) This just asks for the slope again which is 0.84.

1.23 a) The `proc print` prints out the residuals and the `proc means` shows that the variable `resid` has mean very close to 0.

b) The MSE is $\hat\sigma^2 = 0.189$ and the estimate of σ is the Root MSE, 0.435. The RMSE has the same units as σ which has the same units as Y, namely, grade points.

2.13 a) This confidence interval is

$$\hat\beta_0 + 4.7\hat\beta_1 \pm t_{0.025,18}\hat\sigma\sqrt{x^T(X^TX)^{-1}x}$$

where $x^T = (1, 4.7)$. The quantity $\hat\sigma\sqrt{x^T(X^TX)^{-1}x}$ is printed out by the `estimate` statement. It is the standard error 0.106. The actual estimate is 2.248 and the t multiplier is 2.101 so the interval is

$$2.248 \pm 2.101 \times 0.106 \qquad \text{or} \qquad (2.025, 2.471)$$

with one more digit than needed just so you can check your method.

b) The prediction for an individual is the same but the interval is

$$\hat\beta_0 + 4.7\hat\beta_1 \pm t_{0.025,18}\hat\sigma\sqrt{1 + x^T(X^TX)^{-1}x}$$

or

$$2.248 \pm 2.101 \times \sqrt{0.106^2 + 0.189}$$

or just

$$(1.308, 3.188)$$

Notice that this is VERY wide.

2.23 a) The ANOVA table is straight from the output:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.43372807	6.43372807	34.00	0.0001
Error	18	3.40627193	0.18923733		
Corrected Total	19	9.84000000			

b) MSR is an estimate of its expected value. See formula 2.57 in the text:

$$\mathrm{E}(MSE) = \sigma^2 + \beta_1^2 \sum (x_i - \bar{X})^2$$

You can use, for instance, `proc means` to compute the sum of squares at the end or algebraic identities to deduce it from `proc glm` output. The MSE is an unbiased estimate of σ^2. These two expected values are equal if the hypothesis $\beta_1 = 0$ is true.

c) The F statistic in question is printed in the ANOVA table. It is 34 and the P-value is 0.0001 so that at any reasonable level the hypothesis that $\beta_1 = 0$ is rejected. The null hypothesis is $\beta_1 = 0$, the alternative is $\beta_1 \neq 0$ and the conclusion is that the true slope is NOT 0.

2. Working with partitioned matrices. Suppose that the design matrix X is partitioned as $X = [\mathbf{1}|X_1|X_2]$ where X_i has p_i columns.

(a) Write $X^T X$ as a partitioned (3 rows, 3 columns) matrix.

Solution

$$X^T X = \begin{bmatrix} n & 1^T X_1 & 1^t X_2 \\ X_1^T 1 & X_1^T X_1 & X_1^T X_2 \\ X_2^T 1 & X_2^T X_1 & X_2^T X_2 \end{bmatrix}.$$

(b) A matrix

$$A = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{bmatrix}$$

is called *block diagonal*. Show that A^{-1} exists if and only if each A_i^{-1} exists and that then A^{-1} is block diagonal.

Solution

Check by multiplying that

$$\begin{bmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{bmatrix} \begin{bmatrix} (A_1)^{-1} & 0 & 0 \\ 0 & (A_2)^{-1} & 0 \\ 0 & 0 & (A_3)^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}.$$

This shows that if each A_i is invertible then so is A. I notice in marking that many of you assumed that each A_i was a scalar so that $\frac{1}{A_i}$ would make sense. That would make A diagonal, not block diagonal and would be useless for the next part of this question. To do the converse suppose that B is A^{-1} and partition B into a 3×3 array with entries B_{ij}. Multiply AB and set this equal to the identity. You get 9 equations like $A_1 B_{1,1} = I$ and $A_1 B_{1,2} = 0$. The first such equation shows that A_1 must be invertible and that $B_{1,1}$ must be the inverse of A_1. The second equation then shows (because we now know that A_1 is invertible that $B_{1,2} = 0$. Continue like this.

(c) Suppose that $\mathbf{1}^T X_i = 0$ for $i = 1, 2$ and $X_1^T X_2 = 0$. Show that $X^T X$ is block diagonal and give a formula for $(X^T X)^{-1}$.

Solutions

The conditions show that all the off-diagonal blocks are 0, remembering that $(AB)^T = B^T A^T$. Thus

$$(X^T X)^{-1} = \begin{bmatrix} 1/n & 0 & 0 \\ 0 & (X_1^T X_1)^{-1} & 0 \\ 0 & 0 & (X_2^T X_2)^{-1} \end{bmatrix}$$

Many students appeared not to realize that $\mathbf{1}^T \mathbf{1} = n$ because $\mathbf{1}$ is a **vector** all of whose entries equal the number 1.

(d) Suppose $\beta^T = [\beta_0 | \beta_1^T | \beta_2^T]$ is partitioned to conform with the partitioning of X (that is β_0 is a scalar and β_i is a column vector of length p_i for $i = 1, 2$. Let $\tilde{\beta}_0$ be obtained by fitting

$$Y = \mathbf{1}\beta_0 + \epsilon$$

by least square, $\tilde{\beta}_1$ be obtained by fitting

$$Y = X_1 \beta_1 + \epsilon$$

and similarly for $\tilde{\beta}_2$. Let $\hat{\beta}$ be the usual least squares estimate for

$$Y = X\beta + \epsilon \, .$$

Show that $\hat{\beta}^T = [\tilde{\beta}_0 | \tilde{\beta}_1^T | \tilde{\beta}_2^T]$.

Solutions

Multiply out the partitioned matrix $\hat{\beta} = (X^T X)^{-1} X^T Y$ to get

$$\hat{\beta} = \begin{bmatrix} \bar{Y} \\ (X_1^T X_1)^{-1} X_1^T Y \\ (X_2^T X_2)^{-1} X_2^T Y \end{bmatrix} = \begin{bmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{bmatrix} .$$

(e) Let $\hat{\mu}_i$ be the vectors of fitted values corresponding to the estimates $\tilde{\beta}_i$ for $i = 0, 1, 2$. Show that for $i \neq j$ we have $\hat{\mu}_i \perp \hat{\mu}_j$.

Solution

$$\hat{\mu}_1^T \hat{\mu}_2 = \tilde{\beta}_1^T X_1^T X_2 \tilde{\beta}_2 = 0$$

because the centre term $X_1^T X_2 = 0$. You must check similar formulas for $\hat{\mu}_0^T \hat{\mu}_1$ and so on.

(f) For the design matrix X_b of the first assignment identify X_1 and X_2 and verify the orthogonality condition of this problem.

Solution

In the solution set for assignment the first column of X_b is 1, the second is X_1 and the other two columns are X_2. Now just multiply things like $X_1^T X_2$ to make sure you get 0.