56:295 – 001
Multivariate Statistics and Advanced Quality Control
Fall 05

HW3 Due: October 5 (Wednesday), 6:15pm

Solution

1. Solve the following problems from the textbook (J&W)

7.1
Ans.

$$\hat{\underset{\sim}{\beta}} = (Z'Z)^{-1}Z'\underset{\sim}{y} = \frac{1}{120}\begin{bmatrix}120 & -10\\ -10 & 1\end{bmatrix}\begin{bmatrix}72\\ 872\end{bmatrix} = \frac{1}{15}\begin{bmatrix}-10\\ 19\end{bmatrix} = \begin{bmatrix}-.667\\ 1.267\end{bmatrix}$$

$$\hat{\underset{\sim}{y}} = Z\hat{\underset{\sim}{\beta}} = \frac{1}{15}\begin{bmatrix}180\\ 85\\ 123\\ 351\\ 199\\ 142\end{bmatrix} = \begin{bmatrix}12.000\\ 5.667\\ 8.200\\ 23.400\\ 13.267\\ 9.467\end{bmatrix}; \quad \hat{\underset{\sim}{\varepsilon}} = \underset{\sim}{y} - \hat{\underset{\sim}{y}} = \begin{bmatrix}15\\ 9\\ 3\\ 25\\ 9\\ 13\end{bmatrix} - \begin{bmatrix}12.000\\ 5.667\\ 8.200\\ 23.400\\ 13.267\\ 9.467\end{bmatrix} = \begin{bmatrix}3.000\\ 3.333\\ -5.200\\ 1.600\\ -6.267\\ 3.533\end{bmatrix}$$

Residual sum of squares: $\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}} = 101.467$

Fitted equation: $\hat{y} = -.667 + 1.267\ z_1$

7.17(a) (use Minitab or other software; report the fitted/regression equation and the standard deviation of the coefficients for the intercept and the two predictors based on software output).

The data are as follows:
126974 4224 173297
96933 3835 160893
86656 3510 83219
63438 3758 77734
55264 3939 128344
50976 1809 39080
39069 2946 38528
36156 359 51038
35209 2480 34715
32416 2413 25636

Ans.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	6519120.8603	3259560.4302	3.593	0.0844
Error	7	6351059.2397	907294.17709		
C Total	9	12870180.1			

Root MSE	952.51991	R-square	0.5065
Dep Mean	2927.30000	Adj R-sq	0.3655
C.V.	32.53920		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	1464.452545	711.36299816	2.059	0.0785
SALES	1	0.010348	0.02077169	0.498	0.6336
ASSETS	1	0.010069	0.01203479	0.837	0.4304

The fitted equation is
Profit = 1464.45+0.0103SALE+0.0101ASSETS
The standard deviations for Intercept, Sale, and Asset are
711.36, 0.0208, and 0.0120, respectively.

2. For the data and model in 7.1 of J&W, answer the following questions:
a. Find the "hat" matrix.
Ans.

$$\mathbf{H}=\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T=\begin{bmatrix} 0.17 & 0.17 & 0.17 & 0.17 & 0.17 & 0.17 \\ 0.17 & 0.375 & 0.29 & -0.21 & 0.13 & 0.25 \\ 0.17 & 0.29 & 0.24 & -0.06 & 0.14 & 0.22 \\ 0.17 & -0.21 & -0.06 & 0.84 & 0.24 & 0.02 \\ 0.17 & 0.13 & 0.14 & 0.24 & 0.18 & 0.15 \\ 0.17 & 0.25 & 0.22 & 0.02 & 0.15 & 0.2 \end{bmatrix}$$

b. Find the total sum of squares, regression sum of squares and residual sum of squares about the mean and verify the sum of squares decomposition results.
Ans. SST=294
SSR=192.53
SSE=101.47
SST=SSR+SSE, which satisfies the sum of squares decomposition result.

c. Calculate the coefficient of determination R^2.
Ans. R^2=SSR/SST=0.6549

d. Verify that $\hat{\mathbf{y}}$ is in the linear space spanned by the columns of **Z** by writing $\hat{\mathbf{y}}$ as a linear combination of columns of **Z**.
Ans. It can be seen that

$\hat{\mathbf{y}} = -0.67\mathbf{Z}_1 + 1.267\mathbf{Z}_2$, where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the two columns of **Z**.

e. What is the orthogonal projection of the vector 2**y** on the space spanned by columns of **Z**?
Ans. orthogonal projection of 2**y** is **H**·2**y**=[24 11.3 16.4 46.8 26.5 18.9]^T.

f. Find $\hat{\sigma}^2$, the unbiased estimate of σ^2.
Ans.

$$\hat{\sigma}^2 = \frac{SSE}{n-r-1} = 25.3667$$

g. Conduct both the *F*-test and the *t*-test for $H_0 : \beta_1 = 0$. Find the *P*-value for both tests and compare the two results. (If you are not sure of *P*-value, please review Section 3-3.2 of the QC book or other elementary statistics textbook)

Ans. $F = \dfrac{RSS_p - RSS_f}{1} \Big/ \dfrac{RSS_f}{n-r-1} = 7.59$

The p-value for *F*-test= $\Pr(F > 7.59) = 0.0511$. That is, the null hypothesis will be rejected for any alpha value greater than 0.0511.

$t = \dfrac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = 2.775$. The p-value for t-test is $\Pr(|t| > 2.775) = 0.0511$. That is, the null hypothesis will be rejected for any alpha value greater than 0.0511 under the t-test. The p-values for the F-test and t-test are the same. So the two tests are equivalent.