Stat 512 Spring 2009 Midterm Exam 2

Instructor: Bowei Xi

Name: Solution

Department:

To receive full credit:
Show all work that lead to the final answer on the attached pages

In one dataset, the response variable Y and three predictors X_1, X_3, and X_4 are quantitative variables. Another predictor X_2 is qualitative (categorical) and can take two values 0 and 1. Dataset look like:

Y	X1	X2	X3	X4
78.2	75.9	0	1.9881	4.32942
63.3	68.2	1	39.5641	4.22244

......

Based on the SAS outputs, answer the following questions.

1. (2) We fit a regression model $Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_3 + \epsilon$. The design matrix $\mathbf{X}$ has dimensions $30 \times p$. What is the value of p? What is the dimension of the error vector? What is the sample size n?

2. (2) The following is part of the SAS output for the model above:

Variable	DF	Parameter Estimate	Standard Error
Intercept	1	-124.30436	20.88642
X1	1	2.73731	0.27810
X3	1	0.01907	0.08565

Write down the estimated model. And construct 98% confidence intervals using Bonferroni procedure for β_1 and β_2. A T-table is provided.

3. (1) Based on the Bonferroni confidence intervals, do we need both X_1 and X_3? Explain.

(1) $p = 2+1 = 3$ $\quad \varepsilon: 30 \times 1$ $\quad n = 30$

(2.) $\hat{Y} = -124.30 + 2.74 \times X_1 + 0.01907 \times X_2$

98% Bonferroni intervals for β_1 and β_2,

99% for each interval $\quad 1 - \left(\frac{1-0.99}{2}\right) = 0.995$

$t_{30-2-1}(0.995) = 2.771$

β_1 : $2.74 \pm 2.771 \times 0.2781 = (1.96, 3.50)$

β_2 : $0.01907 \pm 2.771 \times 0.08565 = (-0.2182, 0.2564)$

(3) Bonferroni interval for β_2 covers 0. Then β_2 might be 0 and X_3 could be dropped from the model

We have the following SAS output for fitting the model $Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_3 + \epsilon$.

Source	DF	Sum of Squares	F Value	Pr > F
Model	2	??	??	<.0001
Error	27	??		
Corrected Total	29	2353.65367		

Variable	DF	Parameter Estimate	Standard Error	Type I SS	Type II SS
Intercept	1	-124.30436	20.88642	190834	655.70236
X1	1	2.73731	0.27810	1852.90171	1793.49255
X3	1	0.01907	0.08565	0.91803	0.91803

4. (2) Given the Type I and Type II sum of squares, what are the model sum of squares SSM, and the error sum of squares SSE (two missing numbers in the SAS output)? And what is R^2?

5. (1) Is there any predictor variable for which the Type I and Type II sum of squares are the same? Explain why.

6. (2) For the test $H_0 : \beta_1 = \beta_2 = 0$, write down the alternative hypothesis, compute the test statistic (another missing number in the SAS output), and state your conclusion. Is it consistent with the Bonferroni confidence intervals?

④ $SSM = 1852.90171 + 0.91803$ = sum of Type I SS $= 1853.81973$

$SSE = SST - SSM = 499.833$

$R^2 = \frac{SSM}{SST} = 78.76\%$

⑤ For X_3, both Type I SS and Type II SS are the same,

Type I = Type II = $SSM(X_3 | X_1)$

⑥ Ha: at least one of the two βs is not 0.

$F = \frac{SSM/2}{SST/27} = 50.07$ If H_0 is true, the test statistic follows a $F_{2,27}$ distribution

Since p-value $< 0.0001 < \alpha = 0.05$, reject H_0.

It's consistent with Bonferroni intervals. We can drop X_3 but we need X_1.

7. (2) The Type I and Type II partial correlations are based on the corresponding sum of squares. Compute the Type I partial correlation, fill up the SAS output, show all the details, and put your final answers in the ( ):

⑦ $0.78724 = \frac{1852.90171}{SST} = \frac{1852.90171}{2353.65367}$

Variable	DF	Squared Corr Type I	Squared Corr Type II
Intercept	1	.	.
X1	1	(0.78724)	0.78205
X3	1	(0.00183) =	0.00183

8. (2) Based on the Type I and Type II sum of squares on the previous page, if we fit a model $Y = \beta_0 + \beta_1 \times X_1 + \epsilon$, what will be its model sum of squares and R^2?

9. (2) We compute the tolerance TOL. Below is the SAS output.

Variable	DF	Parameter Estimate	Standard Error	Tolerance
Intercept	1	-124.30436	20.88642	.
X1	1	2.73731	0.27810	0.95910
X3	1	0.01907	0.08565	0.95910

What is the VIF? What is the R^2 for the model with X_3 as response and X_1 as predictor ($X_3 = \beta_0 + \beta_1 \times X_1 + \epsilon$)?

⑧ Model SS for $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ is Type I SS for X_1, equals to 1852.90171. Thus $R^2_{new} = \frac{1852.90171}{2353.65367} = 0.78724$

SST is the same for both models.

⑨ $VIF = 1/TOL = 1.042644$

For model $X_3 = \beta_0 + \beta_1 X_1 + \varepsilon$, $R^2 = 1 - TOL = 0.0409$

10. (3) We fit a model using the qualitative variable X_2 and obtain the following SAS output. The model is $Y = \beta_0 + \beta_1 \times X_4 + \beta_2 \times X_2 + \beta_3 \times X_4 \times X_2 + \epsilon$.

Variable	DF	Parameter Estimate	Standard Error
Intercept	1	-780.74431	83.75117
X4	1	199.86642	19.41391
X2	1	299.31929	251.53676
X42	1	-70.75408	58.54282

Write down the two regression models for $X_2 = 0$ and $X_2 = 1$ respectively.

11. (1) For the above model, observation 17 has residual $e_{17} = 6.7007$ and the standard error of the residual is 4.171. What is the value of the studentized residual? Is observation 17 an outlier?

12. (2) Test whether we only need two parallel lines. Write down the null and alternative hypotheses. Compute the value of the test statistic. Based on the T-table state the decision rule and your conclusion.

13. (1) Below is the correlation among all three quantitative variables. If we put all of them in the model, will there be a problem? Explain.

Pearson Correlation Coefficients

	X1	X3	X4
X1	1.00000	-0.20224	0.99977
X3	-0.20224	1.00000	-0.22334
X4	0.99977	-0.22334	1.00000

14. (2) We will try the model with all three quantitative variables $Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_3 + \beta_3 \times X_4 + \epsilon$. But we fix $\beta_1 = 5$. How can we fit this model?

10 $X_2 = 0$: $Y = -780.7 + 199.87 \times X_4$

$X_2 = 1$: $Y = -481.42 + 129.11 \times X_4$

11 studentized residual $= \dfrac{6.7007}{4.171} = 1.606 < 3$.

It is between -3 and 3. Not an outlier.

12 $H_0: \beta_3 = 0$ $H_a: \beta_3 \neq 0$ $t_{n-3-1} = t_{26} = \dfrac{-70.75408}{58.54282} = -1.21$

For $\alpha = 0.05$, $|t_{26}(0.975)| = |t_{26}(0.025)| = 2.056$

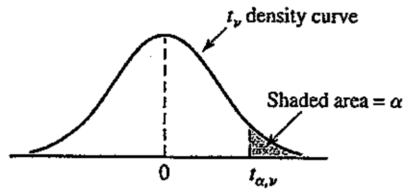
$|-1.21| < 2.056$, Fail to reject H_0. We only need two parallel lines.

(13) There will be a multicollinearity problem.
Correlation between X_1 and X_4 = 0.99977, too big.
We have to drop either X_1 or X_4.

(14) Create a new response $Y^* = Y - 5X_1$, and
fit model $Y^* = \beta_0^* + \beta_2^* X_3 + \beta_3^* X_4 + \varepsilon^*$.

Table A.5 Critical Values for t Distributions

$3.078 = t_1(0.90) = t_1(1-0.10)$



ν \ α	.10	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
32	1.309	1.694	2.037	2.449	2.738	3.365	3.622
34	1.307	1.691	2.032	2.441	2.728	3.348	3.601
36	1.306	1.688	2.028	2.434	2.719	3.333	3.582
38	1.304	1.686	2.024	2.429	2.712	3.319	3.566
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.262	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291