Statistics 104—Fall, 2004— Solutions to practice problems

11.2
a) F = 5.7143
b) 5.7143 is greater than 5.02, which is the entry for a level of 0.001
Therefore, the result is significant at both the 5% and 1%, as well as at the 0.1% level.

Source	df	SS	MS	F
Model	4	20	2.857	5.7143
Error	100	200	0.5	
Total	104	220		

11.4
As a group, the 10 quizzes are useful in predicting final exam scores. However, no single quiz is particular useful.

11.32
Correlations below. The value for IQ is largest in absolute value, so the relationship is closest to a straight line. About 40.2% of the variation in GPA would be explained by IQ.

	GPA
IQ	0.634
Age	-0.389
Sex	-0.097
SC	0.542
C1	0.441
C2	0.601
C3	0..495
C4	0.267
C5	0.472
C6	0.401

11.33
a) R^2 = 45.9%. The t stat for C3 is 2.83 and the p-value is 0.006. Thus, C3 contributes significantly. C3 increase R^2 by 5.7% = 45.9% - 40.2%.

```
The regression equation is
GPA = -2.83 + 0.0822 IQ + 0.163 C3
```

Predictor	Coef	StDev	T	P
Constant	-2.829	1.507	-1.88	0.064
IQ	0.08220	0.01508	5.45	0.000
C3	0.16289	0.05752	2.83	0.006

```
S = 1.564   R-Sqd = 45.9%   R-Sqd (adj) = 44.5%
```

Analysis of variance

Source	DF	SS	MS	F	P
Regression	2	155.943	77.971	31.87	0.000
Residual Error	75	183.484	2.446		
Total	77	339.427			

b) Now, R^2 = 47.5%, which is only a small increase. From the t stats we can see that neither C3 nor SC are significant.

The regression equation is
GPA = -3.49 + 0.0761 IQ + 0.0670 C3 + 0.0369 SC

Predictor	Coef	StDev	T	P
Constant	-3.491	1.558	-2.24	0.028
IQ	0.07612	0.01549	4.91	0.000
C3	0.06701	0.08558	0.78	0.436
SC	0.03691	0.02456	1.50	0.137

S = 1.551 R-Sqd = 47.5% R-Sqd (adj) = 45.4%

Analysis of variance

Source	DF	SS	MS	F	P
Regression	3	161.378	53.793	22.36	0.000
Residual Error	74	178.049	2.406		
Total	77	339.427			

c) The values change because coefficients are quite sensitive to changes in the model, especially when the explanatory variables are highly correlated. (The correlation between SC and C3 is 0.80). In this case, the predictive information of SC and C3 overlap, so that the two of them together add little more than either one separately.

11.34

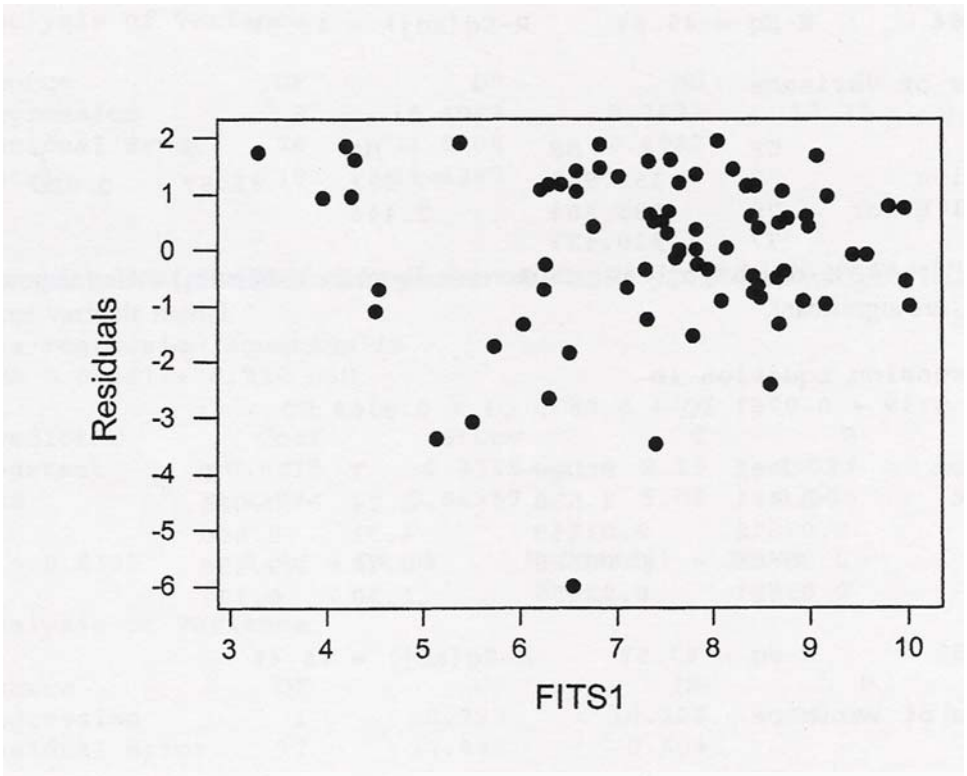
a) GPA = -4.94 + 0.0815 IQ + 0.183 C1 + 0.142 C5.

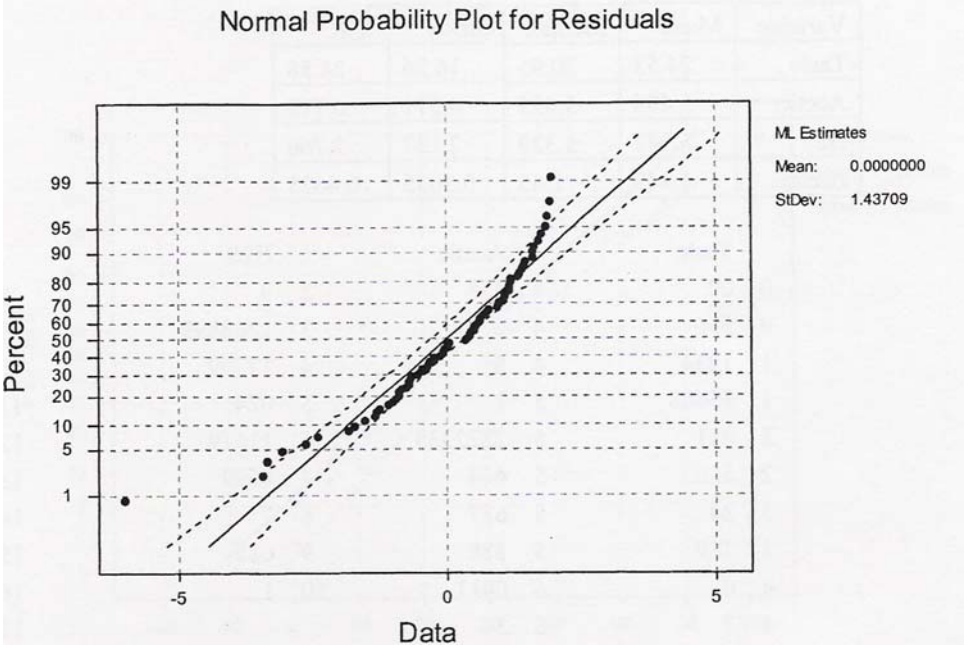
R^2 = 52.5% and s = 1.475. The predicted value is 7.457.

b) The increase would be 0.08145. The interval is 0.0543 to 0.1087.

```
The regression equation is
GPA = -4.94 + 0.0815 IQ + 0.183 C1 + 0.142 C5

Predictor   Coef            StDev                  T        P
Constant    -4.937          1.491                  -3.31    0.001
IQ          0.08145         0.01367                5.96     0.000
C1          0.18308         0.06475                2.83     0.006
C5          0.14205         0.06663                2.13     0.036

S = 1.475   R-Sqd = 52.5%   R-Sqd (adj) = 50.6%

Analysis of variance

Source            DF      SS        MS        F      P
Regression        3       178.340   59.447    27.31  0.000
Residual Error    74      161.087   2.177
Total             77      339.427
```

c) The residual for OBS 55 stands out. This student had the lowest GPA and was the oldest.





d) The equation is now GPA = -4.68 + 0.0805 IQ + 0.197 C1 + 0.109 C5.
R2 = 57.4%. and s = 1.303. The new predicted value is 7.534. Removing this observation did not change the model or the prediction greatly, but now C5 is just barely not significant.

```
The regression equation is
GPA = -4.68 + 0.0805 IQ + 0.197 C1 + 0.109 C5

Predictor   Coef           StDev                 T       P
Constant    -4.678         1.318                 -3.55   0.001
IQ          0.08050        0.01207               6.67    0.000
C1          0.19707        0.05724               3.44    0.001
C5          0.10950        0.05923               1.85    0.069

S = 1.303   R-Sqd = 57.4%   R-Sqd (adj) = 55.7%

Analysis of variance

Source            DF      SS        MS        F      P
Regression        3       167.112   55.704    32.83  0.000
Residual Error    73      123.855   1.697
Total             76      290.967
```

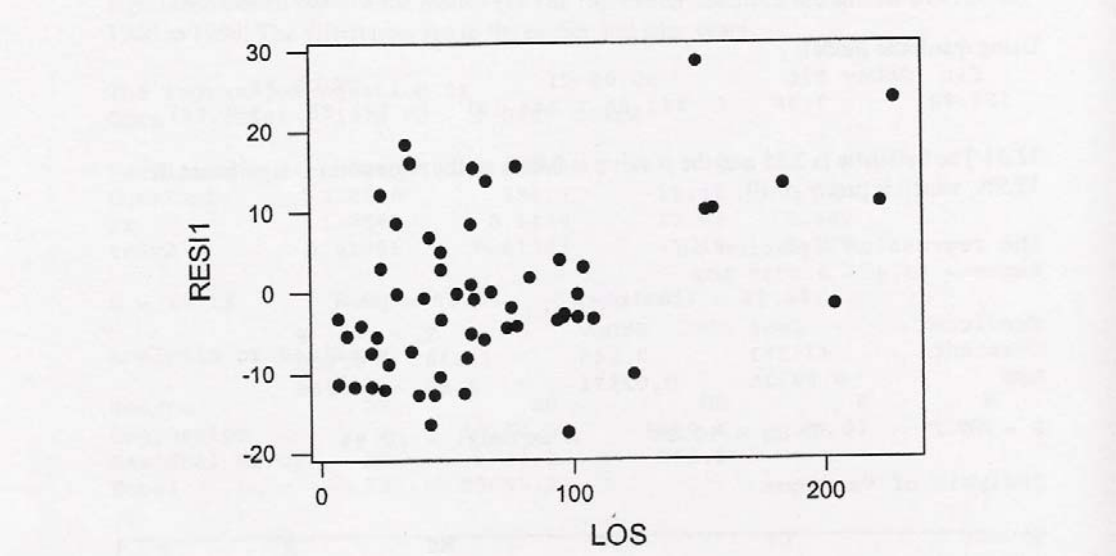
11.52

a) The t statistic is 2.96 and the p-value is 0.004, so the regression is significant. R^2 = 13.4%, which is pretty small.

```
The regression equation is
Wages = 44.0 + 7.93 Size

Predictor   Coef           StDev                 T       P
Constant    43.974         2.032                 21.64   0.000
Size        7.934          2.677                 2.96    0.004

S = 10.16   R-Sqd = 13.4%   R-Sqd (adj) = 11.8%

Analysis of variance

Source            DF      SS        MS        F      P
Regression        1       906.9     906.9     8.78   0.004
Residual Error    57      5885.6    103.3
Total             58      6792.5
```

b) The t stat for the two sample t test is the same, as are df. Note that MSE is the same as the pooled estimate of variance. Because the Size variable takes on only two values, the slope gives the difference in LOS for the two bank sizes. This is the same as the two sample t test.

c) There appears to be positive association between the residuals and LOS. Including LOS in the model could give better results.



11.53

The F stat is 11.50 and the p-value < 0.001, so the regression is significant. Both t statistics are significant, so both variables contribute to the model. R^2 = 29.1% which is quite a bit higher than with either variable alone.

```
The regression equation is
Wages = 37.6 + 0.0829 LOS + 7.93 Size

Predictor    Coef            StDev                  T        P
Constant     37.565          2.596                  14.47    0.000
LOS          0.08289         0.02349                3.53     0.001
Size         8.916           2.459                  3.63     0.001

S = 8.916    R-Sqd = 29.1%   R-Sqd (adj) = 26.6%

Analysis of variance

Source            DF      SS        MS        F       P
Regression        2       1977.67   988.83    11.50   0.000
Residual Error    56      4814.85   85.98
Total             58      6792.51
```

12.6

a) Yes, the ratio of the largest to the smallest is 62/40<2.
b) The squares of the standard deviations are 3844, 1600, 2704, and 2304.
c) 501,660/269 = 1864.91
d) 43.1846
e) The second sample size is much larger than the others.

12.10

a) H_0: $\mu_1 = \mu_2 = \ldots = \mu_I$ vs. H_a: not all μ_I equal.

b)

Source	DF	SS	MS	F
Groups	3	SSG	SSG/3	
Error	196	SSE	SSE/196	
Total	199			

c) F distribution with df = (3,196)

d) Table for df = (3,200) = 2.65

12.32

a)

Source	DF	SS	MS	F
Groups	3	104855.87	34951.96	15.8646
Error	32	70500.59	2203.143	
Total	35	175356.46		

b)H_0: $\mu_1 = \mu_2 = \ldots = \mu_I$ vs. H_a: not all μ_I equal.

c) F distribution with df = (3,32). With df = (3,40) and $\alpha = 0.001$, Table E says 7.05. 15.86 is much larger than any of these, so the p-value is much less than 0.001. We conclude that not all the means are equal.

d) The estimate of pooled variance is the MSE = 2203.143.

Pooled standard deviation is the square root or 46.938.

12.36

a) $\psi_1 = (\mu_1 + \mu_2)/2 - \mu_3$

b) $\psi_2 = (\mu_1 - \mu_2)$

12.38

a) For $\psi_1 = (\mu_1 + \mu_2)/2 - \mu_3$

H_0: $\psi_1 = 0$ vs. H_a: $\psi_1 > 0$, because science majors might have higher SATM scores. For $\psi_2 = (\mu_1 - \mu_2)$, H_0: $\psi_2 = 0$ vs. H_a: $\psi_2 \neq 0$ because we have no prior expectations of the direction of the difference.

b) c1 = (619 + 629)/2 – 575 = 49 and c2 = (619 – 629) = -10

c) SE(c1) = $82.5\sqrt{0.25/103 + 0.25/31 + 1/122} = 11.28$

SE(c2) = $82.5\sqrt{1/103 + 1/31 + 0/122} = 16.90$

d) t1 = 49/11.28 = 4.344 (df = 252) and p-value very small.

We conclude that science majors have a higher mean SATM than other majors.

Then t2 = -10/16.90 = -0.5916 (df = 253). This is not significant. The difference in the mean SATM scores for computer science majors vs. other science majors is not significant.

e) Use t* = 1.984 (df = 100 from table). For ψ_1 this gives 26.6 to 71.4.

For ψ_2 -43.5 to 23.5