Stat 13

http://www.stat.ucla.edu/~dinov/courses_students.html

Chapter 12 Problems/Solutions

All Problems are from: Myra L. Samuels and Jeffrey A. Witmer,
Statistics for the Life Sciences, 3rd edition, Prentice-Hall (2003)

12.5. SOCR (http://www.socr.ucla.edu) output is given below.

(a) cob-wt = 316 - 0.721 plant-density

(b) scatterplot (not shown) shows a strong negative linear association between cob-weight (gm grain/cob) and plant density (# plants / pot).

(c) as plant density increases by 1 plant per plot, cob weight decreases by 0.72 gm of grain per cob, on average.

(d) sY = sqrt(11831.8/19) = 25 gm and sY/X = sqrt(1337.3/18) = 8.6 gm

(e) Predictions of cob weight based on the regression model tend to be off by 8.6 gm on average.

Equivalently, the data points deviate above or below the regression line by 8.6 gm on average.

Regression Analysis

The regression equation is

cob-wt = 316 - 0.721 plant-density

Predictor	Coef	StDev	T	P
Constant	316.376	8.000	39.55	0.000
plant-de	-0.72063	0.06063	-11.89	0.000

S = 8.619 R-Sq = 88.7% R-Sq(adj) = 88.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	10495	10495	141.26	0.000
Residual Error	18	1337	74		
Total	19	11832			

12.6

(a) The slope and intercept of the regression line are

b_1 = -927.75/1303 = -.7120;
b_0 = y-bar - b_1x-bar = 23.64 - (-.7120)(11.5) = 31.83

The fitted regression line is y-hat = 31.83 - .7120X

(c) s_Y|X = sqrt[SS(resid)/df] = sqrt[16.7812/(12-2)] = 1.3

12.8a: b1 = 161.40/50667 = .0003186; b0 = .210 – (.0003186)(433.3) = .0720.

The fitted regression line is Y = .072 - .0003186X.

12.8c: As altitude of origin goes up by 1 m, respiration rate goes up by .0003186 mul/hr-mg, on average.

12.8d: s_{Y|X} = sqrt[.013986/10] = .0374

12.12

The intercept of the regression line b0 is based on all 12 data points, not just on the two point for which X = 0. If there is a linear relationship between X and Y (a scatter plot of the data strongly suggest that there is), then the best estimate of the average for Y at any given value of X is given by the regression line, taking into account all of the data. In contrast, the average (33.3 + 31.0)/2 = 32.15 ignores most of the data.

12.14a: (See Exercise 12.5 for b0 and b1.

(i) plugging in plant-density = 100 plants gives a predicted cob-wt of 316 - 0.721(100) = 244.34 gm

(ii) plugging in plant-density = 120 plants gives a predicted cob-wt of 316 - 0.721(120) =229.93 gm

12.14b:

(i) (224.34)(100) = 24434 gm = 24.43 kg
(ii) (229.928)(12) = 27591 gm = 27.6 kg

12.15

Using the fitted regression line found in Exercise 12.6 above, we substitute X = 15. This yields y-hat = 31.83 - (.7120)(15) = 21.1.

Thus, we estimate that the mean fungus growth would be 21.1 mm at a laetisaric acid concentration of 15 microg/ml.

According to the linear model, the standard deviation of fungus growth does not depend on X. Our estimate of this standard deviation from the regression line is the Residual Standard Deviation sigma_{Y|X} = sqrt[SS(resid)/(n-2)] = sqrt[16.7812/10] = 1.3 mm.

Thus we estimate that the standard deviation of fungus growth would be 1.3 mm at a laetisaric acid concentration of 15 microg/ml.

For X = 15, we have y-hat = 21.1 +/- 1.3 mm.

12.19a: b1 = 81.90/2800 = 0.02925 ng/min (the rate of incorporation)
b0 = 0.83 – (0.02925)(30) = -0.05

s_{y|x} = sqrt[SS(resid)/(n-2)] = sqrt[0.035225/5] = 0.0839

To construct a 95% confident interval, we consult the z-table (Table 4) with df = n-2 = 7-2 =5; the multiplier is t_{4,0.025} = 2.571. The confidence interval is

b1 +/- t_{4,0.025}SEb1 = 0.02925 +/- (2.571)(0.00159)

or 0.0252 < beta1 < 0.033 ng/min

12.19ba: We are 95% confident that the rate at which leucine is incorporated into protein in the population of all Xenopus oocytes is between 0.0252 ng/min and 0.0333 ng/min

12.21a: SEb1 = 8.6/sqrt(20209) = 0.0605, so 95% CI for b1 is

-0.7206 +/- (2.101)(0.0605) or -0.7206 +/- 0.1271 or (-0.848 , -0.593)

12.21b: We are 95% confident that as plant density increases by 1 plant per plot, average cob weight decreases by between 0.848 gm and 0.593 gm of grain per cob.

12.22a: From Exercise 12.6, s_Y|X = sqrt[SS(resid)/df] = sqrt[16.7812/(12-2)] = 1.3

The standard error of the slope is

SE_b1 = s_Y|X / sqrt[sum(x - x-bar)^2] = 1.3/sqrt[1303] = 0.36

12.22b: H0: Leatisaric acid has no effect on fungus growth (beta_1 = 0)

HA: Laetisaric acid inhibits fungus growth (beta_1 < 0)

t_s = -.7120/0.36 = -19.8. With df = 10, the t-table (Table 4) gives t_.0005 = 4.587. Thus the P-value < .0005, so we reject H0. There is sufficient evidence (P-value < .0005) to conclude that laetisaric acid inhibits fungus growth.

12.27a: r = 82.8977/sqrt[(28465.7)(.363708)] = .8147

12.27b: s_Y = sqrt[(.363708/(13-1)] = .1741 gm

s_Y|X = sqrt[SS(resid)/df] = sqrt[.1223/(13-2)] = .1054 gm

.1054/.1741 = .605; sqrt[1 - .8147^2] = .580

12.27c: b_1 = 82.8977/28465.7 = .002912;

b_0 = 2.174 - (.002912)(443.8) = .882

The fitted regession line is y-hat = .882 - .002912X

12.28a: r = -14563.1/sqrt[(20209)*(11831.8)] = -0.942

12.28b: from Exercise 12.5 (d), sY = 25 gm and sY/X = 8.6 gm, so sY/X / sY = 0.344 further, sqrt(1 - r2) = 0.3356, which is nearly equal to 0.344, so the approximate relationship is indeed verified.

12.28c: b1 = -14563.1/20209.0 = -.7206;

b0 = 224.1 – (-.7206)(128.05) = 316.4

The fitted regression line is Y = 316.4 - .7206X.

12.30: Let X = age and let Y = blood pressure. The Residual Standard Deviation is s_{Y|X} = sqrt[1 - r^2](s_Y)sqrt[(n-1)/(n-2)] = sqrt[1 - .43^2](19.5)sqrt[2668/2667] = 17.6 mm Hg.

s_{Y|X} = sqrt[(y - y-hat)^2/(n-2)] is a measure of the variability about the regression line y-hat = b1x + b0.

But s_Y = sqrt[(y - y-bar)^2/(n-1)] is a measure of the variability about the mean y-bar.

12.41a: with (iii),
12.41b: with (ii), and
12.41c: with (i).

12.45a: The slope and intercept of the regression line are

b1 = -.342/.1512 = -2.262
b0 = 1.117 – (-2.262)(.12) = 1.39

The fitted regression line is Y = 1.39 – 2.262X.

12.45c: s_{Y|X} = sqrt[SS(resid)/(n-2)] = sqrt[.2955/10] = .1719 kg.

12.46a: If x = .24, then predicted y = 1.39 – 2.262(.24) = .84512. But the variability of this prediction is given by s_{Y|X} = .17.

If x is unknown, then the best prediction is y-bar = 1.117, and the precision of this prediction is +/- s_Y = .31175. We write y = 1.117 +/- .31175 kg.

However, if x = .24 is known, then the best prediction for y is given by the regression line y = 1.39 – 2.262(.24) = .84512, but the precision of this prediction is +/- s_{Y|X} = +/- .17. We write y = .84512 +/- .17 kg.

12.46b: The condition that sigma_{Y|X} does not depend on X appears to be doubtful. Rather, the scatterplot shows that there is more variability in Y when X is small than when X is large.

X	SD
.00	.21
.06	.28
.12	.11
.30	.06

12.47: The hypotheses are
H0: sulfur dioxide has no effect on yield (beta1 = 0) and
HA: Increasing sulfur dioxide tends to decrease yield (beta1 < 0).

The sample slope is b1 = -.342/.1512 = -2.262

We note that b1 < 0, so the data do deviate from H0 in the direction specified by HA.

The residual standard deviation is s_{Y|X} = sqrt[SS(resid)/(n-2)] = sqrt[.2955/10] = .1719 kg.

The standard error of the slope is SE_{b1} = .1719/sqrt[.1512] = .4421.

The test statistic is ts = (b1 – 0)/SE_{b1} = -2.261/.4421 = -5.12.

Consulting Table 4 with df = n – 2 = 10, we find that P-value < .0005, so we reject H0.

There is strong evidence (P-value < .0005) to conclude that increasing sulfur dioxide tends to decrease yield.

12.54: SE_{b1} = s_{Y|X}/sqrt[n-1]s_X = .0374/sqrt[506667] = .0000525

So 95% CI is .0003186 +/- (2.228)(.0000525) (df=10)
or (.00020,.00044) or .00020 < beta1 < .00044.

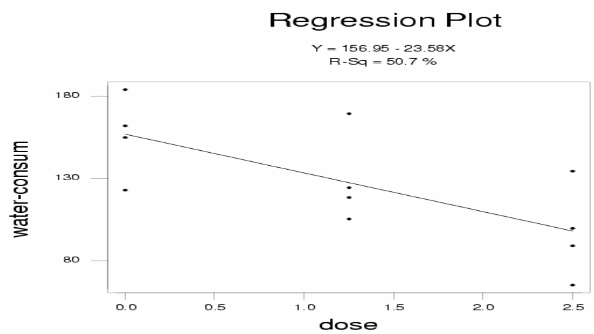
12.59

Regression Analysis

```
The regression equation is
water-consumption = 157 - 23.6 dose

Predictor        Coef       StDev          T        P
Constant       156.95       11.87      13.22    0.000
dose          -23.580       7.358      -3.20    0.009

S = 26.01       R-Sq = 50.7%     R-Sq(adj) = 45.7%

Analysis of Variance

Source             DF          SS          MS         F        P
Regression          1      6950.2      6950.2     10.27    0.009
Residual Error     10      6766.7       676.7
Total              11     13716.9
```

12.59a: b0 = 156.95; b1 = -23.580 (from SOCR, http://www.socr.ucla.edu printout)
The fitted regression line is y-hat = 156.95 – 23.580x

12.59b:



12.59d: H0: Amphetamine dose has no affect on water consumption (beta1 = 0)
HA: Increasing amphetamine dose tends to reduce water consumption (beta1 < 0)

ts = -3.20 (SOCR, http://www.socr.ucla.edu printout) and the P-value = 0.009/2 = 0.0045. Thus we reject H0.

There is strong evidence (P-value = 0.0045) to conclude that increasing amphetamine dose tends to reduce water consumption.

12.59e: **One-way Analysis of Variance**

```
Analysis of Variance for water-co
Source      DF         SS         MS        F        P
dose         2       6972       3486     4.65    0.041
Error        9       6745        749
Total       11      13717
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level        N       Mean      StDev  --+---------+---------+---------+---
-
0.00         4     156.00      25.32                   (--------*-------)
1.25         4     129.38      27.85           (--------*--------)
2.50         4      97.05      28.84   (--------*--------)
                                      --+---------+---------+---------+---

-
Pooled StDev =     27.38               70       105       140       175
```

H0: The three doses produce the same mean water consumption level (mu1 = mu2 = mu3)
HA: The mean water consumption levels are not all equal (the mu's are not all equal)

Fs = 4.65 (SOCR, http://www.socr.ucla.edu printout) and the P-value = 0.041. Note HA cannot be directional because there are three doses). Thus we reject H0.

The conclusion here is similar to that in part (d), in that we reject H0. However, the analysis from (d) gave a smaller P-value, as it made use of the fact that the means are not only different, but they decrease as dose increases.

12.59f: The analysis in part (d) requires linearity; that is, the mean water consumption levels must have a linear relationship to dose for the regression model to make sense. The ANOVA in part (e) does not require this condition.

12.59g: s_{pooled} = 27.38 (from ANOVA printout), which is similar to $s_{Y|X}$ = 26.01 (from regression printout)