Brief Solutions – Homework 2 – STAT 516 – Spring 2016

**1.** (a)From Homework 1,Syy = 42 – 142/5 = 2.8, Sxy = -5 – (0)(14)/5 = -5, Sxx = (10) – 02/4 = 10 for these data. So r = -5/[(10)(2.8)]1/2 = – 0.945. This tells us that there is a *strong, negative linear* relationship between Oxidation and Temperature.

(b) Note TSS = Syy = 2.8 and from Homework 1, SSE = 0.3, so r2 = 1 – (0.3/2.8) = 0.893. Or simply r2 = (– 0.945)2 = 0.893. So around 89% of the sample variability in oxidation can be explained by its linear relationship with temperature.

**2.** (a) Estimated regression equation:  We *estimate* that the *expected* goals made increases by 1.203 goals for each one-inch increase in the player’s height, adjusting for weight and 100-yard dash in the model (or, in the presence of the other independent variables, or, holding constant the other independent variables).

(b) Based on the P-values, using  = 0.05 in each case, we conclude: (i) Weight (t\*= –2.1, p-value=.0487) is (marginally) significantly related to goals made, in the presence of Height and Dash100 in the model; (ii) Height (t\*=19.2, p-value<.0001) is highly significantly related to goals made, in the presence of Weight and Dash100 in the model; (iii) Dash100 (t\*= –0.66, p-value=.5194) is not significantly related to goals made, in the presence of Height and Weight in the model.

(c) For this F-test, SAS gives the F-statistic as 2.43 (note F0.05 with df = 2, 21 is 3.47) and the P-value as 0.1128. Based on this P-value, using  = 0.05 we would fail to reject H0: 1 =  3 = 0. We conclude Weight and Dash100 may be useless in the model, given that Height is in the model.

(d) R2 = 0.9478. So about 94.8% of the sample variability in goals made can be explained by its linear relationship with weight, height, and 100-yard dash time.

**3.** (a) There appears to be model misspecification (nonlinearity), and one severe outlier, based on the curvature in the residual plot. The normality assumption is also somewhat questionable based on the Q-Q plot.

(b) The residual plot looks much better, like random scatter. The normality assumption is still a bit iffy, but possibly plausible, based on the Q-Q plot.

(c) From the SAS results, the requested 90% prediction interval for the log survival time is (4.3854, 4.7586). Reversing the log transformation with the exponential function will produce a 90% prediction interval for the survival time: (e4.3854, e4.7586) = (80.27, 116.58).

**4.**  2-false (The slope indicates this.), 3-false (only Y need be normally distributed)

13-true, 17-true, 19-false (It will only detect only a linear relationship.),

20-false (We can’t assume causation, merely a strong linear association between the variables.)

**5.** #1: TSS = SSR+SSE = 50+100 = 150. R2 = SSR/TSS = 50/150 = 0.33.

#3: (a) Note from p. 408, F = [(n-m-1) R2 ]/[m(1- R2)]. Since F(5,24,0.05)=2.62, and n=30, m=5, we want 24R2/[5(1-R2)] > 2.62. Solving for R2, this implies R2 > 0.353.

(b) Same as above, but n=500. We want 494R2/[5(1-R2)] > 2.25. Solving for R2, this implies R2 > 0.022.

(c) This tells us you don’t need a very large R2 to get a significant regression if you have a lot of observations.

#4: The regression coefficient is measured in pounds (or precisely, pounds per inch).

/\* Example SAS code \*/

/\* problem 2 \*/

data basket;

input id height goalmade dash100 weight;

cards;

01 71 15 11.50 130

02 74 19 12.23 149

<omitted datalines>

24 74 19 11.98 188

25 70 13 12.23 231

;

run;

PROC REG data=basket;

MODEL goalmade = WEIGHT HEIGHT DASH100;

TEST weight=0, dash100=0;

/\* This TEST statement is for part (c) \*/

run;

/\* Problem 3 \*/

data livertrans;

input Obs CLOT PROG ENZ LIV TIME;

LOGTIME = log(TIME);

cards;

1 3.7 51 41 1.55 34

2 8.7 45 23 2.52 58

<omitted datalines>

54 5.8 96 114 3.95 830

;

run;

/\* produces plots on left below \*/

PROC REG DATA=livertrans;

MODEL time = CLOT PROG ENZ LIV / P R;

OUTPUT OUT=NEW P=PRED R=RES;

PROC SGPLOT DATA=NEW;

SCATTER y=RES x=PRED;

REFLINE 0;

PROC UNIVARIATE noprint ;

QQPLOT RES / normal;

RUN;

/\* transformed model \*/

/\* produces plots on right below \*/

PROC REG DATA=livertrans;

MODEL logtime = CLOT PROG ENZ LIV / P R;

OUTPUT OUT=NEW P=PRED2 R=RES2;

PROC SGPLOT DATA=NEW;

SCATTER y=RES2 x=PRED2;

REFLINE 0;

PROC UNIVARIATE noprint ;

QQPLOT RES2 / normal;

RUN;

/\* Problem 3(c) \*/

data xvalues;

input Obs CLOT PROG ENZ LIV TIME;

cards;

. 5.5 57 62 2.63 .

;

data livertrans;

set livertrans xvalues;

run;

PROC REG DATA=livertrans;

MODEL logtime = CLOT PROG ENZ LIV / cli alpha=0.10;

run;





Total points: **1**(a) 6 (b) 5; **2**(a) 5 (b) 6 (c) 4 (d) 4; **3** (a) 4 (b) 4 (c) 5; **4** 12 pts (2 each); **5** #1:3 #3 (a) 2 (b) 2 (c) 2 #4: 2. Plus 5 points for neatness.



