STAT 512: Homework 2
Name: Lei Nie

1. *KNNL Problem 2.17*

   The α level used by analyst is greater than 0.033. The proper conclusion should be: We don't have enough evidence to prove that there is a significant linear relationship between x *and* y.

2. *KNNL Problem 2.22*

   R^2 is the percentage of total variance that can be explained by the independent variables. It is possible that the R^2 for first ten cases being zero and R^2 for complete 30 cases not being zero. In that case, the variance of y can only be partially explained by the later 20 cases. It is also possible that the R^2 for first ten cases not being zero but R^2 for complete 30 cases being zero. For instance, the residual plot for the first ten cases shows some pattern while the residual plot for all 30 cases shows no pattern as all.

3. Calculate power for the slope using the results of text Problem 1.22 as follows. Assume n = 16, $\sigma^2 = MSE$, and $SS_X = 1280$.

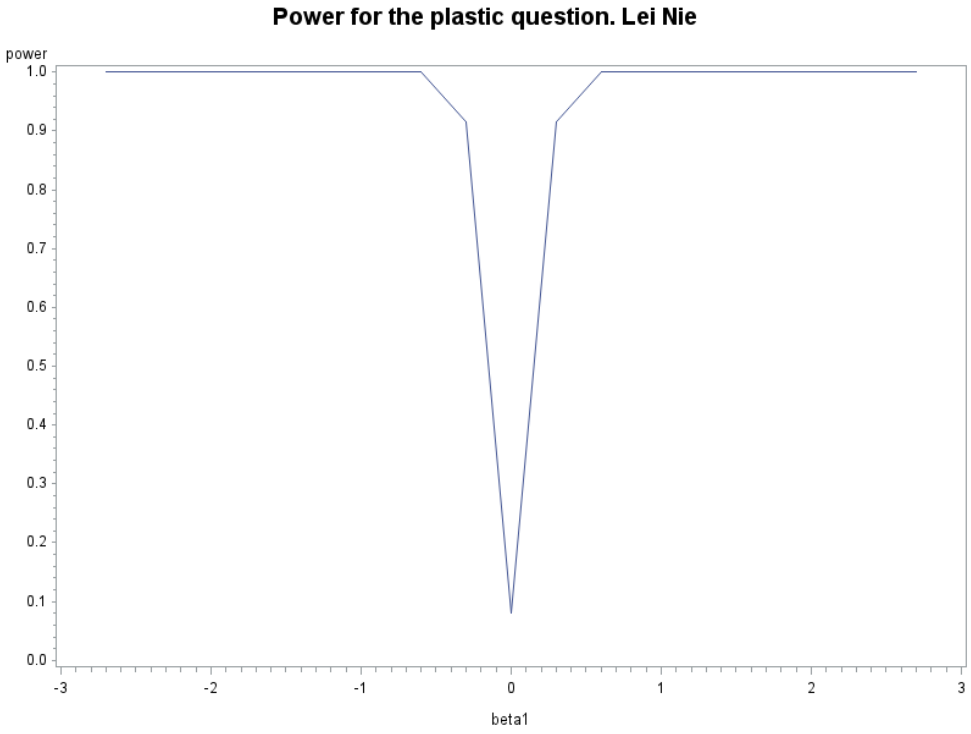
   a. *Find the power for rejecting the null hypothesis that the regression slope is zero using* $\alpha = 0.08$ *significance test when the alternative is* $\beta_1 = 0.9$.

Obs	**n**	**sig2**	**ssx**	**alpha**	**sig2b1**	**df**	**beta1**	**delta**	**tstar**	**power**
1	16	10.4589	1280	0.08	.008171039	14	0.9	9.95644	1.88750	1

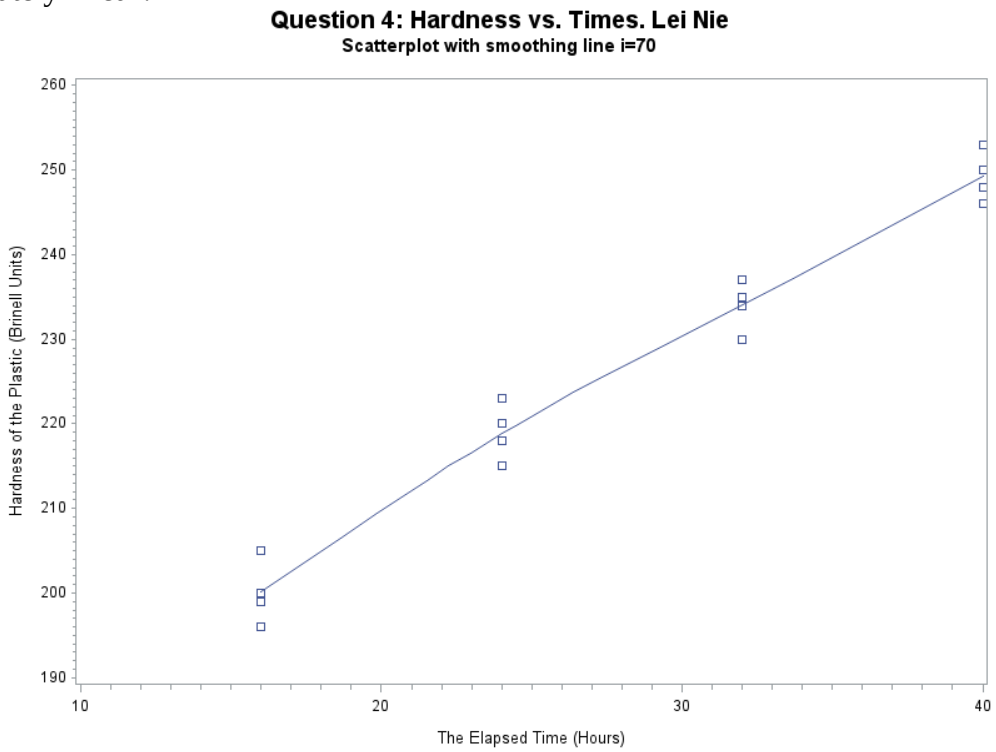
   According to SAS output, the power is 1.

   b. *Plot the power as a function of* β_1 *for values of* β_1 *between -2.7 and +2.7 in increments of 0.3.*

   See the figure below.

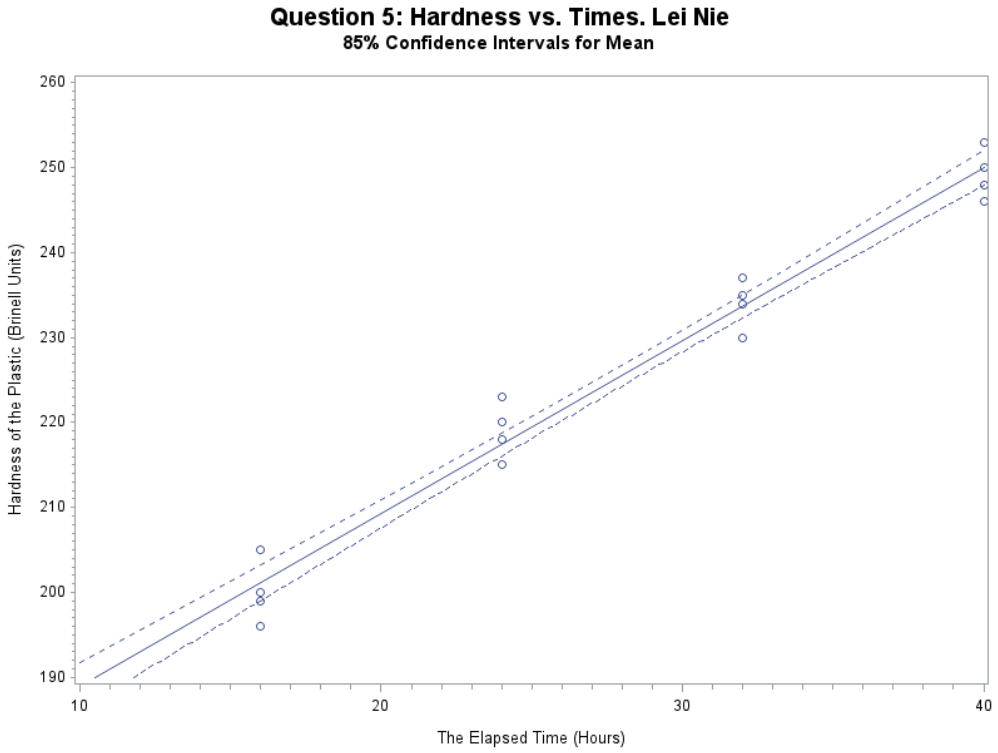


4. Plot the data using proc gplot. Include a smoothed function on the plot by using the i = smnn option on the symbol1 statement, where nn is a number between 1 and 99. Is the relationship approximately linear?

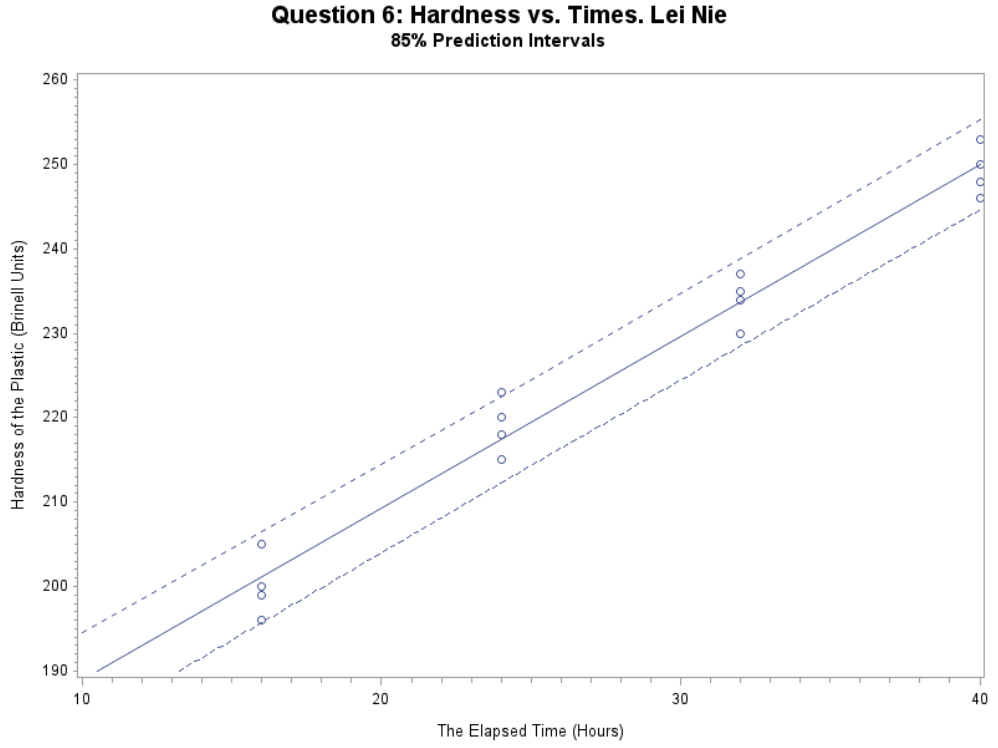


The relationship seems approximately linear.

5. Plot the 85% bounds (confidence band) for the mean (use i=rlclm on the symbol1 statement).



6. Plot the 85% bounds for individual observations (using i=rlcli).



7. Give an estimate of the mean hardness that you would expect after 37 and 41 hours; and an 85% confidence interval for each estimate. Which confidence interval is wider and why is it wider?

Obs	time	Dependent Variable	Predicted Value	Std Error Mean Predict	85% CL Mean		85% CL Predict		Resi dual
17	**37**	.	243.8719	1.1470	242.1249	245.6188	238.6455	249.0982	.
18	**41**	.	252.0094	1.4264	249.8368	254.1819	246.6258	257.3929	.

The SAS output shows that

- After 37 hours, the mean hardness is estimated to be 243.8719;
- After 41 hours, the mean hardness is estimated to be 252.0094;
- 85% confidence interval for the mean after 37 hours is (242.1249, 245.6188);
- 85% confidence interval for the mean after 41 hours is (249.8368, 254.1819);

The confidence interval for the hardness after 41 hours is wider because 41 is further away from the sample mean and, thus, has a larger standard error for prediction.

8. Give a prediction for the hardness that you would expect for an individual piece of plastic after 41 hours; give an 85% prediction interval for this quantity.

   The SAS output shows that, after 41 hours, the mean hardness is predicted to be 252.0094. The 85% confidence interval for the prediction is (246.6258, 257.3929).

9. Given that $R^2 = SSM/SST$, it can be shown that $\frac{R^2}{1-R^2} = \frac{SSM}{SSE}$. If you have n = 25 cases and $R^2 = 0.35$, what is the F-statistic for the test that the slope is equal to zero?

$$F\ statistic = \frac{MSM}{MSE} = \frac{SSM/DFM}{SSE/DFE} = \frac{R^2}{1-R^2} * \frac{23}{1} = \frac{0.35}{0.65} * 23 = 12.3846$$

Hypothesis: $H_0: \beta_1 = 0\ vs. H_a: \beta_1 \neq 0$

Using $\alpha = 0.05$, $F_{1,23} = t_{23}^2 = 4.2808 < F\ statistic$. With 95% confidence, we reject the null hypothesis and conclude that the slope is not equal to zero.

Appendix: SAS code

```
*Question 3;
data plastic;
input hardness time @@;
datalines;
  199.0   16.0
  205.0   16.0
  196.0   16.0
  200.0   16.0
  218.0   24.0
  220.0   24.0
  215.0   24.0
  223.0   24.0
  237.0   32.0
  234.0   32.0
  235.0   32.0
  230.0   32.0
  250.0   40.0
  248.0   40.0
  253.0   40.0
  246.0   40.0
  ;
proc print data=plastic;run;
/*Use proc univariate to calculate
ssx*/;
proc univariate data = plastic;
var time;run;
/*Use proc reg to calculate mse*/;
proc reg data=plastic;
model hardness=time;run;

*Question 3a: Calculate power;
data a1;
n=16; sig2=10.45893; ssx=1280;
alpha=.08;
sig2b1=sig2/ssx; df=n-2;
beta1=0.9;
delta=abs(beta1)/sqrt(sig2b1);
tstar=tinv(1-alpha/2,df);
power=1-
probt(tstar,df,delta)+probt(-
tstar,df,delta);
output;
proc print data=a1;run;

*Question 3b;
data a2;
n=16; sig2=10.45893; ssx=1280;
alpha=.08;
sig2b1=sig2/ssx; df=n-2;
do beta1=-2.7 to 2.7 by .3;
delta=abs(beta1)/sqrt(sig2b1);
tstar=tinv(1-alpha/2,df);
power=1-
probt(tstar,df,delta)+probt(-
tstar,df,delta);
output;end;
proc print data=a2;run;
title1 'Power for the plastic
question. Lei Nie';
symbol1 v=none i=join;
proc gplot data=a2; plot
power*beta1; run;

*Question 4;
proc sort data=plastic; by time;
symbol1 v=square i=sm70;
title1 'Question 4: Hardness vs.
Times. Lei Nie';
title2 'Scatterplot with smoothing
line i=70';
axis1 label=('The Elapsed Time
(Hours)');
axis2 label=(angle=90 'Hardness of
the Plastic (Brinell Units)');
proc gplot data=plastic;
plot hardness*time/ haxis=axis1
vaxis=axis2;run;

*Question 5;
symbol1 v=circle i=rlclm85;
title1 'Question 5: Hardness vs.
Times. Lei Nie';
title2 '85% Confidence Intervals
for Mean';
proc gplot data=plastic;
plot hardness*time/ haxis=axis1
vaxis=axis2;run;

*Question 6;
symbol1 v=circle i=rlcli85;
title1 'Question 6: Hardness vs.
Times. Lei Nie';
title2 '85% Prediction Intervals';
proc gplot data=plastic;
plot hardness*time/ haxis=axis1
vaxis=axis2;run;

*Question 7;
data a1;
time=37; output;
time=41; output;
data plastic1;
set plastic a1;
proc reg data=plastic1;
model hardness=time/cli clm
alpha=0.15;id time;run;
```