

This examination is open book and open notes, and you may use a calculator. There are 10 problems, all pertaining to the same scenario given below. The data are provided in the SAS program `golf.sas`, along with some corresponding analyses. In your problem solutions, please make clear reference to any information used from the SAS output—refer to the SAS output by page number, and circle information used from the output and label it by problem number. (The SAS output is not complete.)

All problem parts will count equally. Show the details of your work.

**Scenario:** In order to compare the durability of four different brands of golf balls, the United States Hackers Association randomly selects five balls of each brand and places each ball into a machine that exerts the force produced by a 180-yard drive. The number of simulated drives needed to crack or chip each ball is recorded. The results of such a test are included in the program `golf.sas`, provided with this examination.

1. The SAS program includes a plot of the raw data. Based on this plot, do you anticipate any significant differences between the brands? Discuss.
2. State the usual model for a one-way analysis. Using the residual plots provided, check as many of the model assumptions as possible.
3. State the hypotheses and decision rule and conduct the test for equality of variances using a significance level of 5%.
4. Test the normality assumption, computing the observed significance level of the test and interpreting the results.
5. What percentage of the variability in the number of drives needed to crack or chip each ball is explained by the model? What is the usual name for this percentage?
6. Test whether or not mean response is the same for all four brands. State the hypotheses, test statistic and decision rule, compute the observed significance level of the test, and state the results.
7. Use Tukey's method of multiple comparisons and a 95% joint confidence level to compare the brands. Is one brand better than all others? Discuss.
8. Brands  $A$  and  $B$  are produced by one manufacturer and brands  $C$  and  $D$  by another. Give the contrast  $L$  which compares average mean response for brands  $A$  and  $B$  to average mean response for brands  $C$  and  $D$ , and construct an individual 95% confidence interval for  $L$ .
9. Use Satterthwaite's approximation to compute a 95% confidence interval for the contrast  $L$  of problem 8. Comparing the results of problems 8 and 9, discuss desirability of the equal variances assumption of the model.
10. Suppose simultaneous 90% confidence intervals are to be constructed for all pairwise comparisons plus the contrast  $L$  of problem 8. Which methods of multiple comparisons are applicable? Of these, which is preferable? Why?

```

* golf.sas, STT 646 Sample Final Exam;
options ls=72;
;
data golf;
  infile 'golf.dat';
  input obs brand $ numhits;
;
proc print;
  var brand numhits;
;
proc plot;
  plot numhits*brand / vpos=20;
;
proc glm;
  class brand;
  model numhits = brand;
  means brand;
  output out=stats r=e p=pred;
;
proc rank normal=Blom;
  var e;
  ranks nscore;
;
proc corr nosimple;
  var numhits e pred nscore;
;
proc plot;
  plot e*pred / vpos=20 vref=0;
  plot e*nscore / vpos=20 vref=0 href=0;

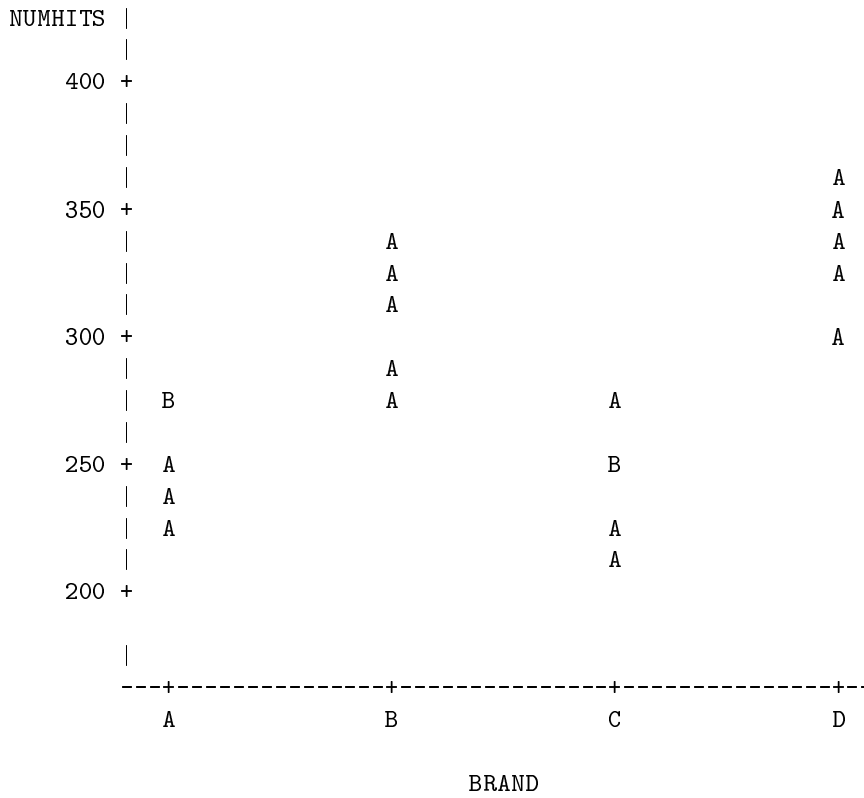
```

The SAS System

1

OBS	BRAND	NUMHITS
1	A	281
2	A	220
3	A	274
4	A	242
5	A	251
6	B	270
7	B	334
8	B	307
9	B	290
10	B	331
11	C	218
12	C	244
13	C	225
14	C	273
15	C	249
16	D	364
17	D	302
18	D	325
19	D	337
20	D	355

Plot of NUMHITS\*BRAND. Legend: A = 1 obs, B = 2 obs, etc.



General Linear Models Procedure  
Class Level Information

Class	Levels	Values
BRAND	4	A B C D

Number of observations in data set = 20

General Linear Models Procedure

Dependent Variable: NUMHITS

Source	DF	Sum of Squares	Mean Square
Model		29860.400	
Error	16	9698.400	606.150
Corrected Total			

General Linear Models Procedure

Level of BRAND	-----NUMHITS-----		
	N	Mean	SD
A	5	253.600000	24.6840029
B	5	306.400000	27.2084546
C	5	241.800000	21.6725633
D	5	336.600000	24.6028454

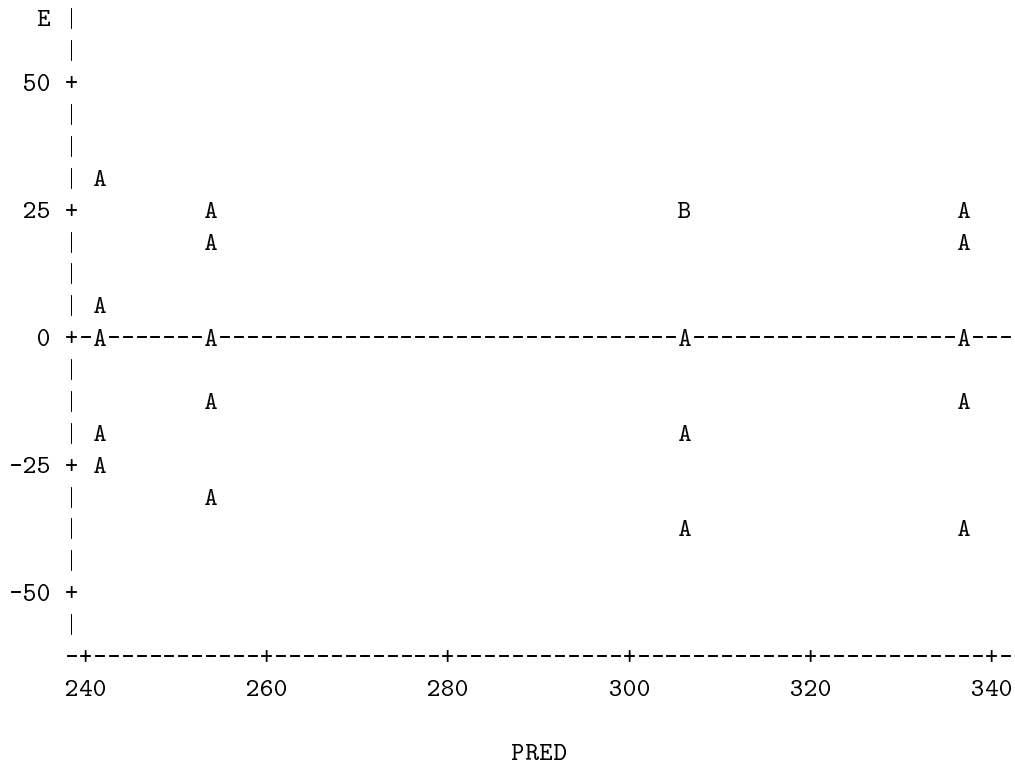
Correlation Analysis

4 'VAR' Variables: NUMHITS E PRED NSCORE

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 20

	NUMHITS	E	PRED	NSCORE
NUMHITS	1.00000 0.0	0.49514 0.0264	0.86881 0.0001	0.41900 0.0659
E	0.49514 0.0264	1.00000 0.0	0.00000 1.0000	0.97346 0.0001
PRED	0.86881 0.0001	0.00000 1.0000	1.00000 0.0	-0.07252 0.7613
NSCORE	0.41900	0.97346	-0.07252	1.00000
RANK FOR VARIABLE E	0.0659	0.0001	0.7613	0.0

Plot of E\*PRED. Legend: A = 1 obs, B = 2 obs, etc.



Plot of E\*NSCORE. Legend: A = 1 obs, B = 2 obs, etc.

