STAT 516 - Spring 2008 - Homework 7
Due: Monday, April 27th

1) This problem concerns the ability of pre-school students to associate objects with words. The experiment randomly assigned 24 pre-schoolers to be exposed to the objects by either TV, Audio Tape, or Picture Books. The time they were exposed to the material was either 5 min, 10 min, 15 min, or 20 min. Unfortunately, some of the children became distracted over the longer time spans and were unable to complete the experiment. The response variable is a score based on how many of the objects they could identify, and how accurate their identifications were. It is desired to analyze this data as an ANCOVA with the time being a continuous variable. The SAS code and output is attached.

	Time of Exposure			
Medium Used	5 min	10 min	15 min	20 min
TV	49	50	43	53
	39	55	38	48
Audio	55	67	53	85
	41	58		
Written	66	85	69	85
	68	92	62	

a) Verify that the slopes of the regression lines are parallel (using α=0.05).

b) Write down the model equation you are using, and get estimates of the parameters in it. Identify these parameters so that a reader could understand what the equation was saying.

c) Check the assumptions.

d) In terms of your model equation and the statement of the problem, explain what hypotheses are being tested by the p-values in the Type I and Type III tests respectively. Assuming that the assumptions were true, say if you would accept or reject the null hypothesis in each case.

2) For this problem we are using the data in Problem 5 (pg. 548) that can be found in Table 11.26 (pg. 551). It is desired to analyze it as a logistic regression. The data and output are attached.

a) Report the parameter estimates, and give the equation you would use to predict the probability of an abnormal EEG.

b) Use the Hosmer-Lemeshow Statistic to judge whether a logistic form seems appropriate for this data set. State your conclusion.

c) Assume that the logistic form is appropriate. Test whether there is a statistically significant relationship between the ventricle measurement and an abnormal EEG at an α=0.05 level.

d) What is the predicted probability of an abnormal EEG for a ventricle of size 15, and why might you doubt this prediction?

```
DATA exposure;
INPUT medium      $ time score @@;
CARDS;
TV    5     49    TV    5     39
TV    10    50    TV    10    55
TV    15    43    TV    15    38
TV    20    53    TV    20    48
AU    5     55    AU    5     41
AU    10    67    AU    10    58
AU    15    53
AU    20    85
WR    5     66    WR    5     68
WR    10    85    WR    10    92
WR    15    69    WR    16    62
WR    20    85
;

PROC INSIGHT;
OPEN exposure;
FIT score = medium time;
```

Parameter Information

Parameter	Variable	medium
1	Intercept	
2	medium	AU
3		TV
4		WR
5	time	

Model Equation

score = 66.9596 - 14.9213 P_2 - 29.0789 P_3 + 0.7195 time

Summary of Fit

Mean of Response	60.0476	R-Square	0.6377
Root MSE	10.5523	Adj R-Sq	0.5738

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	3	3331.9884	1110.6628	9.97	0.0005
Error	17	1892.9640	111.3508		
C Total	20	5224.9524			

Type I Tests

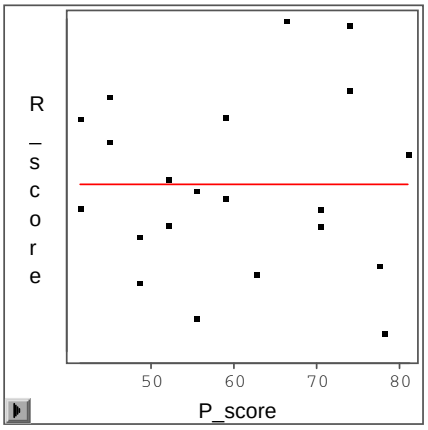
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
medium	2	3013.8155	1506.9077	13.53	0.0003
time	1	318.1729	318.1729	2.86	0.1092

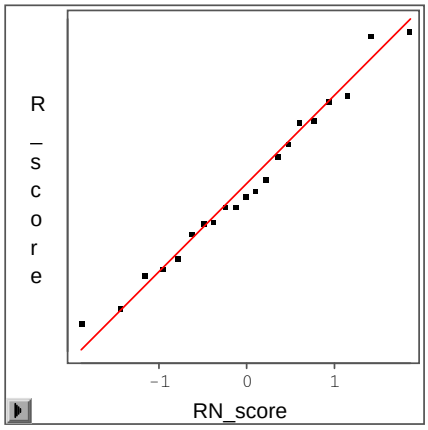
Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
medium	2	3141.0146	1570.5073	14.10	0.0002
time	1	318.1729	318.1729	2.86	0.1092

Parameter Estimates

Variable	medium	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept		1	66.9596	6.3379	10.57	<.0001	.	0
medium	AU	1	-14.9213	5.8792	-2.54	0.0212	0.7517	1.3303
	TV	1	-29.0789	5.4756	-5.31	<.0001	0.7499	1.3335
	WR	0	0	.	.	.	.	.
time		1	0.7195	0.4257	1.69	0.1092	0.9844	1.0158





```
PROC INSIGHT;
OPEN exposure;
FIT score = medium time time*medium;
RUN;
```

Parameter Information

Parameter	Variable	medium
1	Intercept	
2	medium	AU
3		TV
4		WR
5	time	
6	time*medium	AU
7		TV
8		WR

Model Equation

score = 71.1726 - 32.6360 P_2 - 26.1726 P_3 + 0.3555 time + 1.6104 P_6 - 0.2055 P_7

Summary of Fit

Mean of Response	60.0476	R-Square	0.7089
Root MSE	10.0692	Adj R-Sq	0.6119

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	5	3704.1154	740.8231	7.31	0.0012
Error	15	1520.8369	101.3891		
C Total	20	5224.9524			

Type I Tests

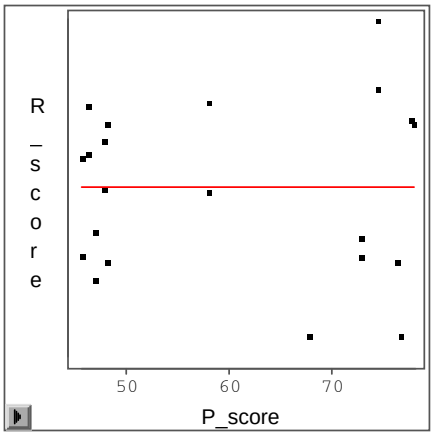
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
medium	2	3013.8155	1506.9077	14.86	0.0003
time	1	318.1729	318.1729	3.14	0.0968
time*medium	2	372.1270	186.0635	1.84	0.1937

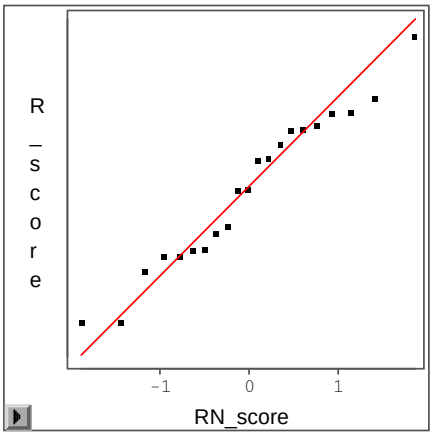
Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
medium	2	717.1741	358.5871	3.54	0.0552
time	1	406.7274	406.7274	4.01	0.0636
time*medium	2	372.1270	186.0635	1.84	0.1937

Parameter Estimates

Variable	medium	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept		1	71.1726	9.1960	7.74	<.0001	.	0
medium	AU	1	-32.6360	13.0812	-2.49	0.0248	0.1383	7.2331
	TV	1	-26.1726	12.6731	-2.07	0.0566	0.1275	7.8449
	WR	0	0	.	.	.	.	.
time		1	0.3555	0.7235	0.49	0.6303	0.3103	3.2227
time*medium	AU	1	1.6104	1.0568	1.52	0.1484	0.1347	7.4226
	TV	1	-0.2055	0.9638	-0.21	0.8341	0.1066	9.3804
	WR	0	0	.	.	.	.	.





```
DATA eeg;
INPUT  vent eeg  @@;
CARDS;
          53   0    37   0    63   0    25   0    60   0    58   0
          56   0    59   0    50   0    58   1    70   0    68   1
          50   0    59   0    51   0    76   0    74   1    62   1
          41   0    65   0    50   0    94   1    73   1    72   0
          45   1    56   0    56   0    75   0    76   0    78   1
          50   0    68   0    47   0    66   0    42   1    76   1
          57   0    65   0    51   0    83   1    51   0    80   1
          70   0    68   1    49   0    56   1    58   1    58   1
          64   1    60   1    57   0    54   0    58   0    63   1
          61   0    70   0    40   0    51   1    58   1    70   1
          57   1    84   0    58   0    51   1    57   0    85   1
          50   0    48   0    67   1    62   0    65   0
;

PROC LOGISTIC DATA=eeg DESCENDING;
MODEL eeg=vent /LACKFIT;
RUN;
```

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	6.8361	1	0.0089
Score	6.5287	1	0.0106
Wald	5.9318	1	0.0149

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.0478	1.4766	7.5151	0.0061
vent	1	0.0569	0.0234	5.9318	0.0149

Partition for the Hosmer and Lemeshow Test

		eeg = 1		eeg = 0	
Group	Total	Observed	Expected	Observed	Expected
1	7	2	1.04	5	5.96
2	7	0	1.59	7	5.41
3	7	2	1.75	5	5.25
4	8	2	2.43	6	5.57
5	7	4	2.25	3	4.75
6	7	2	2.47	5	4.53
7	7	2	2.84	5	4.16
8	8	4	3.75	4	4.25
9	7	3	3.85	4	3.15
10	6	5	4.04	1	1.96

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
6.9665	8	0.5403