MA 2612 **Test 3** **B '03**

NAME: ______________________________

1. A research team is investigating the association between a "gold standard" measure of hearing (the variable RESP_PRE) and several non-invasive measures (signal-noise of responsive emissions at different frequencies). The variables are:

RESP_PRE	The "gold standard" measure
SN2_PRE	Signal/noise ratio at 2 kHz
SN4_PRE	Signal/noise ratio at 4 kHz
SN6_PRE	Signal/noise ratio at 6 kHz
SN10_PRE	Signal/noise ratio at 10 kHz

SAS/INSIGHT output for the regression of RESP_PRE on all four regressors is shown in Figures 1-3.

(a) **(10 points)** What proportion of the variation in RESP_PRE is explained by the regression model?
ANS: $R^2 = 0.6264$.

(b) **(10 points)** At the 0.05 level of significance, conduct a test for $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, versus H_a : not H_0. Tell the observed value of the test statistic, its p-value, and your conclusion.
ANS: *Test statistic:* $F^* = 17.19$, *p-value: less than* 0.0001. *Since this is less than* 0.05, *reject* H_0, *and conclude at least one* β_i *is non-zero.*

(c) **(10 points)** To see which regressor(s) are responsible for the significant relationship with RESP_PRE, for each of $i = 1, 2, 3, 4$, conduct a test of $H_{0i} : \beta_i = 0$, versus $H_{ai} : \beta_i \neq 0$ at the 0.05 level of significance. What do you find?
ANS: *Since all p-values are less than* 0.05, *we reject* H_{0i} *for all* i. *Therefore, we conclude each of the regressors is significantly related to RESP_PRE.*

(d) **(10 points)** Is there evidence of lack of fit or violation of model assumptions? Justify your answer.
ANS: *The studentized residuals show a slight left skewness, indicating that the assumption of normality is questionable. A few of the studentized residuals are slightly smaller than* -2, *but they do not appear to be outliers. The residual plots give no reason to suspect non-randomness.*

(e) **(10 points)** Interpret the coefficient of SN2_PRE in the fitted model
ANS: *If all other regressors are held constant, a one unit increase in SN2_PRE increases the prediction of RESP_PRE by* 0.1599.

(f) **(10 points)** The researchers want to predict a new observation at SN2_PRE=8, SN4_PRE=7, SN6_PRE=−3, SN10_PRE=9. Find the value of the predictor based on the fitted model.
ANS:

$$\widehat{Y}_{new} = -8.9548 + (0.1599)(8) + (0.4205)(7) - (0.2360)(-3) + (0.2255)(9) = -1.9946.$$

2. The researchers considered a second model which includes the above regressors and adds SN3_PRE, the signal-noise ratio at 3 kHz. SAS/INSIGHT output for the resulting model is shown in Figure 4.

(a) **(10 points)** Does the t test for the significance of SN3_PRE indicate this regressor should be included in the model? Justify your answer.
ANS: *No, its p-value of 0.2236 is too large to be significant.*

(b) **(10 points)** What do the adjusted R^2 values suggest about including SN3_PRE in the model? Justify your answer.
ANS: *Since* $R_a^2 = 0.5952$ *for this model, which exceeds its value of* 0.5900 *for the previous model,* R_a^2 *suggests SN3_PRE should be included in the model.*

(c) **(10 points)** Does multicollinearity appear to be a problem for this model? Justify your answer.
ANS: *No, since all VIFs are well below 10.*

3. **(10 points)** Consider again the model from problem 1. If $t_{41,0.975} = 2.0195$, and if a level 0.95 confidence interval for the mean response at SN2_PRE=8, SN4_PRE=7, SN6_PRE=−3, SN10_PRE=9 is $(-3.0533, -0.9364)$, obtain a level 0.95 prediction interval for a new observation atSN2_PRE=8, SN4_PRE=7, SN6_PRE=−3, SN10_PRE=9.

ANS: *We know that* $(-3.0530, -0.9362) = \widehat{Y} \pm \hat{\sigma}(\widehat{Y}) t_{41,0.975} = -1.9946 \pm 2.0195 \hat{\sigma}(\widehat{Y})$, *which implies that* $\hat{\sigma}(\widehat{Y}) = 0.5241$. *We also know that* $\hat{\sigma}(Y_{new} - \widehat{Y}_{new}) = \sqrt{MSE + \hat{\sigma}^2(\widehat{Y})} = \sqrt{6.3979 + 0.5241^2} = 2.5831$. *Thus, the desired interval is* $-1.9946 \pm (2.0195)(2.5831) = (-7.2115, 3.2217)$.

Resp_pre = SN2_pre SN4_pre SN6_pre SN10_pre	
Response Distribution:	Normal
Link Function:	Identity

Model Equation
Resp_pre = - 8.9548 + 0.1599 SN2_pre + 0.4205 SN4_pre - 0.2360 SN6_pre + 0.2255 SN10_pre

Summary of Fit			
Mean of Response	0.1696	R-Square	0.6264
Root MSE	2.5294	Adj R-Sq	0.5900

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	4	439.8435	109.9609	17.19	<.0001
Error	41	262.3139	6.3979		
C Total	45	702.1574			

Type III Tests					
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
SN2_pre	1	45.0996	45.0996	7.05	0.0112
SN4_pre	1	248.1784	248.1784	38.79	<.0001
SN6_pre	1	50.9913	50.9913	7.97	0.0073
SN10_pre	1	27.3612	27.3612	4.28	0.0450

Parameter Estimates							
Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept	1	-8.9548	1.3925	-6.43	<.0001	.	0
SN2_pre	1	0.1599	0.0602	2.66	0.0112	0.8987	1.1127
SN4_pre	1	0.4205	0.0675	6.23	<.0001	0.6561	1.5242
SN6_pre	1	-0.2360	0.0836	-2.82	0.0073	0.8217	1.2170
SN10_pre	1	0.2255	0.1091	2.07	0.0450	0.6995	1.4297

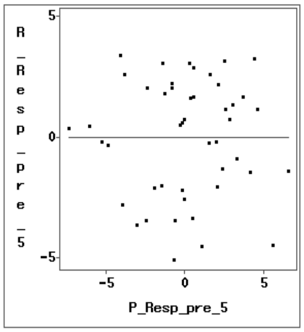
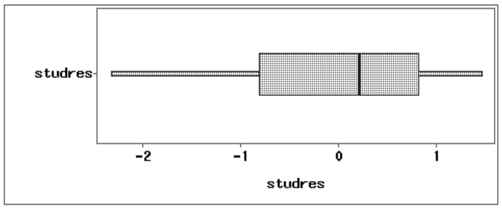
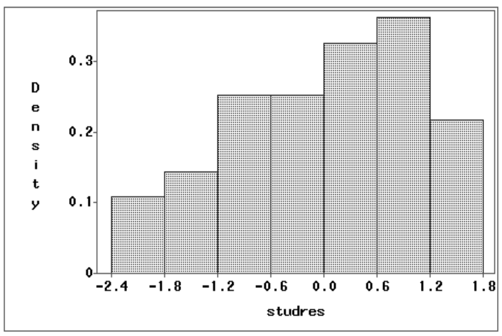
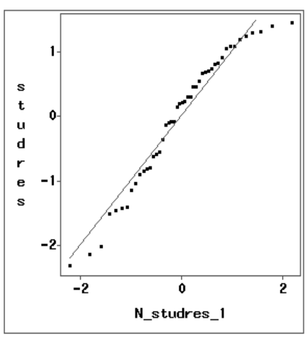


Figure 1: *SAS output for problem 1*

studres







Moments			
N	46.0000	Sum Wgts	46.0000
Mean	-0.0081	Sum	-0.3725
Std Dev	1.0377	Variance	1.0769
Skewness	-0.5045	Kurtosis	-0.6943
USS	48.4637	CSS	48.4607
CV	-12814.179	Std Mean	0.1530

Quantiles			
100% Max	1.4822	99.0%	1.4822
75% Q3	0.8262	97.5%	1.4233
50% Med	0.2176	95.0%	1.3319
25% Q1	-0.8097	90.0%	1.2720
0% Min	-2.3196	10.0%	-1.4585
Range	3.8018	5.0%	-2.0075
Q3-Q1	1.6359	2.5%	-2.1362
Mode	.	1.0%	-2.3196

Normal QQ Ref Lines		
Line	Intercept	Slope
———	0	1.0000

Figure 2: *SAS output for problem 1*

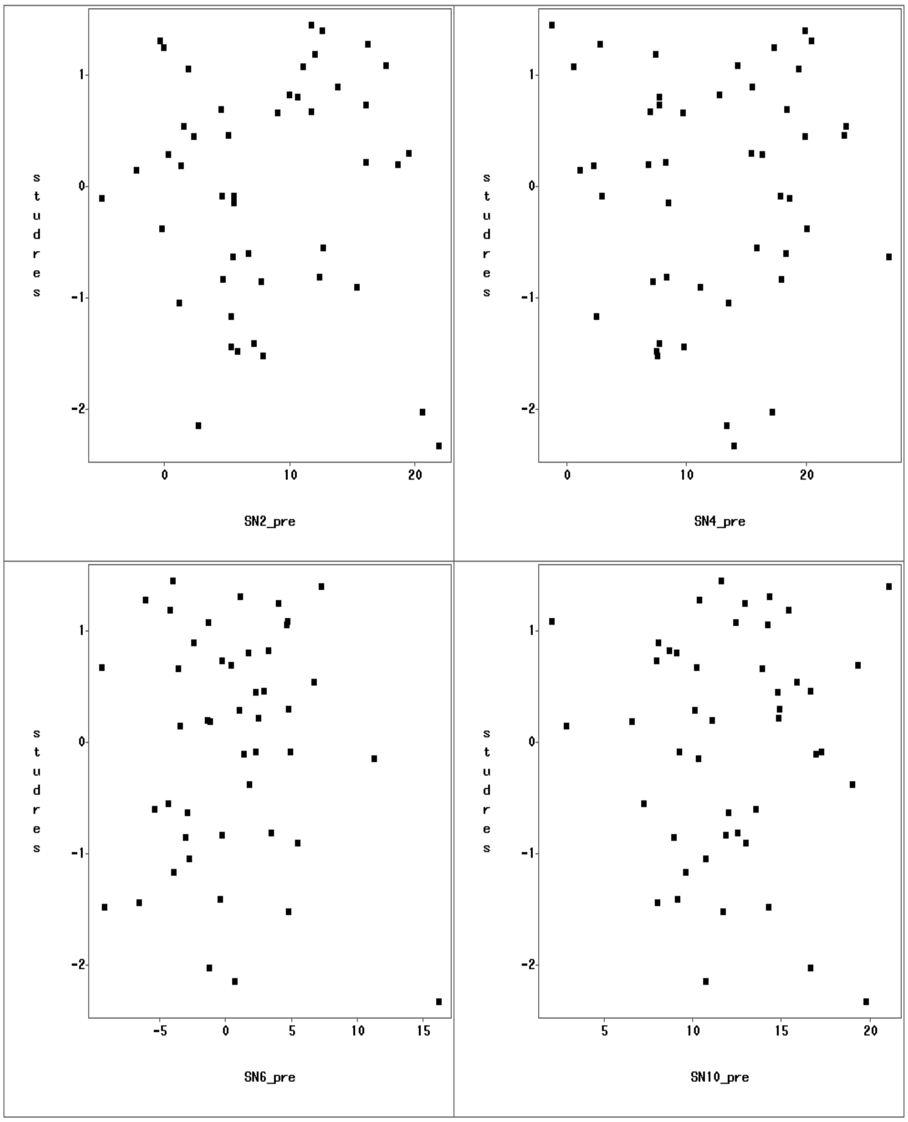


Figure 3: *SAS output for problem 1*

Resp_pre = SN2_pre SN3_pre SN4_pre SN6_pre SN10_pre	
Response Distribution:	Normal
Link Function:	Identity

Model Equation
Resp_pre = - 9.5504 + 0.1307 SN2_pre + 0.1216 SN3_pre + 0.3983 SN4_pre - 0.2258 SN6_pre + 0.1961 SN10_pre

Summary of Fit			
Mean of Response	0.1696	R-Square	0.6402
Root MSE	2.5133	Adj R-Sq	0.5952

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	5	449.4972	89.8994	14.23	<.0001
Error	40	252.6602	6.3165		
C Total	45	702.1574			

Type III Tests					
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
SN2_pre	1	26.1122	26.1122	4.13	0.0487
SN3_pre	1	9.6536	9.6536	1.53	0.2236
SN4_pre	1	207.7148	207.7148	32.88	<.0001
SN6_pre	1	46.2557	46.2557	7.32	0.0100
SN10_pre	1	19.7395	19.7395	3.13	0.0847

Parameter Estimates							
Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept	1	-9.5504	1.4651	-6.52	<.0001	.	0
SN2_pre	1	0.1307	0.0643	2.03	0.0487	0.7781	1.2853
SN3_pre	1	0.1216	0.0984	1.24	0.2236	0.7515	1.3306
SN4_pre	1	0.3983	0.0694	5.73	<.0001	0.6121	1.6337
SN6_pre	1	-0.2258	0.0834	-2.71	0.0100	0.8138	1.2288
SN10_pre	1	0.1961	0.1109	1.77	0.0847	0.6673	1.4986

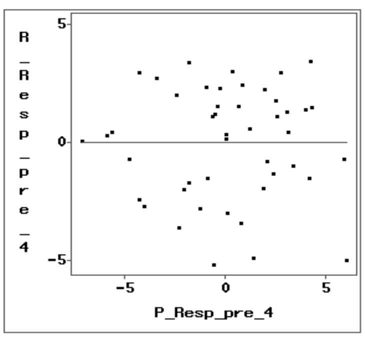


Figure 4: *SAS output for problem 2*