MASTER'S DIAGNOSTIC EXAMINATION

January 2012

Student's Name ____

INSTRUCTIONS FOR STUDENTS:

1. DO NOT put your NAME on the exam. Place the NUMBER assigned to you on the

   UPPER RIGHT HAND CORNER of EACH PAGE of your solutions.

2. Please start your answer to EACH QUESTION on a SEPARATE sheet of paper.
3. Use only one side of each sheet of paper.
4. You must answer all four questions: Questions I, II, III and IV.
5. Be sure to attempt all parts of the four questions. It may be possible to answer a later part of a question without having solved the earlier parts.
6. Be sure to hand in all of your exam. No additional material will be accepted once the exam has ended and you have left the exam room.
7. You may use only a calculator, pencil or pen, and blank paper for this examination. No other materials are allowed.

I attest that I spent no more than 4 hours to complete the exam. I used only the materials described above. I did not receive assistance from anyone during the taking of this exam.

Student's Signature__________________________________

INSTRUCTIONS FOR PROCTOR:

Immediately after the student completes the exam, **fax** the student's solutions to **979-845-6060** or email to **longneck@stat.tamu.edu** Do not send the questions, just send the student's solutions.

(1) I certify that the time at which the student started the exam was ______________________

and the time at which the student completed the exam was ______________________

(2) I certify that the student has followed all the **INSTRUCTIONS FOR STUDENTS** listed above.

(3) I certify that the student's solutions were faxed to **979-845-6060** or

emailed to **longneck@stat.tamu.edu**.

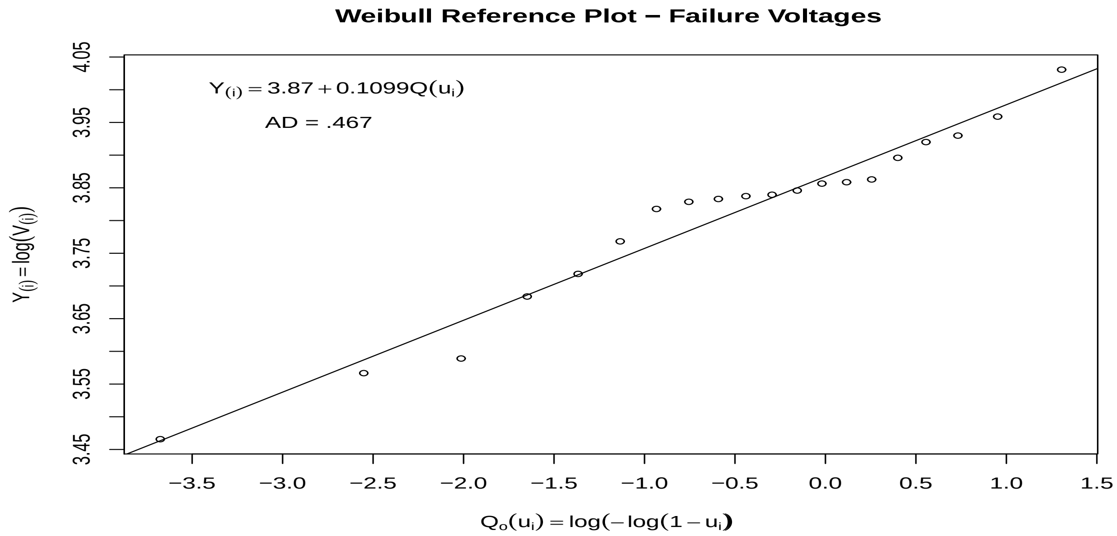
Proctor's Signature__________________________________

QUESTION I.

An electronics firm is testing the insulation used on its electrical cable. A study was conducted to determine the voltage level at which the insulation failed. The study consisted of 20 specimens and the data given in the following table is failure voltages (in kilovolts per millimeter).

32.0	35.4	36.2	39.8	41.2	43.3	45.5	46.0	46.2	46.4
46.5	46.8	47.3	47.4	47.6	49.2	50.4	50.9	52.4	56.3

A plot of $Y_i = log(V_i)$ versus $Q(u_i) = log(-log(1 - u_i))$ is given here, where V_i is the failure voltage of the ith unit and $u_i = (i - .5)/20$ for $i = 1, \cdots, 20$.



1. Explain how the Weibull Reference Plot is able to assess the fit of the Weibull model to the data without specifying the parameters, γ and β.

2. The Anderson-Darling statistic yielded a value of $AD = .467$ as a measure of how well a Weibull model would fit the data. Based on AD and the above plot, assess the fit of a Weibull distribution to the data. Make sure to provide a p-value from the AD statistic (tables attached).

3. Use the following SAS output to obtain the MLE's for the parameters, γ and β from a Weibull distribution:

   $$F(x) = 1 - e^{-x^{\gamma}/\beta}$$

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.8667	0.0251	3.8176	3.9159	23792.0	<.0001
Scale	1	0.1065	0.0184	0.0759	0.1495		
Weibull Scale	1	47.7866	1.1979	45.4954	50.1931		
Weibull Shape	1	9.3863	1.6224	6.6891	13.1712		

Hint: SAS uses the Weibull cdf in the form $F(x) = 1 - e^{-(x/\alpha)^{\gamma}}$

4. The above table contains 95% Confidence Limits for α and γ. Using the information in the above table, provide approximate 95% Lower Limits for α and γ.

   Hint: The following upper percentiles are from the standard normal distribution:

   $Z_{.005} = 2.576;\quad Z_{.01} = 2.326;\quad Z_{.025} = 1.96;\quad Z_{.05} = 1.645;\quad Z_{.10} = 1.282$

5. The researcher was uncertain about using a Weibull distribution in the analysis and decides to use a distribution-free method of estimating the survival function. The Kaplan-Meier Product-Limit estimator of the survival function is displayed in the SAS output given below.

   a. Obtain the estimate of the 78th percentile provided by the Product-Limit estimator.

   b. Obtain the MLE estimate of the 78th percentile based on fitting the Weibull distribution. Hint: Use your results from part 3.

   c. Which of the two estimators would you recommend?

Product-Limit Survival Estimates

V	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	20
32.0000	0.9500	0.0500	0.0487	1	19
35.4000	0.9000	0.1000	0.0671	2	18
36.2000	0.8500	0.1500	0.0798	3	17
39.8000	0.8000	0.2000	0.0894	4	16
41.2000	0.7500	0.2500	0.0968	5	15
43.3000	0.7000	0.3000	0.1025	6	14
45.5000	0.6500	0.3500	0.1067	7	13
46.0000	0.6000	0.4000	0.1095	8	12
46.2000	0.5500	0.4500	0.1112	9	11
46.4000	0.5000	0.5000	0.1118	10	10
46.5000	0.4500	0.5500	0.1112	11	9
46.8000	0.4000	0.6000	0.1095	12	8
47.3000	0.3500	0.6500	0.1067	13	7
47.4000	0.3000	0.7000	0.1025	14	6
47.6000	0.2500	0.7500	0.0968	15	5
49.2000	0.2000	0.8000	0.0894	16	4
50.4000	0.1500	0.8500	0.0798	17	3
50.9000	0.1000	0.9000	0.0671	18	2
52.4000	0.0500	0.9500	0.0487	19	1
56.3000	0	1.0000	.	20	0

Table 1: Percentiles for GOF Measures (Completely Specified Distributions)

		Upper Percentiles							
Statistic	Modified Statistic	.25	.15	.10	.05	.025	.01	.005	.001
D_n	$D_n(\sqrt{n} + .12 + .11/\sqrt{n})$	1.019	1.138	1.224	1.358	1.480	1.628	1.731	1.950
W_n^2	$(W_n^2 - \frac{.4}{n} + \frac{.6}{n^2})(1 + \frac{1}{n})$	0.209	0.284	0.347	0.461	0.581	0.743	0.869	1.167
A_n^2	For all $n \geq 5$	1.248	1.610	1.933	2.492	3.070	3.857	4.500	6.000

Table 2: CDF for Anderson-Darling (Completely Specified Distributions)

z	G(z)	z	G(z)	z	G(z)	z	G(z)	z	G(z)	z	G(z)
0.05	0.0000	0.75	0.4815	1.45	0.8111	2.15	0.9239	2.85	0.9674	3.80	0.9891
0.10	0.0000	0.80	0.5190	1.50	0.8235	2.20	0.9285	2.90	0.9692	3.90	0.9902
0.15	0.0000	0.85	0.5537	1.55	0.8350	2.25	0.9328	2.95	0.9710	4.00	0.9913
0.20	0.0096	0.90	0.5858	1.60	0.8457	2.30	0.9368	3.00	0.9726	4.25	0.9934
0.25	0.0296	0.95	0.6154	1.65	0.8556	2.35	0.9405	3.25	0.9795	4.50	0.9950
0.30	0.0618	1.00	0.6427	1.70	0.8648	2.40	0.9441	3.30	0.9807	4.60	0.9955
0.35	0.1036	1.05	0.6680	1.75	0.8734	2.45	0.9474	3.35	0.9818	4.70	0.9960
0.40	0.1513	1.10	0.6912	1.80	0.8814	2.50	0.9504	3.40	0.9828	4.80	0.9964
0.45	0.2019	1.15	0.7127	1.85	0.8888	2.55	0.9534	3.45	0.9837	4.90	0.9968
0.50	0.2532	1.20	0.7324	1.90	0.8957	2.60	0.9561	3.50	0.9846	5.00	0.9971
0.55	0.3036	1.25	0.7503	1.95	0.9021	2.65	0.9586	3.55	0.9855	5.50	0.9983
0.60	0.3520	1.30	0.7677	2.00	0.9082	2.70	0.9610	3.60	0.9863	6.00	0.9990
0.65	0.3930	1.35	0.7833	2.05	0.9138	2.75	0.9633	3.65	0.9870	7.00	0.9997
0.70	0.4412	1.40	0.7973	2.10	0.9190	2.80	0.9654	3.70	0.9878	8.00	0.9999

Table 3: Modifications and Percentiles for GOF Measures for Normal Distributions with μ and σ Unknown

		Upper Percentiles							
Statistic	Modified Statistic	.50	.25	.15	.10	.05	.025	.01	.005
D_n	$D_n(\sqrt{n} - .01 + .85/\sqrt{n})$	-	-	0.775	0.819	0.895	0.995	1.035	-
W_n^2	$W_n^2(1 + \frac{.5}{n})$	0.051	0.074	0.091	0.104	0.126	0.148	0.179	0.201
A_n^2	$A_n^2(1 + \frac{.75}{n} + \frac{2.25}{n^2})$	0.341	0.470	0.561	0.631	0.752	0.873	1.035	1.159

Table 4: Modifications and Percentiles for GOF Measures for Exponential Distribution with β Unknown

		Upper Percentiles							
Statistic	Modified Statistic	.25	.20	.15	.10	.05	.025	.01	.005
D_n	$(D_n - \frac{0.2}{n})(\sqrt{n} + 0.26 + \frac{0.5}{\sqrt{n}})$	-	-	0.926	0.995	1.094	1.184	-	-
W_n^2	$W_n^2(1.0 + \frac{0.16}{n})$	0.116	0.130	0.148	0.175	0.222	0.271	0.338	0.390
A_n^2	$A_n^2(1.0 + \frac{0.6}{n})$	0.736	0.816	0.916	1.062	1.321	1.591	1.959	2.244

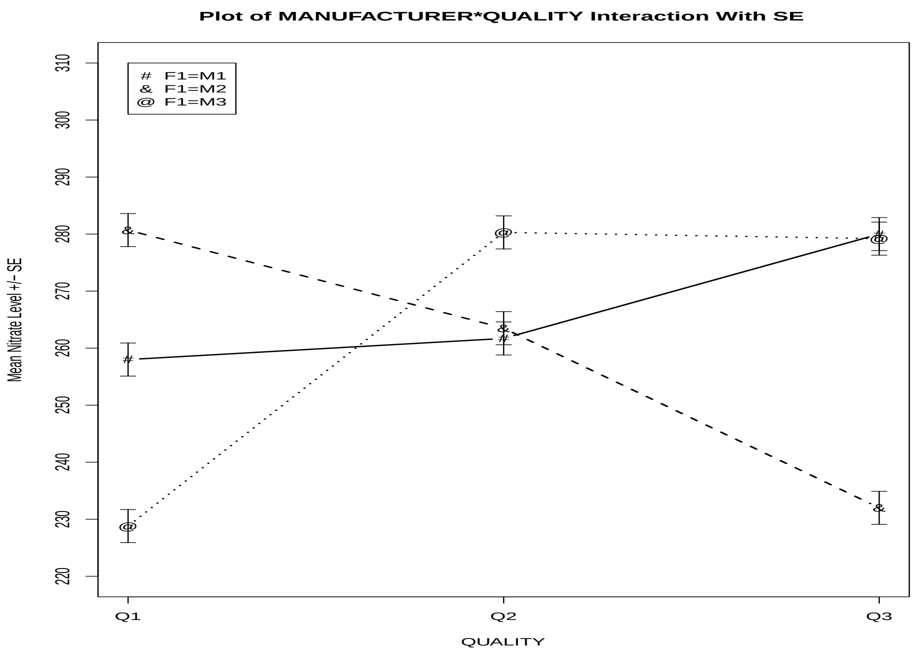
Table 5: Modifications and Percentiles for A-D Measure for Extreme Value Distribution with Unspecified Parameters

		Upper Percentiles				
Statistic	Modified Statistic	.25	.10	.05	.025	.01
A_n^2	$A_n^2(1.0 + \frac{0.2}{\sqrt{n}})$	0.474	0.637	0.757	0.877	1.038

QUESTION II.

The USDA is evaluating the level of nitrate (NO_3) from sausages obtained from three largest manufacturers, M1, M2, M3 in the US. Each manufacturer produces three grades of quality, either Q1, Q2, or Q3 of their sausage. The processing of different grades of sausage from a common production run may involve different sources of raw materials and processing environments, and these factors sometimes are problematic. Each manufacturer submits two sausages of each grade from each of three production runs. The amount of NO_3 is determined by an USDA lab and is reported in the following table. The three manufacturers are the only manufacturers under evaluation, the production runs were randomly selected, and are representative of general production runs of each manufacturer.

	Manufacture								
	M1			M2			M3		
	Run			Run			Run		
Grade	R1	R2	R3	R4	R5	R6	R7	R8	R9
Q1	253	265	253	230	234	231	225	228	232
	256	270	251	226	239	232	229	227	232
Q2	262	263	255	257	268	265	277	276	289
	260	266	264	267	258	266	276	277	287
Q3	279	285	277	275	286	284	280	278	282
	279	288	272	272	283	284	276	277	282



Use the above plots, data, and the attached SAS output to answer the following questions.

1. Do the necessary conditions for testing hypotheses and constructing confidence intervals appear to be satisfied? Justify your answers.

   C_1 Normality:

   C_2 Equal Variance:

   C_3 Independence:

2. At the $\alpha = .05$ level, which main effects and interactions are significant? Justify your answer by including the relevant p-values along with their pair of degrees of freedom $(df_{NUM.}, df_{DEN.})$.

3. What is the expected value of $MS_{Manufacturer}$?

4. Separate the three Grade Levels of Quality into groups of levels such that all levels in a group are not significantly different from any other member of the group with respect to their mean NO_3 level. Use an experimentwise error rate of $\alpha = .05$.

5. Provide a 95% confidence on the mean NO_3 level of a sausage having Quality Grade Q1 produced by Manufacturer M1.

6. Identify each of the following sums of squares formulas with its correct source of variation in the AOV table for the experiment. For example, if y_{ijkl} is the NO_3 level of sausage l of Quality Grade i from Manufacturer j on production run k, then

$\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=1}^{2}(y_{ijkl} - \bar{y}_{....})^2$ is SS__$TOTAL$__

a. $18\sum_{i=1}^{3}(\bar{y}_{i...} - \bar{y}_{....})^2$ is SS________________________

b. $6\sum_{i=1}^{3}\sum_{j=1}^{3}(\bar{y}_{ij..} - \bar{y}_{i...} - \bar{y}_{.j..} + \bar{y}_{....})^2$ is SS________________________

c. $6\sum_{j=1}^{3}\sum_{k=1}^{3}(\bar{y}_{.jk.} - \bar{y}_{.j..})^2$ is SS________________________

```
OPTIONS LS=90 PS=55 nocenter nodate;
TITLE 'SAS OUTPUT FOR QUESTION II';
DATA MANU;
INPUT M $ Q $ R $ Y @@;
TRT=COMPRESS (M) || COMPRESS (Q);
LABEL M="MANUFACTURER" Q="QUALITY";
CARDS;
M1 Q1 R1 253  M1 Q1 R2 265  M1 Q1 R3 253 M1 Q1 R1 256  M1 Q1 R2 270  M1 Q1 R3 251
M1 Q2 R1 262  M1 Q2 R2 263  M1 Q2 R3 255 M1 Q2 R1 260  M1 Q2 R2 266  M1 Q2 R3 264
M1 Q3 R1 279  M1 Q3 R2 285  M1 Q3 R3 277 M1 Q3 R1 279  M1 Q3 R2 288  M1 Q3 R3 272
M2 Q3 R4 230  M2 Q3 R5 234  M2 Q3 R6 231 M2 Q3 R4 226  M2 Q3 R5 239  M2 Q3 R6 232
M2 Q2 R4 257  M2 Q2 R5 268  M2 Q2 R6 265 M2 Q2 R4 267  M2 Q2 R5 258  M2 Q2 R6 266
M2 Q1 R4 275  M2 Q1 R5 286  M2 Q1 R6 284 M2 Q1 R4 272  M2 Q1 R5 283  M2 Q1 R6 284
M3 Q1 R7 225  M3 Q1 R8 228  M3 Q1 R9 232 M3 Q1 R7 229  M3 Q1 R8 227  M3 Q1 R9 232
M3 Q2 R7 277  M3 Q2 R8 276  M3 Q2 R9 289 M3 Q2 R7 276  M3 Q2 R8 277  M3 Q2 R9 287
M3 Q3 R7 280  M3 Q3 R8 278  M3 Q3 R9 282 M3 Q3 R7 276  M3 Q3 R8 277  M3 Q3 R9 282
PROC GLM;
CLASS M Q R;
MODEL Y = M Q M*Q R(M) Q*R(M);
RANDOM R(M) Q*R(M)/TEST;
LSMEANS M Q M*Q/STDERR PDIFF ADJUST=TUKEY;
RUN;
PROC MIXED CL ALPHA=.05 COVTEST;
CLASS M Q R;
MODEL Y = M Q M*Q ;
RANDOM R(M) Q*R(M);
LSMEANS M Q M*Q/ ADJUST=TUKEY;
RUN;
PROC GLM;
CLASS TRT;
MODEL Y = TRT;
MEANS TRT/HOVTEST=BF;
OUTPUT OUT=ASSUMP R=RESID P=MEANS;
PROC GPLOT; PLOT MEANS*TRT ;PLOT RESID*TRT/VREF=0;
PROC UNIVARIATE DEF=5 PLOT NORMAL;
VAR RESID;
RUN;
```

SAS OUTPUT FOR QUESTION II

Class	Levels	Values
M	3	M1 M2 M3
Q	3	Q1 Q2 Q3
R	9	R1 R2 R3 R4 R5 R6 R7 R8 R9

Number of Observations Read 54

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	26	20826.14815	801.00570	92.62	<.0001
Error	27	233.50000	8.64815		
Corrected Total	53	21059.64815			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
M	2	552.48148	276.24074	31.94	<.0001
Q	2	1473.03704	736.51852	85.16	<.0001
M*Q	4	17878.96296	4469.74074	516.84	<.0001
R(M)	6	729.00000	121.50000	14.05	<.0001
Q*R(M)	12	192.66667	16.05556	1.86	0.0887

Source	Type III Expected Mean Square
M	Var(Error) + 2 Var(Q*R(M)) + 6 Var(R(M)) + Q(M,M*Q)
Q	Var(Error) + 2 Var(Q*R(M)) + Q(Q,M*Q)
M*Q	Var(Error) + 2 Var(Q*R(M)) + Q(M*Q)
R(M)	Var(Error) + 2 Var(Q*R(M)) + 6 Var(R(M))
Q*R(M)	Var(Error) + 2 Var(Q*R(M))

Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: Y

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
*	M	2	552.481481	276.240741	2.27	0.1841
	Error: MS(R(M))	6	729.000000	121.500000		

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
*	Q	2	1473.037037	736.518519	45.87	<.0001
	M*Q	4	17879	4469.740741	278.39	<.0001
	R(M)	6	729.000000	121.500000	7.57	0.0016
	Error: MS(Q*R(M))	12	192.666667	16.055556		

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q*R(M)	12	192.666667	16.055556	1.86	0.0887
Error: MS(Error)	27	233.500000	8.648148		

SAS OUTPUT FOR QUESTION II

The GLM Procedure

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

M	Y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
M1	266.555556	0.693147	<.0001	1
M2	258.722222	0.693147	<.0001	2
M3	262.777778	0.693147	<.0001	3

Least Squares Means for effect M
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Y

i/j	1	2	3
1		<.0001	0.0018
2	<.0001		0.0009
3	0.0018	0.0009	

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

Q	Y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
Q1	255.833333	0.693147	<.0001	1
Q2	268.500000	0.693147	<.0001	2
Q3	263.722222	0.693147	<.0001	3

Least Squares Means for effect Q
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Y

i/j	1	2	3
1		<.0001	<.0001
2	<.0001		0.0001
3	<.0001	0.0001	

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

M	Q	Y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
M1	Q1	258.000000	1.200566	<.0001	1
M1	Q2	261.666667	1.200566	<.0001	2
M1	Q3	280.000000	1.200566	<.0001	3
M2	Q1	280.666667	1.200566	<.0001	4
M2	Q2	263.500000	1.200566	<.0001	5
M2	Q3	232.000000	1.200566	<.0001	6
M3	Q1	228.833333	1.200566	<.0001	7
M3	Q2	280.333333	1.200566	<.0001	8
M3	Q3	279.166667	1.200566	<.0001	9

Least Squares Means for effect M*Q
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Y

i/j	1	2	3	4	5	6	7	8	9
1		0.4574	<.0001	<.0001	0.0659	<.0001	<.0001	<.0001	<.0001
2	0.4574		<.0001	<.0001	0.9724	<.0001	<.0001	<.0001	<.0001
3	<.0001	<.0001		1.0000	<.0001	<.0001	<.0001	1.0000	0.9999
4	<.0001	<.0001	1.0000		<.0001	<.0001	<.0001	1.0000	0.9921
5	0.0659	0.9724	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001
6	<.0001	<.0001	<.0001	<.0001	<.0001		0.6414	<.0001	<.0001
7	<.0001	<.0001	<.0001	<.0001	<.0001	0.6414		<.0001	<.0001
8	<.0001	<.0001	1.0000	1.0000	<.0001	<.0001	<.0001		0.9986
9	<.0001	<.0001	0.9999	0.9921	<.0001	<.0001	<.0001	0.9986	

SAS OUTPUT FOR QUESTION II

The Mixed Procedure

Model Information

Data Set	WORK.MANU
Dependent Variable	Y
Covariance Structure	Variance Components
Estimation Method	REML

Class Level Information

Class	Levels	Values
M	3	M1 M2 M3
Q	3	Q1 Q2 Q3
R	9	R1 R2 R3 R4 R5 R6 R7 R8 R9

Number of Observations

Number of Observations Read	54
Number of Observations Used	54

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr > Z	Alpha	Lower	Upper
R(M)	17.5741	11.7423	1.50	0.0672	0.05	6.5803	122.55
Q*R(M)	3.7037	3.4822	1.06	0.1438	0.05	1.0582	101.54
Residual	8.6481	2.3537	3.67	0.0001	0.05	5.4058	16.0224

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
M	2	6	2.27	0.1841
Q	2	12	45.87	<.0001
M*Q	4	12	278.39	<.0001

SAS OUTPUT FOR QUESTION II

The Mixed Procedure

Least Squares Means

Effect	MANUFACTURER	QUALITY	Estimate	Standard Error	DF	t Value	Pr > |t|
M	M1		266.56	2.5981	6	102.60	<.0001
M	M2		258.72	2.5981	6	99.58	<.0001
M	M3		262.78	2.5981	6	101.14	<.0001
Q		Q1	255.83	1.6866	12	151.69	<.0001
Q		Q2	268.50	1.6866	12	159.20	<.0001
Q		Q3	263.72	1.6866	12	156.36	<.0001
M*Q	M1	Q1	258.00	2.9213	12	88.32	<.0001
M*Q	M1	Q2	261.67	2.9213	12	89.57	<.0001
M*Q	M1	Q3	280.00	2.9213	12	95.85	<.0001
M*Q	M2	Q1	280.67	2.9213	12	96.08	<.0001
M*Q	M2	Q2	263.50	2.9213	12	90.20	<.0001
M*Q	M2	Q3	232.00	2.9213	12	79.42	<.0001
M*Q	M3	Q1	228.83	2.9213	12	78.33	<.0001
M*Q	M3	Q2	280.33	2.9213	12	95.96	<.0001
M*Q	M3	Q3	279.17	2.9213	12	95.56	<.0001

SAS OUTPUT FOR QUESTION II

The Mixed Procedure

Differences of Least Squares Means

Effect	MANUFACTURER	QUALITY	MANUFACTURER	QUALITY	Pr > |t|	Adjustment	Adj P
M	M1		M2		0.0770	Tukey	0.1632
M	M1		M3		0.3435	Tukey	0.5878
M	M2		M3		0.3120	Tukey	0.5463
Q		Q1		Q2	<.0001	Tukey-Kramer	<.0001
Q		Q1		Q3	<.0001	Tukey-Kramer	0.0002
Q		Q2		Q3	0.0038	Tukey-Kramer	0.0098
M*Q	M1	Q1	M1	Q2	0.1390	Tukey-Kramer	0.7963
M*Q	M1	Q1	M1	Q3	<.0001	Tukey-Kramer	<.0001
M*Q	M1	Q2	M1	Q3	<.0001	Tukey-Kramer	<.0001
M*Q	M2	Q1	M2	Q2	<.0001	Tukey-Kramer	0.0002
M*Q	M2	Q1	M2	Q3	<.0001	Tukey-Kramer	<.0001
M*Q	M2	Q2	M2	Q3	<.0001	Tukey-Kramer	<.0001
M*Q	M3	Q1	M3	Q2	<.0001	Tukey-Kramer	<.0001
M*Q	M3	Q1	M3	Q3	<.0001	Tukey-Kramer	<.0001
M*Q	M3	Q2	M3	Q3	0.6232	Tukey-Kramer	0.9998

SAS OUTPUT FOR QUESTION II

Brown and Forsythe's Test for Homogeneity of Y Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
TRT	8	59.5926	7.4491	0.50	0.8510
Error	45	672.8	14.9519		

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.984389	Pr < W	0.7025
Kolmogorov-Smirnov	D	0.070834	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.037375	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.237817	Pr > A-Sq	>0.2500

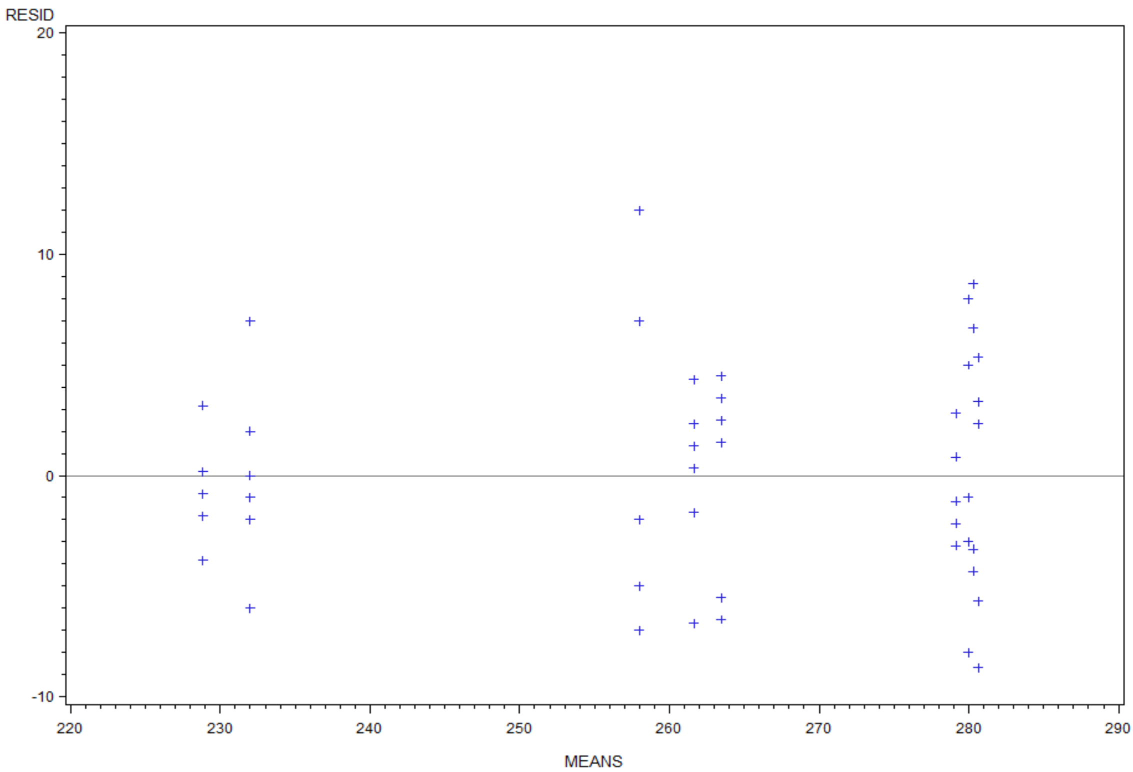
```
Stem Leaf                     #  Boxplot
  12 0                        1     |
  11                                |
  10                                |
   9                                |
   8 07                       2     |
   7 00                       2     |
   6 7                        1     |
   5 03                       2     |
   4 35                       2     |
   3 22335                    5  +-----+
   2 033588                   6  |     |
   1 35                       2  |     |
   0 0238                     4  |  +  |
  -0 8                        1  *-----*
  -1 872000                   6  |     |
  -2 200                      3  |     |
  -3 83320                    5  +-----+
  -4 33                       2     |
  -5 7500                     4     |
  -6 750                      3     |
  -7 0                        1     |
  -8 70                       2     |
     ----+----+----+----+
```

SAS OUTPUT FOR PROBLEM II

Variable: RESID

```
                   Normal Probability Plot
 12.5+                                         *
     |                                          ++
     |                                        ++
     |                                      ++
     |                                   ***
     |                                ****+
     |                               **+
  5.5+                             **
     |                           **
     |                        ****
     |                     ****
     |                    ***+
     |                  ***
     |                ***
 -1.5+              ***
     |            ++*
     |          +**
     |        ***
     |     ***
     |   * ***+
     | * +++
 -8.5+ *   ++
      +----+----+----+----+----+----+----+----+----+----+
          -2        -1         0        +1        +2
```

SAS OUTPUT FOR PROBLEM II



QUESTION III.

1. Let $X_1, \ldots, X_7$ be independent standard normal random variables. Identify the distribution of each of the following random variables. Be sure to justify your answers.

   (a) $U = X_1^2 + X_3^2 + X_5^2 + X_6^2 + X_7^2$.

   (b) $W = X_7/\sqrt{[X_1^2 + X_2^2 + X_3^2 + X_4^2 + X_5^2 + X_6^2]/6}$.

   (c) $Y = W^2 = 6X_2^2/[X_1^2 + X_3^2 + X_4^2 + X_5^2 + X_6^2 + X_7^2 + X_8^2]$.

   (d) $T = X_1/X_4$.

   (e) $S = 3(X_2^2 + X_4^2)/[2(X_1^2 + X + 3^2 + X_5^2)]$.

2. Let $X \sim N(2, 8)$ and $Y \sim N(-3, 5)$ be independent normal random variables. (Note: The notation $N(a, b)$ indicates a normal distribution with mean a and variance b.)

   (a) Let $U = 2X + 3Y - 5$ and $V = X - CY$, where C is a constant. Identify the distributions of U and V.

   (b) For U and V defined in part (a), what is the value of C that makes U and V independent?

   (c) Let
   $$W = C_1(X + C_2)^2 + C_3(Y + C_4)^2$$
   Find values of $C_1,\ C_2,\ C_3,\ C_4$, and C_5 (with $C_1 \neq 0$ and $C_3 \neq 0$) so that W has a χ^2 distribution with C_5 degrees of freedom.

   (d) Let
   $$T = \frac{C_1(X + C_2)^{C_3}}{(Y + C_4)^{C_5}}.$$
   Find values of $C_1,\ C_2,\ C_3,\ C_4,\ C_5$ and C_6 so that T has a t distribution with C_6 degrees of freedom.

   (e) Let
   $$T = \frac{C_1(X + C_2)^{C_3}}{(Y + C_4)^{C_5}}.$$
   Find values of $C_1,\ C_2,\ C_3,\ C_4,\ C_5,\ C_6$ and C_7 so that T has a F distribution with (C_6, C_7) degrees of freedom.

QUESTION IV:

1. The multiple regression matrix formulation is given by:

$$Y_{nx1} = X_{nx(p+1)}\beta_{(p+1)x1} + \varepsilon$$

Where:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

$E(\epsilon) = 0_{nx1}$ $\quad$ $V(\epsilon) = \sigma^2 I_{nxn}$

a. What distribution is usually assumed for ε?

b. Is it correct to test that Y has a normal distribution using a univariate test such as Sharipo-Wilks? Explain your answer.

c. What is I_{nxn}?

d. Assuming the conditions for MLR are met, how many unknown parameters are there? Display the unknown parameters.

e. If $\hat{\beta} = (X^t X)^{(-1)} X^t Y$ derive the expectation and variance of $\hat{\beta}$.

f. Is $\hat{\beta}$ an unbiased estimator of β? Explain your answer.

2. A researcher states that he has two models:

Model 1: $Y_{nx1} = X_{nx(p+1)}\beta_{(p+1)x1} + \varepsilon$ with R-squared = .7

Model 2: log(Y_{nx1}) $= X_{nx(p+1)}\beta_{(p+1)x1} + \varepsilon$ with R-squared = .8

Note: n/p = 50.

He also states that since the R-squared for Model 2 is greater than the R-squared for Model 1, then Model 2 is better than Model 1 .

Do you agree? Explain your answer.