MASTER'S DIAGNOSTIC EXAMINATION
August 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 the following:
   - Calculator which does not have capability to phone, text, or access the Web
   - Pencil or pen
   - Blank paper for the solutions 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 exam booklet or SAS output, 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.

A randomized trial was conducted to investigate the relationship between a continuous response y and four treatments `A`, `B`, `C`, and `D`. The sample size was $n = 200$, with 50 observations in each of the four treatment groups. Let $\boldsymbol{y}$ be the 200×1 vector of response values, ordered so that the first 50 entries are for treatment group `A`, the next 50 for `B`, then `C`, and finally `D`. The regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ was fit, where $\boldsymbol{X}$ is the 200×4 design matrix given by

$$\boldsymbol{X} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix},$$

and where each entry is a column vector of length 50. The estimated regression coefficients were

$$\hat{\boldsymbol{\beta}}' = (37.5, \ -11.5, \ 1.0, \ -27.7), \quad \text{with standard errors} \ \ 2.75, \ 3.89, \ 3.89, \ 3.89$$

and residual standard deviation $\hat{\sigma} = 19.45$.

(1.) Interpret each of the **four** regression parameters. As in, "the intercept, β_o, is the mean response when ...".

(2.) What assumptions are required for the regression model?

(3.) What is an approximate 95% confidence interval for the mean response in treatment group `B`?

`Hint`: If $\boldsymbol{v}$ is a 4×1 column vector, then the variance of

$\boldsymbol{v}'\hat{\boldsymbol{\beta}}$ is equal to $\hat{\sigma}^2 \boldsymbol{v}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{v}$.

In our example,

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.02 & -0.02 & -0.02 & -0.02 \\ -0.02 & 0.04 & 0.02 & 0.02 \\ -0.02 & 0.02 & 0.04 & 0.02 \\ -0.02 & 0.02 & 0.02 & 0.04 \end{pmatrix}$$

(4.) What is an approximate 95% confidence interval for the mean difference in response between treatment groups `B` and `A` (so, the difference $\mu_{\mathbf{B}} - \mu_{\mathbf{A}}$)?

(5.) Suppose the observations in treatment group `A` were positively correlated with those in treatment group `B`, and you correctly fit a correlated-data regression model. How would a 95% confidence interval for the mean difference in response between treatment groups `B` and `A` compare to the one reported in (4.) above, and why?

QUESTION II.

(1.) Define, using formulas in addition to words, the least squares criterion.

(2.) Show that the least squares criterion applied to the "intercept-only" model, i.e.,

$$y_i = \beta_o + \epsilon_i, \quad i = 1, 2, \ldots, n$$

results in the least squares estimator of β_o : $\hat{\beta}_o = \bar{y}$.

(3.) Consider the analysis of variance table for the simple linear regression model:

$$y_i = \beta_o + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n$$

What sum of squares in the analysis of variance table for the simple linear regression model would correspond to the sum of squares error, SSE, from the "intercept-only" model in part (2.)?

(4.) With reference to your answers in the above questions, **briefly** explain in words why statisticians almost always perform a "corrected" analysis of variance, a partition of $\sum_{i=1}^{n}(y_i - \bar{y})^2$, rather than an "uncorrected" analysis of variance, a partition of $\sum_{i=1}^{n} y_i^2$

QUESTION III

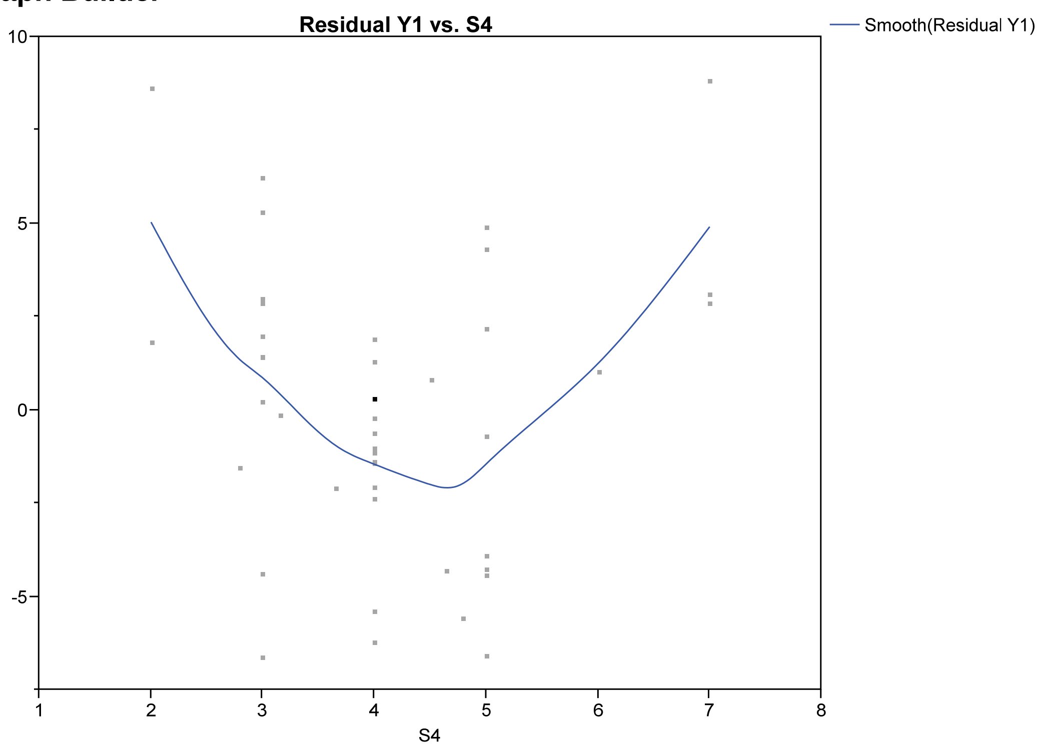
Part 1:

We have a dataset with y1 (a measure of diabetes & dependent variable) and 3 predictors BMI (Body mass index); BP (Blood pressure) and s4 a measure of glucose. We begin by considering the classical linear model relating y1 to the three explanatory variables.

Given the graph below (Residual is the "raw" residual), answer the following questions:

1. Based on the graph can you determine if the residuals are normally distributed? Explain.
2. Based on the graph below can you conclude that this is a valid model? Explain.
3. Does the graph suggest a change in the model? (That is: should you add a term; transform y1 etc.). Explain.

Graph Builder



Part 2:

Suppose that we want to predict whether a person will get the Tourette Syndrome based on Gender, HDL, Glucose & Chol. Note: Let Y = 1 if the Syndrome is detected and Y = 0 if not.

A logistic regression was run and the following output was obtained:

Gender	M	F
HDL	40	40
Glucose	125	125
Chol	190	190
Prob Y = 1	0.001	0.0001

1. What are the odds that a male with those characteristics will get the Syndrome? **Justify your answer.**
2. What are the odds that a female with those characteristics will get the Syndrome? **Justify your Answer.**
3. What is the odds ratio of a male getting the Syndrome as opposed to a female? **Justify your answer.**
4. The local newspaper front page story is: "Males are 10.01 times more likely to get Tourette Syndrome than Females. More research money is needed to help the males…"
   Is there something other than the odds ratio that needs to be considered here? Please explain your answer.

Part 3.

This part is similar to Part 1. Here we have y (a measure of diabetes & dependent variable) and 10 predictors: Age, Sex, BMI,BP, s1,s2,s3,s4,s5,s6.

A stepwise regression was run with the following results:

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-483.9509	71.29641	-6.79	<.0001*
SEX	-31.32679	16.06386	-1.95	0.0586
BP	1.7477267	0.568201	3.08	0.0039*
S2	0.6081179	0.320221	1.90	0.0652
S5	52.425839	16.93224	3.10	0.0037*
S6	2.11306	0.72281	2.92	0.0058*

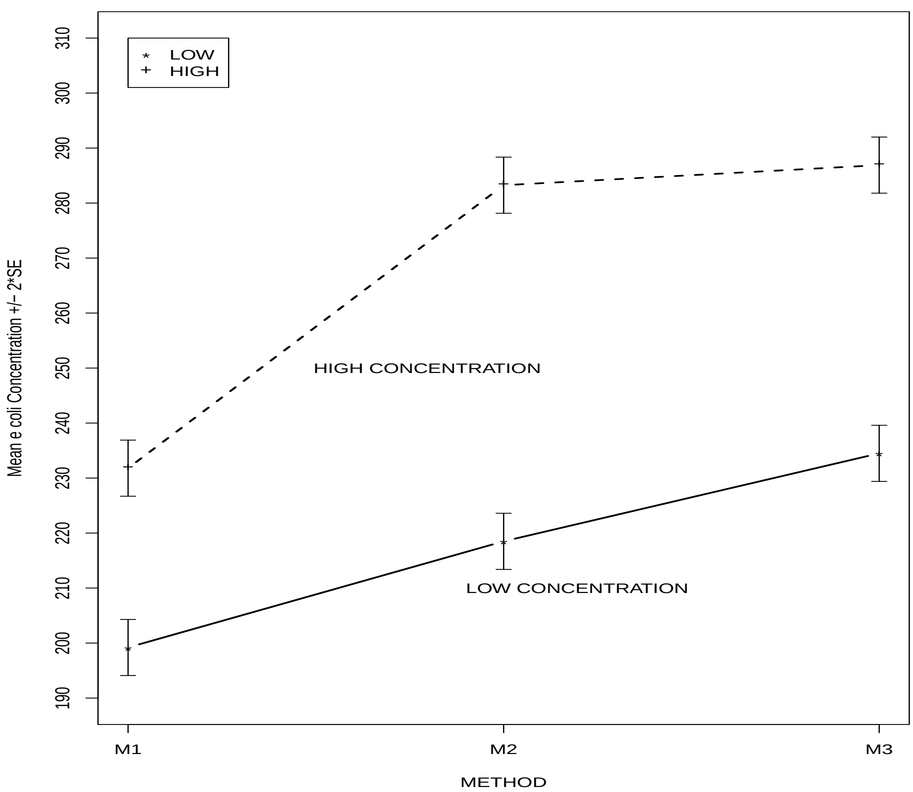
Since SEX and S2 are not significant at alpha= .05, can we remove them both from the model at the same time? Explain.

QUESTION IV.

The FDA is investigating methods to control *e coli* in beef products. There are three methods under consideration: M1, M2, and M3. The scientist at FDA decide to evaluate the three methods under two levels of *e coli* contamination, Low ($<$ 300 cfu/g) and High ($\geq$ 300 cfu/g). The researchers randomly selected six herds of cattle having a Low level of *e coli* contamination and six herds of cattle having a High level of *e coli* contamination. From each of the herds, six cattle were randomly selected for slaughter and then a meat sample from two cattle were randomly assigned to each of the three methods for *e coli* treatment. The amount of *e coli* was determined in a FDA lab and is reported in the following table. The three methods are the only methods under evaluation. The FDA is interested in determining which method provides the lowest *e coli* concentration after treatment.

	LEVEL OF CONTAMINATION											
	LOW						HIGH					
	HERD						HERD					
METHOD	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12
M1	203	215	203	190	194	191	225	228	232	235	233	237
	206	210	201	186	199	192	229	227	232	231	238	235
M2	222	223	215	217	218	215	277	276	289	284	285	286
	230	216	224	217	209	216	276	277	287	286	287	289
M3	239	245	237	235	236	234	280	278	282	292	296	295
	229	238	222	232	233	234	276	277	282	294	294	297

Profile Plot of METHOD*e coli Concentration



Use the above plot, 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.) Write a model for $y_{ijk\ell}$, the level of *e coli* in the ℓth meat sample from Herd k of Concentration i treated by Method j.

(3.) 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 $(df_{NUM.}, df_{DEN.})$.

(4.) Compute that the estimated standard error in the estimated mean difference of *e coli* concentrations between Methods M1 and M2 from a Low concentration.

(5.) Separate the three Methods of treating *e coli* 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 concentration of e coli. Use an experimentwise error rate of $\alpha = .05$.

(6.) Provide a 95% confidence on the mean *e coli* level of a High concentration meat sample treated by Method M3.

(7.) The researchers were criticized for using as their response variable the level of *e coli* concentration in the meat after the meat was treated by the three methods. Suggest a more appropriate modeling of the data still using the level of *e coli* concentration in the meat after the meat was treated as the response.

Methods Exam August 2012 SAS PROGRAM FOR PROBLEM II

```
ods html;ods graphics on;
OPTIONS LS=90 PS=55 nocenter nodate;
TITLE 'SAS OUTPUT FOR PROBLEM II';
DATA MANU;
INPUT C $ M $ H  $ Y @@;
TRT=COMPRESS (C) || COMPRESS (M);
LABEL  C="LEVEL  OF CONTAMINATION" M="METHOD OF CONTROL";
CARDS;

LOW M1 H1 203    LOW M1 H2  215   LOW M1 H3  203
LOW M1 H1 206    LOW M1 H2  210   LOW M1 H3  201
LOW M2 H1 222    LOW M2 H2  223   LOW M2 H3  215
LOW M2 H1 230    LOW M2 H2  216   LOW M2 H3  224
LOW M3 H1 239    LOW M3 H2  245   LOW M3 H3  237
LOW M3 H1 229    LOW M3 H2  238   LOW M3 H3  222
LOW M1 H4  190   LOW M1 H5 194    LOW M1 H6 191
LOW M1 H4  186   LOW M1 H5 199    LOW M1 H6 192
LOW M2 H4  217   LOW M2 H5 218    LOW M2 H6 215
LOW M2 H4  217   LOW M2 H5 209    LOW M2 H6 216
LOW M3 H4  235   LOW M3 H5 236    LOW M3 H6 234
LOW M3 H4  232   LOW M3 H5 233    LOW M3 H6 234
HIGH M1 H7 225   HIGH M1 H8 228   HIGH M1 H9 232
HIGH M1 H7 229   HIGH M1 H8 227   HIGH M1 H9 232
HIGH M2 H7 277   HIGH M2 H8 276   HIGH M2 H9 289
HIGH M2 H7 276   HIGH M2 H8 277   HIGH M2 H9 287
HIGH M3 H7 280   HIGH M3 H8 278   HIGH M3 H9 282
HIGH M3 H7 276   HIGH M3 H8 277   HIGH M3 H9 282
HIGH M1 H10 235  HIGH M1 H11 233  HIGH M1 H12 237
HIGH M1 H10 231  HIGH M1 H11 238  HIGH M1 H12 235
HIGH M2 H10 284  HIGH M2 H11 285  HIGH M2 H12 286
HIGH M2 H10 286  HIGH M2 H11 287  HIGH M2 H12 289
HIGH M3 H10 292  HIGH M3 H11 296  HIGH M3 H12 295
HIGH M3 H10 294  HIGH M3 H11 294  HIGH M3 H12 297
RUN;

PROC GLM;
CLASS C M H;
MODEL  Y = C M C*M H(C) M*H(C);
RANDOM H(C) M*H(C)/TEST;
LSMEANS C M C*M/STDERR PDIFF ADJUST=TUKEY;
RUN;

PROC MIXED CL ALPHA=.05 COVTEST;
CLASS C M H;
MODEL  Y = C M C*M /ddfm=satterth;
RANDOM H(C) M*H(C);
LSMEANS C M C*M/ ADJUST=TUKEY;
RUN;

PROC GLM;
CLASS TRT;
MODEL  Y = TRT;
MEANS TRT/HOVTEST=BF;
OUTPUT OUT=ASSUMP R=RESID P=MEANS;
PROC GPLOT; PLOT RESID*MEANS/VREF=0;
PROC UNIVARIATE DEF=5 PLOT NORMAL;
VAR RESID;
RUN;
ods graphics off;ods html close;
```

SAS OUTPUT FOR PROBLEM II

The GLM Procedure

Class Level Information		
Class	Levels	Values
C	2	HIGH LOW
M	3	M1 M2 M3
H	12	H1 H10 H11 H12 H2 H3 H4 H5 H6 H7 H8 H9

Number of Observations Read	72
Number of Observations Used	72

SAS OUTPUT FOR PROBLEM II

The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	35	77515.61111	2214.73175	184.99	<.0001
Error	36	431.00000	11.97222		
Corrected Total	71	77946.61111			

R-Square	Coeff Var	Root MSE	Y Mean
0.994471	1.427659	3.460090	242.3611

Source	DF	Type I SS	Mean Square	F Value	Pr > F
C	1	44900.05556	44900.05556	3750.35	<.0001
M	2	27135.02778	13567.51389	1133.25	<.0001
C*M	2	3143.02778	1571.51389	131.26	<.0001
H(C)	10	1611.22222	161.12222	13.46	<.0001
M*H(C)	20	726.27778	36.31389	3.03	0.0018

Source	DF	Type III SS	Mean Square	F Value	Pr > F
C	1	44900.05556	44900.05556	3750.35	<.0001
M	2	27135.02778	13567.51389	1133.25	<.0001
C*M	2	3143.02778	1571.51389	131.26	<.0001
H(C)	10	1611.22222	161.12222	13.46	<.0001
M*H(C)	20	726.27778	36.31389	3.03	0.0018

SAS OUTPUT FOR PROBLEM II

The GLM Procedure

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

SAS OUTPUT FOR PROBLEM II

The GLM Procedure
Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: Y

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
*	C	1	44900	44900	278.67	<.0001
	Error: MS(H(C))	10	1611.222222	161.122222		

*** This test assumes one or more other fixed effects are zero.**

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
*	M	2	27135	13568	373.62	<.0001
	C*M	2	3143.027778	1571.513889	43.28	<.0001
	H(C)	10	1611.222222	161.122222	4.44	0.0022
	Error: MS(M*H(C))	20	726.277778	36.313889		

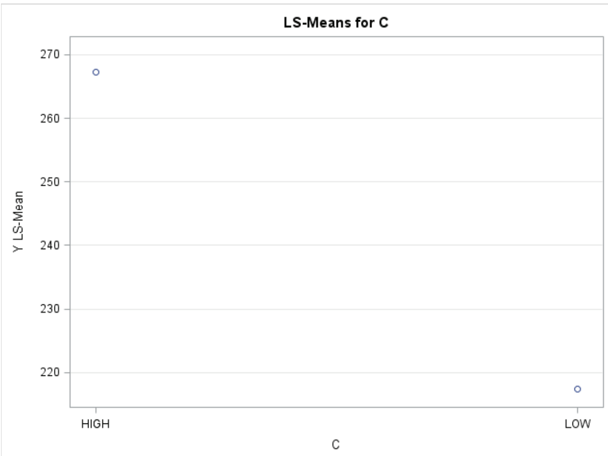
*** This test assumes one or more other fixed effects are zero.**

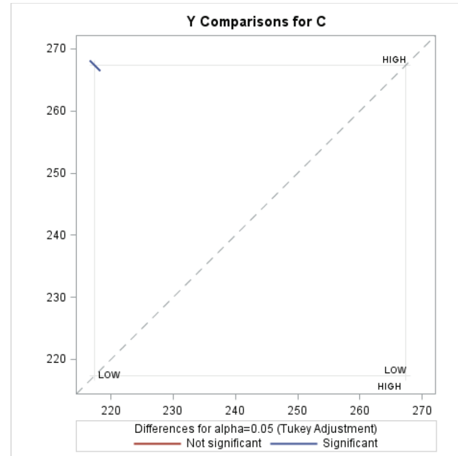
Source	DF	Type III SS	Mean Square	F Value	Pr > F
M*H(C)	20	726.277778	36.313889	3.03	0.0018
Error: MS(Error)	36	431.000000	11.972222		

SAS OUTPUT FOR PROBLEM II

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

			H0:LSMEAN=0	H0:LSMean1=LSMean2
C	Y LSMEAN	Standard Error	Pr > |t|	Pr > |t|
HIGH	267.333333	0.576682	<.0001	<.0001
LOW	217.388889	0.576682	<.0001	



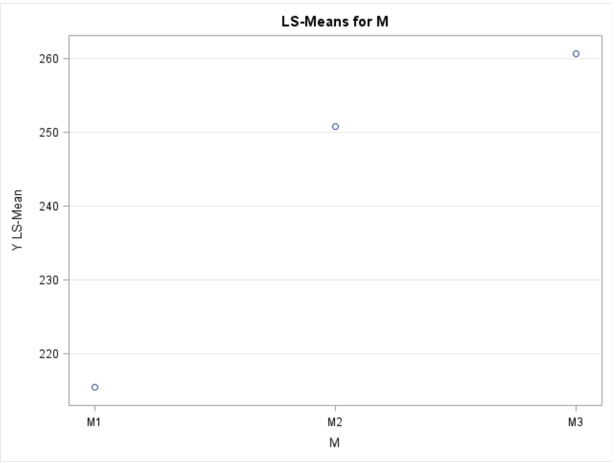


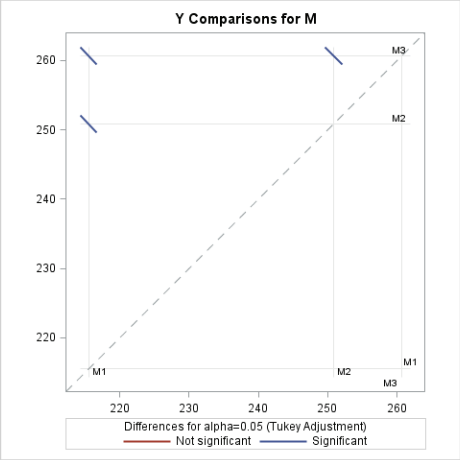
SAS OUTPUT FOR PROBLEM II

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

M	Y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
M1	215.500000	0.706288	<.0001	1
M2	250.875000	0.706288	<.0001	2
M3	260.708333	0.706288	<.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	<.0001
2	<.0001		<.0001
3	<.0001	<.0001	





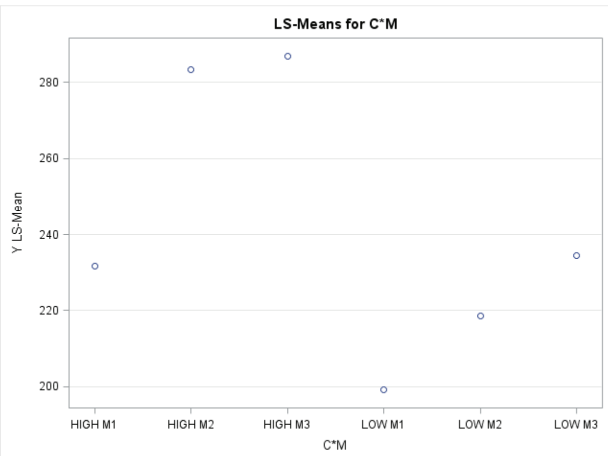
SAS OUTPUT FOR PROBLEM II

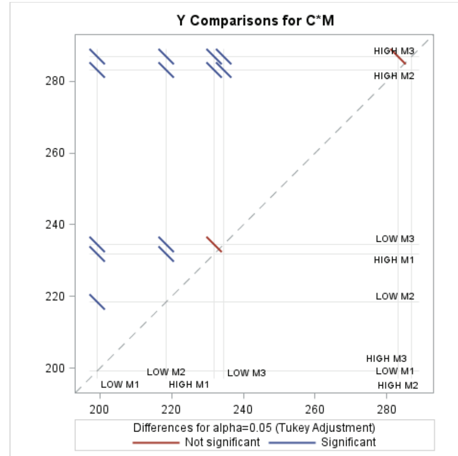
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

C	M	Y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
HIGH	M1	231.833333	0.998842	<.0001	1
HIGH	M2	283.250000	0.998842	<.0001	2
HIGH	M3	286.916667	0.998842	<.0001	3
LOW	M1	199.166667	0.998842	<.0001	4
LOW	M2	218.500000	0.998842	<.0001	5
LOW	M3	234.500000	0.998842	<.0001	6

Least Squares Means for effect C*M
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: Y

i/j	1	2	3	4	5	6
1		<.0001	<.0001	<.0001	<.0001	0.4257
2	<.0001		0.1245	<.0001	<.0001	<.0001
3	<.0001	0.1245		<.0001	<.0001	<.0001
4	<.0001	<.0001	<.0001		<.0001	<.0001
5	<.0001	<.0001	<.0001	<.0001		<.0001
6	0.4257	<.0001	<.0001	<.0001	<.0001	





SAS OUTPUT FOR PROBLEM II

The Mixed Procedure

Model Information	
Data Set	WORK.MANU
Dependent Variable	Y
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Satterthwaite

Class Level Information

Class	Levels	Values
C	2	HIGH LOW
M	3	M1 M2 M3
H	12	H1 H10 H11 H12 H2 H3 H4 H5 H6 H7 H8 H9

Dimensions	
Covariance Parameters	3
Columns in X	12
Columns in Z	48
Subjects	1
Max Obs Per Subject	72

Number of Observations	
Number of Observations Read	72
Number of Observations Used	72
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	448.81213510	
1	1	414.24817360	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr > Z	Alpha	Lower	Upper
H(C)	20.8014	12.1609	1.71	0.0436	0.05	8.5641	103.69
M*H(C)	12.1708	5.9125	2.06	0.0198	0.05	5.6533	42.5624
Residual	11.9722	2.8219	4.24	<.0001	0.05	7.9174	20.2007

Fit Statistics	
-2 Res Log Likelihood	414.2
AIC (smaller is better)	420.2
AICC (smaller is better)	420.6
BIC (smaller is better)	421.7

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
C	1	10	278.67	<.0001
M	2	20	373.62	<.0001
C*M	2	20	43.28	<.0001

Least Squares Means

Effect	LEVEL OF CONTAMINATION	METHOD OF CONTROL	Estimate	Standard Error	DF	t Value	Pr > |t|
C	HIGH		267.33	2.1156	10	126.36	<.0001
C	LOW		217.39	2.1156	10	102.76	<.0001
M		M1	215.50	1.8018	19.1	119.60	<.0001
M		M2	250.87	1.8018	19.1	139.23	<.0001
M		M3	260.71	1.8018	19.1	144.69	<.0001
C*M	HIGH	M1	231.83	2.5481	19.1	90.98	<.0001
C*M	HIGH	M2	283.25	2.5481	19.1	111.16	<.0001
C*M	HIGH	M3	286.92	2.5481	19.1	112.60	<.0001
C*M	LOW	M1	199.17	2.5481	19.1	78.16	<.0001
C*M	LOW	M2	218.50	2.5481	19.1	85.75	<.0001
C*M	LOW	M3	234.50	2.5481	19.1	92.03	<.0001

Differences of Least Squares Means

Effect	LEVEL OF CONTAMINATION	METHOD OF CONTROL	LEVEL OF CONTAMINATION	METHOD OF CONTROL	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
C	HIGH		LOW		49.9444	2.9919	10	16.69	<.0001	Tukey	<.0001
M		M1		M2	-35.3750	1.7396	20	-20.34	<.0001	Tukey-Kramer	<.0001
M		M1		M3	-45.2083	1.7396	20	-25.99	<.0001	Tukey-Kramer	<.0001
M		M2		M3	-9.8333	1.7396	20	-5.65	<.0001	Tukey-Kramer	<.0001
C*M	HIGH	M1	HIGH	M2	-51.4167	2.4601	20	-20.90	<.0001	Tukey-Kramer	<.0001
C*M	HIGH	M1	HIGH	M3	-55.0833	2.4601	20	-22.39	<.0001	Tukey-Kramer	<.0001
C*M	HIGH	M1	LOW	M1	32.6667	3.6036	19.1	9.06	<.0001	Tukey-Kramer	<.0001
C*M	HIGH	M1	LOW	M2	13.3333	3.6036	19.1	3.70	0.0015	Tukey-Kramer	0.0154
C*M	HIGH	M1	LOW	M3	-2.6667	3.6036	19.1	-0.74	0.4683	Tukey-Kramer	0.9743
C*M	HIGH	M2	HIGH	M3	-3.6667	2.4601	20	-1.49	0.1517	Tukey-Kramer	0.6739
C*M	HIGH	M2	LOW	M1	84.0833	3.6036	19.1	23.33	<.0001	Tukey-Kramer	<.0001
C*M	HIGH	M2	LOW	M2	64.7500	3.6036	19.1	17.97	<.0001	Tukey-Kramer	<.0001
C*M	HIGH	M2	LOW	M3	48.7500	3.6036	19.1	13.53	<.0001	Tukey-Kramer	<.0001
C*M	HIGH	M3	LOW	M1	87.7500	3.6036	19.1	24.35	<.0001	Tukey-Kramer	<.0001
C*M	HIGH	M3	LOW	M2	68.4167	3.6036	19.1	18.99	<.0001	Tukey-Kramer	<.0001

C*M	HIGH	M3	LOW	M3	52.4167	3.6036	19.1	14.55	<.0001	Tukey-Kramer	<.0001
C*M	LOW	M1	LOW	M2	-19.3333	2.4601	20	-7.86	<.0001	Tukey-Kramer	<.0001
C*M	LOW	M1	LOW	M3	-35.3333	2.4601	20	-14.36	<.0001	Tukey-Kramer	<.0001
C*M	LOW	M2	LOW	M3	-16.0000	2.4601	20	-6.50	<.0001	Tukey-Kramer	<.0001

SAS OUTPUT FOR PROBLEM II

The GLM Procedure

Class Level Information		
Class	Levels	Values
TRT	6	HIGHM1 HIGHM2 HIGHM3 LOWM1 LOWM2 LOWM3

Number of Observations Read	72
Number of Observations Used	72

SAS OUTPUT FOR PROBLEM II

The GLM Procedure

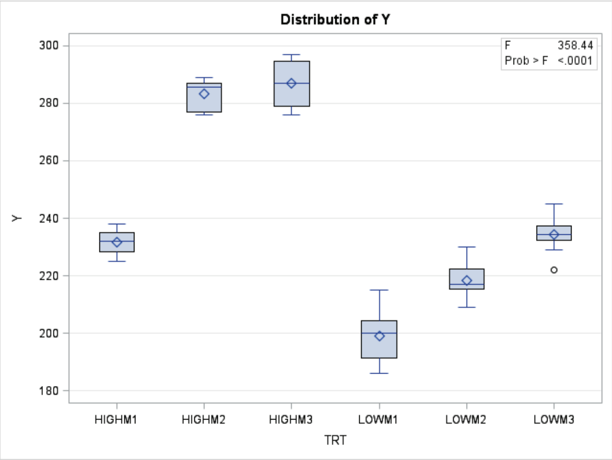
Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	75178.11111	15035.62222	358.44	<.0001
Error	66	2768.50000	41.94697		
Corrected Total	71	77946.61111			

R-Square	Coeff Var	Root MSE	Y Mean
0.964482	2.672313	6.476648	242.3611

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRT	5	75178.11111	15035.62222	358.44	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	5	75178.11111	15035.62222	358.44	<.0001



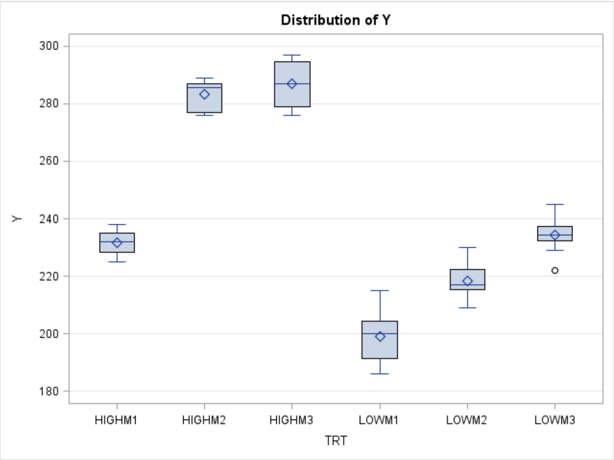
SAS OUTPUT FOR PROBLEM II

The GLM Procedure

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	5	230.1	46.0222	3.55	0.0066
Error	66	854.8	12.9520		

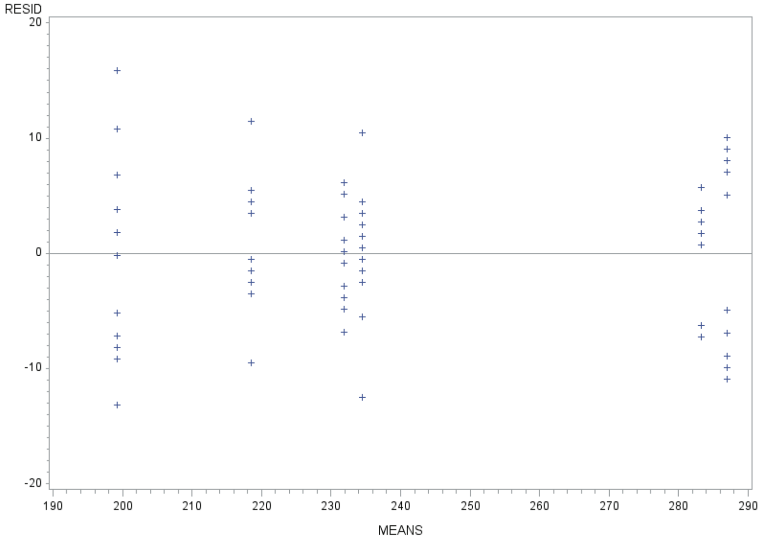
SAS OUTPUT FOR PROBLEM II

The GLM Procedure



Level of TRT	N	Y Mean	Y Std Dev
HIGHM1	12	231.833333	4.04145188
HIGHM2	12	283.250000	5.18958748
HIGHM3	12	286.916667	8.36071260
LOWM1	12	199.166667	8.78876694
LOWM2	12	218.500000	5.45227225
LOWM3	12	234.500000	5.61653403

SAS OUTPUT FOR PROBLEM II



SAS OUTPUT FOR PROBLEM II

The UNIVARIATE Procedure
Variable: RESID

Moments			
N	72	Sum Weights	72
Mean	0	Sum Observations	0
Std Deviation	6.24443414	Variance	38.9929577
Skewness	0.02301717	Kurtosis	-0.4374699
Uncorrected SS	2768.5	Corrected SS	2768.5
Coeff Variation	.	Std Error Mean	0.73591362

Basic Statistical Measures			
Location		Variability	
Mean	0.00000	Std Deviation	6.24443
Median	0.16667	Variance	38.99296
Mode	-7.25000	Range	29.00000
		Interquartile Range	9.08333

Note: The mode displayed is the smallest of 14 modes with a count of 2.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	1	Pr >= |M|	0.9063
Signed Rank	S	-2	Pr >= |S|	0.9911

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.990551	Pr < W	0.8709
Kolmogorov-Smirnov	D	0.059062	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.03672	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.222575	Pr > A-Sq	>0.2500

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	15.833333
99%	15.833333
95%	10.500000
90%	7.083333
75% Q3	4.166667
50% Median	0.166667
25% Q1	-4.916667
10%	-8.166667
5%	-9.916667
1%	-13.166667
0% Min	-13.166667

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-13.16667	22	10.0833	72
-12.50000	18	10.5000	14
-10.91667	52	10.8333	5
-9.91667	53	11.5000	10
-9.50000	29	15.8333	2

```
   Stem Leaf                     #  Boxplot
     14 8                        1     |
     12                                |
     10 1585                     4     |
      8 11                       2     |
      6 2811                     4     |
      4 5512588                  7  +-----+
      2 58822558888             11  |     |
      0 22582588                 8  *--+--*
     -0 55585552                 8  |     |
     -2 8558555                  7  |     |
     -4 52998                    5  +-----+
     -6 2229822                  7     |
     -8 95292                    5     |
    -10 9                        1     |
    -12 25                       2     |
        ----+----+----+----+


                    Normal Probability Plot
     15+                                                +*+
       |                                             +++
       |                                         **+*+*
       |                                       **++
       |                                    ****
       |                                 ****
       |                             *****
      1+                          ***+
       |                       ***
       |                    ****
       |                 +***
       |              *****
       |          ****
       |       +*+
    -13+ *  ++*
        +----+----+----+----+----+----+----+----+----+----+
            -2        -1         0        +1        +2
```