85/100
+2
=
87/100
D.A

STATISTICS 642 - EXAM I

March 2, 2022

Student's Name Jack Rodoni

Student's Email Address jsodoni@tamu.edu

INSTRUCTIONS FOR STUDENTS:

(1) The exam consists of the coverpage, 6 pages of Questions, 1 page of SAS code, and 5 pages of SAS output.

(2) You may start the exam at 8am (Texas Time) on March 2, 2022.

(3) You must submit your solutions prior to 9am (Texas Time) on March 2, 2022.

(4) Make sure to write using a black pen or write using a pencil in such a manner that I can clearly read your solutions.

(5) Do not discuss or provide information to anyone concerning the questions on this exam or your solutions until I post the solutions to the exam.

(6) You may use the following:

- Calculator - Your device cannot facilitate a connection to the internet or to send text messages
- Summary Sheets - **3-pages, 8.5"x11" (you may write/type/cut-paste on both sides of the three sheets)**
- Tables for STAT 642 which will be attached or you can bring a copy of the tables to the exam.

(7) Do not use any other written material except for your summary sheets and the STAT 642 tables.

(8) Do not use a computer, cell phone, or any other electronic device (other than a calculator) except to look at Stat 642 tables or your own formulas. No typing is allowed on the computer.

I attest that I spent no more than the designated time 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

3

Problem I. (30 points) **CIRCLE** (A, B, C, D, or E) corresponding to the **BEST** answer. Only **ONE LETTER** should be **CIRCLED** for each of the 10 questions.

$t=9$ $\quad b=12,\ k=3$

(1) We have 9 different detergents (A,B,C,D,E,F,G,H). There are only three basins that are used simultaneously at the same rate with a different detergent in each session. Response is number of plates until foam disappears in a basin. Our research question includes determining the differences (if any) in average plates of detergent. Here is the screen shot of the data layout in this blocked design.

1	2	3	4	5	6	7	8	9	10	11	12
A	D	G	A	B	C	A	B	C	A	B	C
B	E	H	D	E	F	E	F	D	F	D	E
C	F	I	G	H	I	I	G	H	H	I	G

$r=4$ $\quad N=36$

$N = 9(4) = (12)(3) = bk$

$\lambda = \frac{4(2)}{8} = 1$ $\quad 1<4<12$

$12 \geq 9$

Is this a Balanced Incomplete Block Design (BIBD)?

A. This is a not a BIBD design

B. This is a BIBD design with $\lambda = 2$.

✓ C. This is a BIBD design with $\lambda = 1$.

D. This is a BIBD design with $\lambda = 0.5$.

E. It is not possible to answer this question with this limited information.

(2) A company is designing a new product. There are 4 producers of raw material. The goal of a study was to evaluate the consistency of the raw material's physical properties across multiple batches of material from each producer and then study the variation within each batch. The experiment consisted of randomly select 7 batches of material from each of the 4 producers. From each of the 28 batches, 4 samples are taken and the physical properties were then determined in the company's lab. The factors are Producer (P), Batch (B), Sample (S). Which of the following best represent the experiment modeling.

A. P, B, S, P*B, P*S, B*S, P*B*S

B. P, B, S, B*S

C. P, B, S(B)

✓ D. P, B(P), S(B,P)

E. P, B(P), S

(3) Which of the following is an incorrect statement?

A. Sum of SS for t-1 mutually orthogonal contrast always add up to SSTreatment when there are equal number of replications

✓ B. Sum of SS for t-1 mutually orthogonal contrast always add up to SSTreatment when there are unequal number of replications

C. t-1 mutually orthogonal contrasts may include the polynomial trends

D. mutually orthogonal contrasts aid proving independence

E. if all pairwise contrasts are orthogonal then they are mutually orthogonal

(4) Which of the following is not about Fisher's LSD?

A. lack of an exact value of familwise error rate

✓ B. high power

C. indicate more differences than you may really have

D. experimentwise error rate is neither exact nor bounded by nominal values

E. highly recommended when determining mean differences

(5) Suppose we have a completely randomized design with 4 treatments which are randomly selected a population of treatments. The number of replications per treatment are $n_1 = 6, n_2 = 4, n_3 = 7, n_4 = 5$. In computing the power of the AOV F-test,

the distribution of the ratio of MS_{TRT} to MSE is

A. central F

B. noncentral F

C. central chisquare

D. noncentral chisquare

E. none of the above

degrees of freedom is

A. df1=3 and df2=22

B. df1=3 and df2=18

C. df=22

D. df1=4 and df2=18

E. df1=4 and df2=22

(6) In a completely randomized design with $t = 5$ fixed effects treatments and $n_1 = 6,\ \ n_2 = 3,\ \ n_3 = 5,\ \ n_4 = 3,\ \ n_5 = 3,$ reps/treatment. The following model is fit to the data: $Y_{ij} = \mu + \tau_i + e_{ij}$. A statistical program which has the standard constraint on the parameters yields the following least squares estimates of five of the six parameters in the model:

$$\hat{\mu} = 2.7; \quad \hat{\tau}_1 = -1.6; \quad \hat{\tau}_2 = 2.5; \quad \hat{\tau}_3 = 3.6; \quad \hat{\tau}_4 = -6.2$$

The least squares estimate of τ_5 is given by

A. $\hat{\tau}_5 = 0$

B. $\hat{\tau}_5 = 1.7$

C. $\hat{\tau}_5 = 0.9$

D. The program has an error, $\hat{\tau}_5$ is not estimable if sample sizes are unequal.

E. Not enough information has been provided to answer the question.

need to know constraint. $\tau_t = 0$ or $\sum n_i \tau_i = 0$.

(7) In an experiment to compare the tensile strengths of 5 different types of copper wire, a pooled t-test with $\alpha = .05$ was used to group the types of wire according to equal pairwise mean strength. The major problem that will result from this procedure is

A. the experimentwise Type I error rate will be much larger than the nominal $\alpha = .05$ of the t-test.

B. the number of tests that need to be conducted is excessive.

C. the assumption of independence will be violated.

D. the Type II error rate will not be constant.

E. none of the above, this is the proper procedure.

✓ (8) If a F-test is used to test the significance of 12 contrasts constructed from 22 treatment means, at what value must α_{PC} be set in order to achieve a familywise error rate of α_F =0 .01?

A. 0.01 because the F-test is insensitive to multiple testing.

B. 0.01/12 because there were 12 tests conducted.

C. 0.01/22 because there were 22 population means

D. 0.01/264 to adjust for both the number of tests and population means.

E. none of the above

(9) A completely randomized design with t = 4 fixed effects treatments and n_1 = 6; n_2 = 3; n_3 = 5; n_4 = 3, reps/treatment was run. The experimenter constructed two contrasts in the treatment means:
C1 = $-3\mu_1 - \mu_2 + \mu_3 + 3\mu_4$ $-3 + 1 - 1 + 3 = 0$
C2 = $\mu_1 - \mu_2 - \mu_3 + \mu_4$
Which one of the following statements is True?

✓ A. The two contrasts are orthogonal.

B. The sample estimators of the two contrasts are uncorrelated.

C. The sample estimators of the two contrasts are independent.
→ not independent b/c unequal sample sizes.

D. The two contrasts are not orthogonal.

E. none of the above statements are true

Problem II. (40 points) Steel is heat-treated by heating above the critical temperature, soaking, and then air cooling. This process increases the strength of the steel, refines the grain, and homogenizes the structure. An experiment is performed to determine the effect of temperature on the strength of heat-treated steel. There are five temperatures of interest 1500, 1600, 1700, 1800, 1900F. The experiment is performed by heating the oven to one of the selected temperature and then inserting a specimen of the steel which is randomly selected from a population of specimens. The specimen is removed, a new temperature is selected, and the process is repeated until three specimens have been treated by each of the five temperatures for a total of 15 treated specimens. The individual specimens were cut into four equal sized squares and the strength (MPa) for each of the four squares is determined. The results are given in the following table: The data was analyzed yielding the following summary statistics for the strenght at each of the temperatures:

1500	5.0, 5.5, 4.0, 3.5	3.5, 3.5, 3.0, 4.0	4.5, 4.0, 4.0, 5.0
1600	5.0, 4.5, 5.0, 4.5	5.5, 6.0, 5.0, 5.0	5.5, 4.5, 6.5, 5.5
1700	7.0, 9.0, 8.5, 8.5	6.0, 7.0, 7.0, 7.0	11.0, 7.0, 9.0, 8.0
1800	8.5, 6.0, 9.0, 8.5	6.5, 7.0, 8.0, 6.5	7.0, 7.0, 7.0, 7.0
1900	6.0, 5.5, 3.5, 7.0	6.0, 8.5, 4.5, 7.5	6.5, 6.5, 8.5, 7.5

Temp	$\bar{y}_{i.}$	S_i	n_i
1500	4.1250	0.7424	15
1600	5.2083	0.6201	15
1700	7.9167	1.3624	15
1800	7.3333	0.9374	15
1900	6.4583	1.4994	15
Overall	6.2083	1.7474	75

Use the attached SAS output to assist you in answering the following questions. You must provide a justification for each of your answers.

1. At the $\alpha = 0.05$ level, does there appear to be a difference in the mean strength of the steel for the different heat-treat temperatures? Justify your answer. If you find differences, clearly identify the differences and again justify your answer.

Yes, there do seem to be differences. First check overall F

-2

$H_0: \mu_i = \mu_k \quad i \neq k \qquad H_a: \mu_i \neq \mu_k \quad i \neq k.$

• Looking at the tukey procedure we see that the mean responses for the treatment pairs (1500, 1700), (1500, 1800), (1500, 1900), (1600, 1700), (1600, 1900) are all significantly different at the $\alpha = 0.05$ level w/ p-values ≤ 0.0001

Also, the mean respons for the treatment pair (1700, 1900) ~~are~~ significantly different at the $\alpha = 0.05$ level w/ p-value = 0.0148.

2. At the $\alpha = 0.05$ level, is there a significant trend in the average strength as the heat-treatment temperature is increased? Justify your answer.

Yes, there is a significant trend in the average strength as the heat treatment temp is increased. Looking at our fit of orthogonal polynomials we see that both the linear trend and the quadratic trends are significant w/ p-values < 0.0001. Also, the quartic trend is significant at the $\alpha = 0.05$ level w/ p-value = 0.0029 > $\frac{0.05}{4}$ = 0.0125.

see the solution

$\alpha_{PC} = 1-(1-0.05)^{\frac{1}{4}} = 0.01274$ -2

We have model: $y_{ij} = \beta_0 + \beta x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$

our hypotheses are $H_{01}: \beta_1 = 0$, $H_{02}: \beta_2 = 0$ $H_{03}: \beta_3 = 0$ $H_{04}: \beta_4 = 0$

at the $\alpha = 0.05$ level we are rejecting H_{01}, H_{02} & H_{04} and concluding that these polynomial terms are significant.

b/c contrasts are orthogonal

3. Place a 95% confidence interval on the average strength of a specimen heat-treated with a temperature of 1900F.

$v = 55$, $\alpha = 0.025$

CI: $6.458\bar{3} \pm t_{0.05/2,\,60-5}\sqrt{\frac{1.1837121}{12}}$

CI $\approx 6.458\bar{3} \pm 2.021\sqrt{\frac{1.1837121}{12}}$ $= (5.8236, 7.0931)$

note the CI is actually a little wider than required as $t_{0.025,55} < 2.021$

Q: Where are we supposed to get $t_{0.025,55} = 2.004$? In the tables I gave the closest approximation

4. A new experiment is to be conducted using the five heat-treatment temperatures in order to evaluate the differences in the strength of specimens more precisely. The metallurgists want at least a 90% chance of detecting a difference of at least 1.25 MPa in the mean strength of any pair of temperatures using an $\alpha = 0.05$ test. The process engineer states that r = 20 would be the minimum rep size needed to achieve these specifications. Is the engineer correct? Assume $\sigma_e \approx 1.1$. Justify your answer.

$\gamma = 0.90$ $D = 1.25$ $\alpha = 0.05$ $r = 20$ $v_1 = 4$, $v_2 = 95$

$\Rightarrow \lambda = \frac{20(1.25)^2}{1.1^2} = 25.82644628$ -1

I forgot to divide by 2: $\lambda = \frac{rD^2}{2\sigma_e^2}$ *

$\Phi = \sqrt{\lambda/r} = \sqrt{\frac{25.82644628}{20}} = 1.1363\bar{6}$

No. The engineer is not correct. Looking at table IX w/ $v_1 = 4$, $v_2 = 95$ we see that the power of this test would only be ≈ 0.45. Well short of the specified power of 0.90.

Problem III. (30 points) We wish to study the effects of anxiety and muscular tension on four different types of memory. Twelve subjects are assigned to one of four anxietytension combinations at random. The low-anxiety group is told that they will be awarded \$5 for participation and \$10 if they remember sufficiently accurately, and the high-anxiety group is told that they will be awarded \$5 for participation and \$100 if they remember sufficiently accurately. Everyone must squeeze a spring-loaded grip to keep a buzzer from sounding during the testing period. The high-tension group must squeeze against a stronger spring than the low-tension group. All subjects then perform four memory trials in random order, testing four different types of memory. The response is the number of errors on each memory trial. The following table includes Number of memory errors by type, tension, and anxiety level; subjects are columns. Subject's age also is available and can be used for data analysis if needed.

Anxiety	1	1	1	1	1	1	2	2	2	2	2	2
Tension	1	1	1	2	2	2	1	1	1	2	2	2
Type 1	18	19	14	16	12	18	16	18	16	19	16	16
Type 2	14	12	10	12	8	10	10	8	12	16	14	12
Type 3	12	8	6	10	6	5	8	4	6	10	10	8
Type 4	6	4	2	4	2	1	4	1	2	8	9	8

Provide the details for each of the following items:

1. Type of Randomization:

   Completely randomized design -2

2. Type of Treatment Structure:

   2x2 factorial design w/ ~~crossover design treatment structure~~

   -4 explain

3. Identify each of the factors as being Fixed or Random:

   Treatment factors: Anxiety (2 levels, fixed)
   Tension (2 levels, fixed)

   type? -1

   Blocking factor: sequence in which memory trial conducted (12 levels, fixed)
   Test subjects (12 levels, random)

4. Describe the experimental units:

   - EU is a test subject nested w/in a sequence.

5. Describe the measurement units:

   - MU is a test subject evaluated on a particular memory test during a particular ~~testing~~ period

6. Identify any covariates:

   Subjects age

SAS Program:

```
ods html;ods graphics on;
option ls=120 ps=60 nocenter nodate;
title 'STAT 642 EXAM I - Spring 2022 ';
DATA HEATTREAT;
ARRAY Y Y1-Y4;
INPUT T $ S $ Y1-Y4; DO OVER Y; H=Y; OUTPUT; END;
LABEL T = 'TEMP' H = 'STRENGTH';
CARDS;
1500 1 5.0 5.5 4.0 3.5
1500 2 3.5 3.5 3.0 4.0
1500 3 4.5 4.0 4.0 5.0
1600 1 5.0 4.5 5.0 4.5
1600 2 5.5 6.0 5.0 5.0
1600 3 5.5 4.5 6.5 5.5
1700 1 7.0 9.0 8.5 8.5
1700 2 6.0 7.0 7.0 7.0
1700 3 11.0 7.0 9.0 8.0
1800 1 8.5 6.0 9.0 8.5
1800 2 6.5 7.0 8.0 6.5
1800 3 7.0 7.0 7.0 7.0
1900 1 6.0 5.5 3.5 7.0
1900 2 6.0 8.5 4.5 7.5
1900 2 6.5 6.5 8.5 7.5
RUN;
PROC GLM ORDER = DATA;
CLASS T;
MODEL H=T /ss3;
LSMEANS T/STDERR PDIFF ADJUST=TUKEY CL;
MEANS T / HOVTEST=BF;
contrast 'LINEAR' T -2 -1 0 1 2;
contrast 'QUADRATIC' T 2 -1 -2 -1 2;
contrast 'CUBIC' T -1 2 0 -2 1;
contrast 'QUARTIC' T 1 -4 6 -4 1;
estimate 'LINEAR' T -2 -1 0 1 2;
estimate 'QUADRATIC' T 2 -1 -2 -1 2;
estimate 'CUBIC' T -1 2 0 -2 1;
estimate 'QUARTIC' T 1 -4 6 -4 1;
output out=ASSUMP r=RESID p=MEANS;
RUN;
proc gplot; plot H*T='*';run;
proc univariate def=5 plot normal;
var RESID;
RUN;
ods graphics off;
ods html close;
```

Jack Rodan

'STAT 642 EXAM I - Spring 2022 '

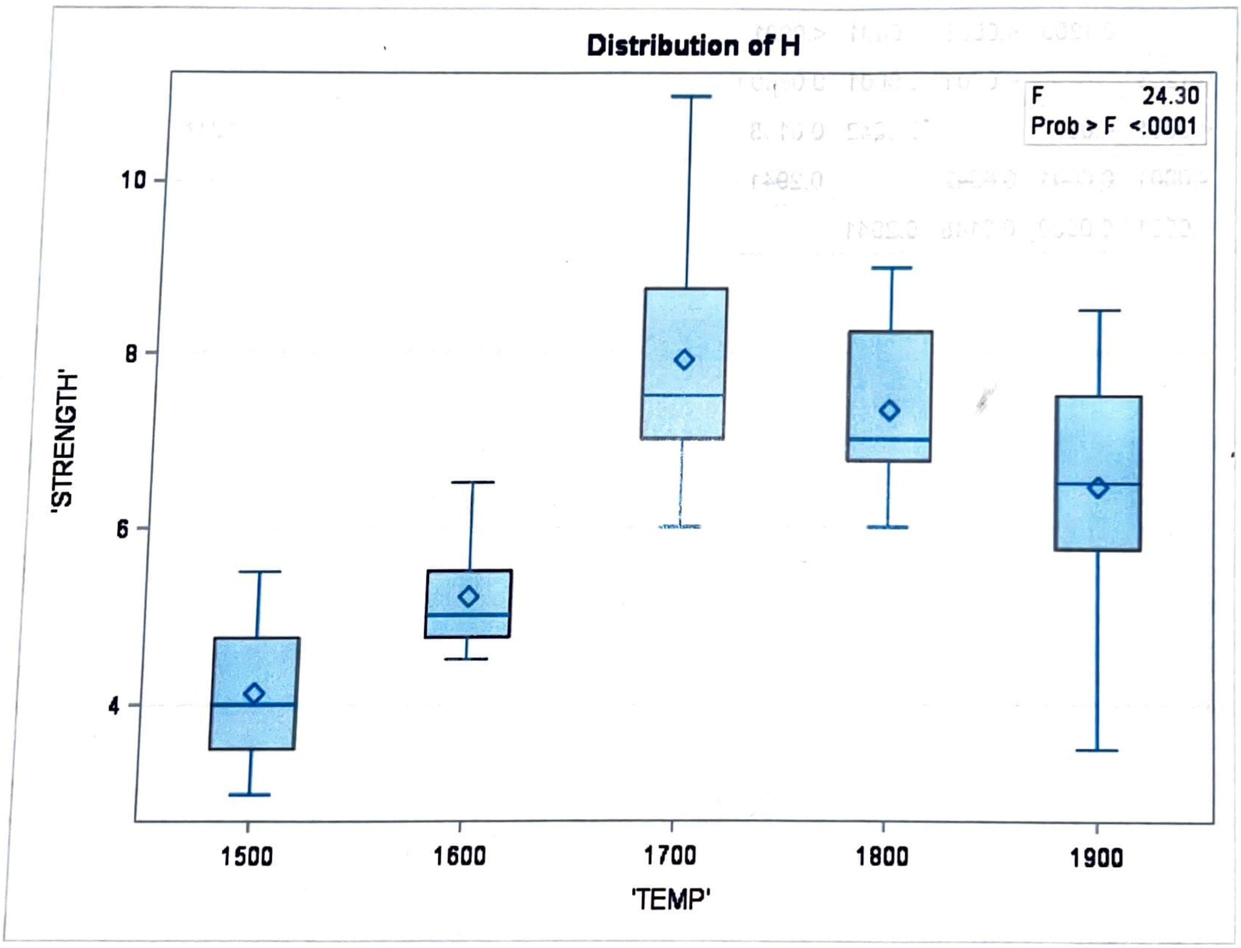
The GLM Procedure

Dependent Variable: H 'STRENGTH'

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	115.0416667	28.7604167	24.30	<.0001
Error	55	65.1041667	1.1837121		
Corrected Total	59	180.1458333			

R-Square	Coeff Var	Root MSE	H Mean
0.638603	17.52460	1.087985	6.208333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
T	4	115.0416667	28.7604167	24.30	<.0001



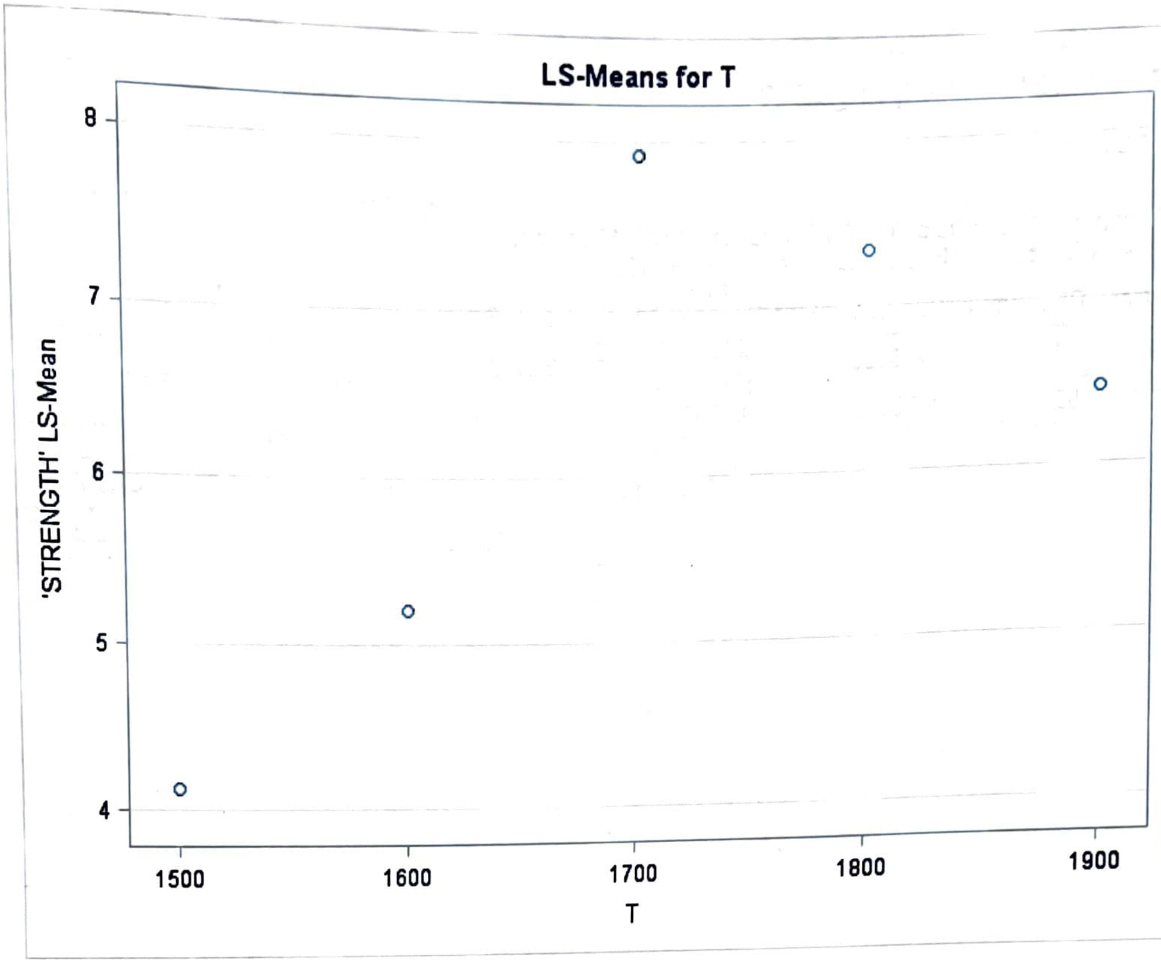
'STAT 642 EXAM I - Spring 2022 '

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

T	H LSMEAN	LSMEAN Number
1500	4.12500000	1
1600	5.20833333	2
1700	7.91666667	3
1800	7.33333333	4
1900	6.45833333	5

Least Squares Means for effect T
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: H

i/j	1	2	3	4	5
1		0.1203	<.0001	<.0001	<.0001
2	0.1203		<.0001	0.0001	0.0508
3	<.0001	<.0001		0.6842	0.0148
4	<.0001	0.0001	0.6842		0.2941
5	<.0001	0.0508	0.0148	0.2941	


LS-Means for T
'STRENGTH' LS-Mean
8
7
6
5
4
1500
1600
1700
1800
1900
T

'STAT 642 EXAM I - Spring 2022 '

The GLM Procedure

Brown and Forsythe's Test for Homogeneity of H Variance ANOVA of Absolute Deviations from Group Medians					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
T	4	4.6083	1.1521	2.22	0.0792
Error	55	28.6042	0.5201		

'STAT 642 EXAM I - Spring 2022 '

The GLM Procedure

Dependent Variable: H 'STRENGTH'

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
LINEAR	1	55.35208333	55.35208333	46.76	<.0001
QUADRATIC	1	44.53720238	44.53720238	37.63	<.0001
CUBIC	1	4.40833333	4.40833333	3.72	0.0588
QUARTIC	1	10.74404762	10.74404762	9.08	0.0039
H matrix	2	15.37847222	7.68923611	6.50	0.0029