STATISTICS 642 - EXAM I

June 26, 2014

Student's Name _____

Student's Email Address _____

INSTRUCTIONS FOR STUDENTS:

(1) The Exam consists of 12 pages including this cover page and SAS output.

(2) You have exactly 120 minutes to complete the exam.

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

(4) You may use the following:

- Textbook for **ONLY the TABLES** and not any of the other parts of the book.
- Calculator
- Summary Sheets: 6-Pages (3 letter sized paper ($8.5'' \times 11''$), front and back)

I attest that I spent no more than 120 minutes to complete the exam. I used only the materials described above. I did not receive assistance from anyone during the taking of this exam and I did not provide assistance to another student in the completion of this exam.

Student's Signature_____

INSTRUCTIONS FOR PROCTOR:

Immediately after the student completes the exam:

- Collect all portions of this exam. Do not allow the student to take any portion of the exam with her/him.
- Please send **JUST pages 1, 2, 3, 4, 5, and 6** of the exam to me in the following manner:
  - Scan **ONLY pages 1, 2, 3, 4, 5, and 6** of the exam to a pdf file and then have the student upload the pdf file into WebAssign

(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, and that the exam was scanned into a pdf and uploaded to webassign in my presence.

Proctor's Signature_____

Disclaimer: This exam has been reviewed for consistency with the posted lectures by Dr. Michael Longnecker.

Problem I. (30 points) A medical specialist wants to compare two different methods (M1, M2) for treating a particular illness. She will use 8 hospitals for the study. Because there may be differences in the response between hospitals, she will block on hospitals. Each hospital has four wards of patients. She will randomly select four patients in each ward to take part in the study. Within each hospital, the wards are segregated by the sex of the patient, with two wards for female patients and two for male patients. Within each hospital, one ward of female patients and one ward of male patients will be randomly selected to receive method M1. The other wards will receive method M2. All patients in a ward receive the same treatment. Values of the response variable will be recorded for each of the 128 patients.

Provide the details for each of the following items:

1. Type of Randomization: CRD, RCBD, LSD, Split-Plot, Crossover, etc.

2. Type of Treatment Structure: Single factor, Crossed, Nested, Fractional, etc.

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

4. Describe the Experimental Units and Measurement Units:

5. Describe the Measurement Process: Response Variable, Subsampling, Repeated Measures, Covariates, etc.

6. How many replications are in this experiment?

Problem II. (30 points) The wrinkle resistance of synthetic fiber used to make cloth for men's shirts is of interest to a manufacturer. The researcher wanted to determine the degree to which wrinkle resistance is affected by the percentage of cotton in the fiber. Five levels of cotton percentage are used in the study: 15%, 20%, 25%, 30%, 35%. Five specimens of fabric are selected for each level of cotton percentage. The data was analyzed yielding the following summary statistics for the wrinkle resistance of the synthetic fiber (psi):

% Cotton	Measurements					$\bar{y}_{i.}$	S_i^2	n_i
15	7	7	9	11	15	9.80	11.20	5
20	12	17	12	18	18	15.40	9.80	5
25	14	18	18	19	19	17.60	4.30	5
30	19	25	22	19	23	21.60	6.80	5
35	12	10	11	15	11	11.80	3.70	5
Overall						15.24	7.16	25

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

1. Separate the cotton percentages into groups of percentages such that all members of the group are not significantly different from any other member of the group with respect to their average wrinkle resistance Use $\alpha_F = .05$.

2. The researcher wants to model the average wrinkle resistance as a function of percent cotton in the fabric. Which ones, if any, of the trends in the average wrinkle resistance as a function of percentage of cotton are significant? Use $\alpha_F = 0.05$ in reaching your conclusions?

3. Display the Hypothesis matrix, H, needed to simultaneously test the linear, quadratic, cubic, and quartic trend contrasts in the average wrinkle resistance across the five percentages of cotton.

4. A second experiment similar to the cotton percentage experiment was planned to evaluate a different type of fabric using the same five percentages of cotton. The textile researcher states that she wants the probability of detecting a difference of at least 4 units in the average wrinkle resistance between at least one pair of cotton percentages to be at least 0.90 when using an $\alpha = .05$ F-test. If the experiment uses 13 reps of each of the 5 percentages of cotton, will the researcher's goal be achieved? Assume $\sigma_e^2 \approx 8$. Justify your answer.

Problem III. (40 points) Answer each of the following questions using at most **20 words**.

1. If a Bonferroni F-test is used to test the significance of 8 orthogonal contrasts constructed from 15 treatment means, at what value must α_{PC} be set in order to obtain an familywise error rate of $\alpha_F = .01$?

2. A researcher conducts a CRD experiment with four treatments. The number of replications are
   $n_1 = 2, \quad n_2 = 3, \quad n_3 = 2, \quad n_4 = 4$ for the four treatments. She constructs two contrasts:
   $$C_1 = -3\mu_1 - \mu_2 + \mu_3 + 3\mu_4 \quad \text{and} \quad C_2 = -\mu_1 + 3\mu_2 - 3\mu_3 + \mu_4$$
   Are C_1 and C_2 orthogonal? Justify your answer. Are their corresponding sums of squares, SS_{C_1} and SS_{C_2} independent? Justify your answer.

3. An experiment was conducted in which there is a single factor having $t = 4$ levels but each EU is measured at $b = 10$ specified times after receiving the treatment. The researcher states that the design was a split-plot design with the 10 times at which the measurements are taken designated as the split-plot factor. She states that the EU can be considered as being split into 10 parts which serve as the split-plot EU's. Why is this experiment not a valid split-plot experiment?

4. Both the Scheffé procedure and the Tukey HSD procedure have exact familywise error rates. Why is the number of pairs of treatments declared to be significantly different by the Tukey HSD procedure generally greater than the number of pairs which would be declared different using the Scheffé procedure?

5. A researcher conducts a CRD experiment with three treatments. The number of replications are

$$n_1 = 3, \quad n_2 = 2, \quad n_3 = 4 \qquad \text{for the three treatments.}$$

The **effects** model for this experiment, $y_{ij} = \mu + \tau_i + e_{ij}$ can be expressed as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. Under the standard constraint on the τ_i's, display the design matrix $\mathbf{X}$ explicitly for this version of the effects model. Write the vector $\boldsymbol{\beta}$ in terms of the treatment means, μ_1, μ_2, μ_3.

6. A senior statistician at your firm states that she would never use LSD or SNK as a multiple comparison procedure. What would be a possible reason for her making this statement?

7. In a complete randomized design with a single factor having t levels with n_i reps for the ith treatment, the following sum of squares: $\sum_{i=1}^{t}\sum_{j=1}^{n_i}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$ is computed. What source of variation is this sum of squares measuring?

8. A paper is discussing a completely randomized design with $t = 5$ fixed effects treatments and sample sizes $n_1 = 6, \quad n_2 = 3, \quad n_3 = 2, \quad n_4 = 3, \quad n_5 = 4$ reps/treatment. The following model is fit to the data: $Y_{ij} = \mu + \tau_i + e_{ij}$. The researcher uses SAS to compute the least squares estimates of the six parameters in the model but the paper only reports the following five estimates:

$$\hat{\mu} = 3; \quad \hat{\tau}_1 = 1; \quad \hat{\tau}_2 = 2; \quad \hat{\tau}_3 = 4; \quad \hat{\tau}_4 = -8$$

Compute the least squares estimate of the five treatment means $\mu_1,\ \mu_2,\ \mu_3,\ \mu_4,\ \mu_5$.

SAS Program:

```
* examI 2014.sas;
ods html; ods graphics on;
option ls=120 ps=60 nocenter nodate;
title ' ';

data new; array Y Y1-Y5;
input P $ Y1-Y5; do over Y; S=Y; output; end;
      drop Y1-Y5;
      label P = '% of Cotton' S = 'Strength of Fabric';
cards;
15         7   7  15  11   9
 20        12  17  12  18  18
 25        14  18  18  19  19
 30        19  25  22  19  23
 35         12  10  11  15  11
RUN;
proc glm ORDER = DATA;
class P;
model S = P/ss3;
MEANS P/TUKEY DUNNETT('15') ALPHA=.05;
LSMEANS P/PDIFF ADJUST = TUKEY;
contrast 'LINEAR' P -2 -1 0 1 2;
contrast 'QUADRATIC' P 2 -1 -2 -1 2;
contrast 'CUBIC' P -1 2 0 -2 1;
contrast 'QUARTIC' P 1 -4 6 -4 1;
estimate 'LINEAR' P -2 -1 0 1 2;
estimate 'QUADRATIC' P 2 -1 -2 -1 2;
estimate 'CUBIC' P -1 2 0 -2 1;
estimate 'QUARTIC' P 1 -4 6 -4 1;
run;
proc gplot;
plot S*P='*';
run;
ods graphics off; ods html close;
```