STATISTICS 642 - ASSIGNMENT 3

DUE DATE: 8am Central, THURSDAY, February 17, 2022

Name (**Typed**) ____

Email Address (**Typed**) ____

Make this cover sheet the first page of your Solutions.

STATISTICS 642 - ASSIGNMENT #3

- Read Handout 3
- Supplemental Reading: Chapter 3 in the Design & ANOVA book
- Hand in the following problems by 8am Central, THURSDAY, February 17, 2022:

Problem 1. (15 points) - An animal physiologist designed a study to evaluate the effect of several forced molt regimens used by egg producers to bring hens back into egg production. Twenty-five hens were randomly selected to be used in the experiment. There are five stages in the regimen: S1-Premolt (control); S2-Fasting for 8 days; S3- 60 grams of bran per day for 10 days; S4- 80 grams of bran per day for 10 days; S5- Laying mash for 42 days. The objective was to follow various physiological responses associated with the pituitary function on the hens during the regimen to aid in explaining why the hens come back into egg production. One of the measurements was the serum T3 concentration in the pituitary gland. Five hens were randomly assigned to be sacrificed at the end of each of the five stages. The serum T3 measurements of the five hens sacrificed at the end of each stage of the regimen are given in the following table.

Treatment	Serum T3 (ng/dl)$\times 10^{-1}$				
Premolt	94.09	90.45	99.38	73.56	74.39
Fasting	98.81	103.55	115.23	129.06	117.61
60 g Bran	197.18	207.31	177.50	226.05	222.74
80 g Bran	102.93	117.51	119.92	112.01	101.10
Laying Mash	83.14	89.59	87.76	96.43	82.94

a. Write a model for this study and explain all terms in your model.

b. State the assumptions needed to perform an analysis of variance for this study.

c. Produce the AOV table for this study.

d. Compute the least squares estimators of the treatment means and their estimated standard errors.

e. Compute 95% confidence intervals for the treatment means.

f. Is there significant evidence at the $\alpha = .05$ level that the average T3 level differs across the five stages?

Problem 2. (10 points) - Refer to the previous problem. Suppose that during the experiment several of the 25 chickens were not available for measurement resulting in the following data.

Treatment	Serum T3 (ng/dl)$\times 10^{-1}$				
Premolt	94.09	90.45	99.38	73.56	
Fasting	98.81	103.55	115.23	129.06	117.61
60 g Bran	197.18	207.31	177.50		
80 g Bran	102.93	117.51	119.92	112.01	101.10
Laying Mash	83.14	89.59	87.76	82.94	

a. Produce the AOV table for this study.
b. Compute the least squares estimators of the treatment means and their estimated standard errors.
c. Compute 95% confidence intervals for the treatment means.
d. Is there significant evidence at the $\alpha = .05$ level that the average T3 level differs across the five stages?

Problem 3. (15 points) - A traffic engineering firm contacted you to assist in designing a study of traffic delay at intersections with signal on urban streets. Three types of signals were to be used in the study: TS1- Pretimed; TS2- Semi-actuated; TS3- Fully actuated. The measure of traffic delay used in the study will be the average stopped time per vehicle (seconds/vehicle) on a Tuesday during the period 7 am to 6 pm at each of r randomly selected intersections. What is the minimum number of intersections, r, that would be needed for each of the three signaling devices so that an $\alpha = .01$ test would have a probability of 90% to detect that the true mean traffic delays were $\mu_1 = 20$, $\mu_2 = 18$, and $\mu_3 = 16$ seconds per vehicle? Use $\hat{\sigma}_e^2 = 12$ in your solution

Problem 4. (15 points) - Refer to Problem 1. What is the minimum number of chickens, r, that the researcher should assign to each of the five treatments in order that an $\alpha = .05$ test would have probability of at least 90% to detect a difference of at least 30 ng/dl$\times 10^{-1}$ units of T3 serum between any pair of treatments? Use $\hat{\sigma}_e^2 = 150$ in your solution

Problem 5. (20 points) - For an experiment with four treatments and number of reps given by

$n_1 = 3$, $n_2 = 4$, $n_3 = 5$, $n_4 = 3$, the following models were proposed:

- Cell means Model: $y_{ij} = \mu_i + e_{ij}$
- Effects Model: $y_{ij} = \mu + \tau_i + e_{ij}$

Answer the following questions:

a. Write out the Design Matrix for the Cell Means Model
b. Write out the Design Matrix for the Effects Models with no constraints
c. Write out the Design Matrix for the Effects Models with the constraint: $\tau_4 = 0$
d. For the Effects Model with no constraint express τ_i in terms of μ and μ_i
e. For the Effects Model with the constraint $\tau_4 = 0$ express τ_i in terms of μ and μ_i

Problem 6. (10 points) - The Cell Means model, $y_{ij} = \mu_i + e_{ij}$ has as one of its assumptions that the y_{ij}'s are random, independent observations from the treatment populations.

a. What could the statistician do during the conduct of the experiment in order to ensure that the condition of random, independent observations is reasonable valid?
b. What condition is required of the y_{ij}'s in order for the least squares estimators of μ_i to be best linear unbiased estimators?
c. What condition is required of the y_{ij}'s in order to validly use the F-test in AOV hypotheses and to place confidence intervals on the μ_{ij}'s?
d. Suppose the condition from part c. is not valid. How could you test for differences in the treatment means?

Problem 7. (15 points) - A completely randomized experiment was conducted to investigate the potential difference in four treatments `A`, `B`, `C`, and `D`. The sample size was $n = 200$, with 50 observations in each of the four treatment groups. Let $\mathbf{y}$ 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 $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$ was fit, where $\mathbf{X}$ is the 200×4 design matrix given by

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

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

$$\hat{\beta}' = (\hat{\beta}_o,\ \hat{\beta}_1,\ \hat{\beta}_2,\ \hat{\beta}_3) = (37.5,\ -11.5,\ 1.0,\ -27.7), \quad \text{with standard errors } 2.75,\ 3.89,\ 3.89,\ 3.89$$

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

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

b. Compute a 95% confidence interval for the mean response in treatment group `B`?

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

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

   In our example,

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

c. Compute a 95% confidence interval for the mean difference in response between treatment groups `B` and `A` that is, the difference $\mu_{\mathbf{B}} - \mu_{\mathbf{A}}$?