Jack Rodoni
STAT 642
HW 03
2/17/22

Read H.O.3, Ch 3 in Design & ANOVA books.

1.) (a) Write a model for this study and explain all the terms in your model

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \quad <=> \quad y_{ij} = \mu_i + \varepsilon_{ij}$$

See H.O.3 pg 19

(b) State the assumptions needed to perform an analysis of variance in this study.

- The ε_{ij}'s $\overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$
- The $y_{ij} \sim N(\mu_i, \sigma_\varepsilon^2)$ under each treatment i
- The y_{ij}'s are independent

H.O.3 pg 21 → (c) Produce the ANOV table for this study

see code pg. 24 →

Source of variation	df	Sum of Squares	Mean Square	F	P-value
Treatment	4	48569	12142	78.08	6.48×10^{-12}
Error	20	3110	156		
Total	24	51679			

see code pg 24 → (d) Compute the least squares estimates of the treatment means and their estimated std errors

- [Use LSmeans() function in R]

$$\underline{\mu} = \begin{bmatrix} \text{Pretreat} \\ \text{Fasting} \\ \text{60g bran} \\ \text{80g bran} \\ \text{Laxative plain} \end{bmatrix} = [86.4, 112.9, 206.2, 110.7, 88.0]'$$

$$SE(\underline{\mu}) = [5.58, 5.58, 5.58, 5.58, 5.58]'$$

(e) Compute 95% CI for the treatment means:

μ_1: (74.7, 98.0)

μ_2: (101.2, 124.5)

μ_3: (194.5, 217.8)

μ_4: (99.1, 122.3)

μ_5: (76.3, 99.6)

(f) Is there significant evidence at the $\alpha = 0.05$ level that the average T3 level differs across the 5 stages?

- Yes, we can see from the ANOVA table above that we have significant evidence, $P(F \geq 78.08) \leq 0.05$

2) (a) Produce the ANOVA table for this study.

Source of Variation	df	Sum of Squares	Mean Square	F	P-value
Treatment	4	24121	6030	54.06	4.29×10^{-9}
Error	16	1785	112		
Total	20				

(b) Compute the least squares estimates of the treatment means & their estimated std errors.

$\underline{\mu} = [89.4, 112.9, 194.0, 110.7, 89.2]'$

$SE(\underline{\mu}) = [5.28, 4.72, 6.10, 4.72, 5.28]'$

(c) Compute 95% confidence intervals for the treatment means.

μ_1: (78.2, 101)

μ_2: (102.8, 123)

μ_3: (181.1, 207)

μ_4: (100.7, 121)

μ_5: (78.0, 100)

(d) Yes, we can see from the anova table that we have significant evidence $P(F > 54.06) \leq 0.05$

3.) What is the number of intersections, r, that would be needed for each of the three cycling directions so that an $\alpha = 0.01$ test would have a probability of 90% to detect that the true mean traffic delays were $\mu_1 = 20$, $\mu_2 = 18$, & $\mu_3 = 16$ seconds per vehicle? Use $\hat{\sigma}_\varepsilon^2 = 12$ in your solution.

- using power.anova.test in R: (see H.O.3 pg 34)
  - means = c(20, 18, 16)
  - power.anova.test(groups = 3, n = , between.var = var(means), within.var = 12, sig.level = 0.01, power = 0.90)
    $\Rightarrow$ r = 28

4.) Refer to problem 1. What is the minimum number of chickens, r, that the researcher should assign to each of the 5 treatments in order that an $\alpha = 0.05$ test would have probability of at least 90% to detect a difference of at least 30 ng/dl $\times 10^{-1}$ unit of T3 serum between any pair of treatments? Use $\hat{\sigma}_\varepsilon^2 = 150$ in your solution.

(see discussion of approaches on H.O.3 pgs 39-41 (see code at pg 45))

- using code from H.O.3 pg 45: (Approach 5)
  we get $r = n_i = 7$ $\Rightarrow$ 35 total chickens

5.) For an experiment w/ four treatments and number of exp 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 cells means model

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$\underline{x}_1$: a column starting w/ 3 (1)s then 12 (0)s

$\underline{x}_2$: a comm starting w/ 3 (0)s, 4 (1)'s then 8 (0)'s

$\underline{x}_3$: a column starting w/ 7 (0)'s, then 5 (1)'s followed by 3 (0)'s

$\underline{x}_4$: a column starting w/

(b) Write out the design matrix for the effects model w/ No constraints.

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(c) write out the design matrix for the effects model w/ constraint: $\tau_4 = 0$

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

see #0.3 pg 49

(d) For the effects model w/ no constraint express τ_i interms of μ_i; μ_i

$$\mu_i = E[Y_{ij}] = \mu + \tau_i \iff \boxed{\tau_i = \mu_i - \mu \quad i = 1,2,3,4}$$

see H.O.3 pg TO

(e) For the effects model w/ the constraint $\tau_4 = 0$, express τ_i interms of μ; μ_i

$$\tau_i = \mu_i - (\mu + \tau_i) = \boxed{\tau_i = \mu_i - \mu_4 \quad i = 1,2,3}$$

6.) 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 reasonably valid.

- To ensure the condition of random independent observations is reasonably valid, I would randomly assign our EUs to the treatments.

HW 5 see Pg 19

(b) What condition is required of the y_{ij}'s in order for the least squares estimators of μ_i to be their unbiased estimators?

- The e_{ij}'s are iid w/ a normal distribution.

(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_j's?

HW 3 Pg 10

- $y_{ij} \sim N(\mu_i, \sigma_e^2)$ and are independent.

(d) If the condition from part c is not valid. How could you test for differences in the treatment means?

- We could increase our sample size. Then, by the CLT, each μ_i would be approximately normally distributed.

see p. 50

7.) (a) Interpret each of the 4 regression parameters

β_0: the mean response of treatment group A.

β_1: the difference between the treatment mean of group B and treatment mean of group A

β_2: " " " C " " "

β_3: " " " D " " "

see p. 52, 29

(b) compute a 95% CI for the mean response in treatment group B?

$\mu_B = \beta_0 + \beta_1 \Rightarrow \underline{v}'\underline{\beta}$ w/ $\underline{v}' = [1\ 1\ 0\ 0]$

$$\text{Var}(\underline{v}'\hat{\beta}) = \sigma_\varepsilon^2 \underline{v}'(X'X)^{-1}\underline{v} = \sigma_\varepsilon^2 (0.02) = 19.45^2(0.02) = 0.378$$

95% CI: $\mu_B \pm t_{0.025,196} \frac{\hat{\sigma}_\varepsilon}{\sqrt{50}} = (31.5 - 11.5) \pm q_t(0.975,196) \frac{19.45}{\sqrt{50}}$

$= (20.57534, 31.42466)$

see p. 29

(c) Compute a 95% CI for the mean difference in response between treatment groups B & A, that is the difference $\mu_B - \mu_A$?

[Note: $\mu_B - \mu_A = \beta_1$]

95% CI: $\mu_B - \mu_A \pm t_{0.025,196} \frac{\hat{\sigma}_\varepsilon}{\sqrt{\frac{1}{50} + \frac{1}{50}}} = \mu_B - \mu_A \pm t_{0.025,196} \frac{19.45}{5} =$

$= (-11.5) \pm t_{0.025,196} (19.45/5)$

$= (-19.207124, -3.792872)$

[Note:

$\mu_B - \mu_A = \beta_0 + \beta_1 - \beta_0 = \beta_1$]