Stat 642 Spring 2022 - Solutions for Homework 3

P1. (15 points) Chicken Problem

(a) $y_{ij} = \mu_i + e_{ij},\ i = 1, \cdots, 5,\ j = 1, \cdots, 5$, where μ_i is the mean of ith treatment population and e_{ij} is the random deviation from the mean.

(b) The e_{ij} must be independent identically distributed as $N(0, \sigma_e^2)$ random variables, that is, we have independent observations from 5 normal populations which have the same standard deviation, σ_e, but possibly different means, μ_i.

(c) With $N = 25,\ t = 5$, the following values are obtained from SAS output:

Source	df	SS	MS	F-value	p-value
Model	4	48569	12142.3	78.08	6.48×10^{-12}
Error	20	3110	155.5		
Total	24	51679			

(d) (i) LSE of μ_i: $\hat{\mu}_i = \bar{y}_{i\cdot},\ i = 1, 2, 3, 4.$
Thus, $\hat{\mu}_1 = 86.37, \hat{\mu}_2 = 112.85, \hat{\mu}_3 = 206.16, \hat{\mu}_4 = 110.69, \hat{\mu}_5 = 87.97.$

(ii) Estimated standard error of $\hat{\mu}_i$ is $\sqrt{\frac{\hat{\sigma}_e^2}{r}} = \sqrt{\frac{155.5}{5}} = 5.5767$

(e) A 95% C.I. on μ_i is given by $\hat{\mu}_i \pm (t_{.025,N-t})(\hat{\sigma}_{\hat{\mu}_i})$ with $t_{.025,N-t} = t_{0.025,20} = 2.086$:

μ_1 : $86.37 \pm (2.086)(5.5767) = (74.74,\ 98.00)$

μ_2 : $112.85 \pm (2.086)(5.5767) = (101.22,\ 124.48)$

μ_3 : $206.16 \pm (2.086)(5.5767) = (194.53,\ 217.79)$

μ_4 : $110.69 \pm (2.086)(5.5767) = (99.06,\ 122.32)$

μ_5 : $87.97 \pm (2.086)(5.5767) = (76.34,\ 99.60)$

SAS code for obtaining standard errors and 95% C.I.'s :

LSMEANS TRT/STDERR CL ALPHA=.05;

(f) $F = \frac{MS_{Model}}{MSE} = \frac{12142.3}{155.5} = 78.08 > F_{0.05,4,20} = 2.866 \Rightarrow$
Reject $H_0 : \mu_1 = \cdots = \mu_5$ with p-value$= P[F_{4,20} \geq 78.08] = 1 - pf(78.08, 4, 20) = 6.4 \times 10^{-12} < .05.$
Thus, we conclude that there is significant evidence (p-value$<$ .0001) in the data that the mean T3 serum levels differ across the five regimens.

P2. (10 points)

a. SAS Output:

```
                                  Sum of
Source                      DF       Squares    Mean Square   F Value   Pr > F
Model                        4   24852.09337     6213.02334     57.56   <.0001
Error                       16    1726.89814      107.93113
Corrected Total             20   26578.99151

Least Squares Means
                               Standard
S          T3 LSMEAN              Error    Pr > |t|
PRE         89.370000           5.194495      <.0001
FAST       112.852000           4.646098      <.0001
60G        193.996667           5.998087      <.0001
80G        110.694000           4.646098      <.0001
LM          85.857500           5.194495      <.0001
```

b. (i) LSE of μ_i: $\hat{\mu}_i = \bar{y}_{i\cdot}$, $i = 1, 2, 3, 4, 5$.
Thus, $\hat{\mu}_1 = 89.37,\ \hat{\mu}_2 = 112.852,\ \hat{\mu}_3 = 193.997,\ \hat{\mu}_4 = 110.694,\ \hat{\mu}_5 = 85.858$.

(ii) Standard error: $\hat{\sigma}_{\hat{\mu}_i} = \sqrt{\frac{107.93}{n_i}}$, $\Rightarrow$

$\hat{\sigma}_{\hat{\mu}_1} = 5.1945,\ \hat{\sigma}_{\hat{\mu}_2} = 4.6461,\ \hat{\sigma}_{\hat{\mu}_3} = 5.9981,\ \hat{\sigma}_{\hat{\mu}_4} = 4.6461,\ \hat{\sigma}_{\hat{\mu}_5} = 5.1945$

The LSE's have changed somewhat (one increased, four decreased). Also, the LSE's now have unequal standard errors whereas whereas with equal reps the LSE have the same value for their standard errors.

c. A 95% C.I. on μ_i is given by $\hat{\mu}_i \pm (t_{.025,n-t})(\hat{\sigma}_{\hat{\mu}_i})$ with $t_{.025,n-t} = t_{0.025,16} = 2.120$:

μ_1 :$89.370 \pm (2.12)(5.1945) = (78.36,\ 100.38)$

μ_2 : $112.852 \pm (2.12)(4.6461) = (103.00,\ 122.70)$

μ_3 : $193.997 \pm (2.12)(5.9981) = (181.28,\ 206.71)$

μ_4 : $110.694 \pm (2.12)(4.6461) = (100.84,\ 120.54)$

μ_5 : $85.858 \pm (2.12)(5.1945) = (74.85,\ 96.87)$

d. $F = \frac{MST}{MSE} = \frac{6213.0}{107.9} = 57.56 > F_{0.05,4,16} = 3.01 \Rightarrow$ Reject $H_0 : \mu_1 = \cdots = \mu_5$ with
p-value= $P[F_{4,16} \geq 57.56] = 1 - pf(57.56, 4, 16) = 2.69 \times 10^{-9}$.

Thus we conclude that there is significant evidence (p-value $< .0001$) of a difference in the mean serum T3 concentration for the five stages of molting.

P3. (15 points) The μ_i's are specified as $\mu_1 = 20, \mu_2 = 18, \mu_3 = 16$. Also, $\alpha = 0.01, \gamma_0 = 0.90, t = 3, \hat{\sigma}_e^2 = 12$. The noncentrality parameter of the F-distribution is

$$\lambda = \frac{r}{\sigma^2}\sum_{i=1}^{t}(\mu_i - \bar{\mu}_\cdot)^2 \ \Rightarrow L = \frac{r}{\hat{\sigma}^2}\{(20-18)^2 + (18-18)^2 + (16-18)^2\} = \frac{2r}{3}$$

Using the SAS code for Approach 3: repsize_Approach3.sas,
we obtain the following results.

Obs	r	n1	n2	Fcr	Lambda	POWER
8	27	2	78	4.88809	18.0000	0.88983
9	28	2	81	4.87720	18.6667	0.90395
10	29	2	84	4.86713	19.3333	0.91646

Select smallest value of r such that the power satisfies: $\gamma(L) \geq 0.90 = \gamma_o \ \Rightarrow \ r = 28$. To meet the power specification, it would be necessary to use 28 intersections at each of the three types of signals.

Using the following R Code and D=4, the values of r using Approach 4 and 5 are obtained:

```
r = c(4,5,6,7,8,26,27,28)
L4 =  4*r
L5 = 2*r/3
f = qf(.99,2,3*(r-1))
p4 = 1-pf(f,2,3*(r-1),L4)
p5 = 1-pf(f,2,3*(r-1),L5)
cbind(r,p4,p5)
```

Using Approach 4 with D = 4 yields r = 7

Using Approach 5 with D = 4 yields r = 28

P4. (15 points) The problem specifies $\alpha = 0.05, \gamma_0 = 0.90, t = 5, \hat{\sigma}_e^2 = 150$, $D = 30$. By using the SAS code for Approach 5:

```
Using the SAS program: repsize_Approach5.sas,
we obtain the following power values for selected values of r

Obs   t     r     u1     u2     L      phi         p
1     5     2      4      5     6    1.09545    0.22103
2     5     3      4     10     9    1.34164    0.45296
3     5     4      4     15    12    1.54919    0.65342
4     5     5      4     20    15    1.73205    0.79720
5     5     6      4     25    18    1.89737    0.88866
6     5     7      4     30    21    2.04939    0.94201
7     5     8      4     35    24    2.19089    0.97111
```

With $L = \lambda \geq \frac{rD^2}{2\hat{\sigma}^2} = \frac{r(30)^2}{2(150)} = 3r$, select the smallest value of r such that the power satisfies: $\gamma(L) \geq \gamma_0 = 0.90$

From the SAS output the required number of chickens per treatment would be r=7.

Using the graphs in the textbook on page 607 ($\nu_1 = t - 1 = 4$), with $\nu_2 = 5(r-1)$ and $\lambda = 3r$, we have

If r=6, $\phi = \sqrt{\lambda/t} = \sqrt{(3)(6)/5)} \approx 1.9$. From the curve with $\alpha = .05$ and $\nu_2 = 5(6-1) = 25$

we estimate the power to be about $.89 < .9$

If r=7, $\phi = \sqrt{\lambda/t} = \sqrt{(3)(7)/5)} \approx 2.05$. From the curve with $\alpha = .05$ and $\nu_2 = 5(7-1) = 30$

we estimate the power to be about $.94 > .9$

Therefore, take r=7 and the minimum number of chickens needed per treatment is 7.

P5. (20 points) For an experiment with four treatments, $t = 4$, and reps/treatment given by $n_1 = 3, n_2 = 4, n_3 = 5, n_4 = 3$ we have the following results.

a. For the Cell Means model: $Y_{ij} = \mu_i + e_{ij},\ i = 1, 2, 3, 4,\ j = 1, \cdots, n_i$, the Design Matrix is

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}_{15 \times 4}$$

The treatment means are μ_i

b. For the Effect models with no constraints: $Y_{ij} = \mu + \tau_i + e_{ij},\ i = 1, 2, 3, 4,\ j = 1, \cdots, n_i.$

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}_{15 \times 5}$$

The treatment means are $\mu_i = \mu + \tau_i$ but the τ_i are not estimable because $\mathbf{X}'\mathbf{X}$ is not of full rank and hence the LSE of the model parameters, $\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'Y$ are not computable because $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ does not exist. A generalized inverse could be used but then the model estimates are not unique.

c. For the Effect models with constraint $\tau_4 = 0 :\ Y_{ij} = \mu + \tau_i + e_{ij},\ i = 1, 2, 3, 4,\ j = 1, \cdots, n_i$ reduces to $Y_{ij} = \mu + \tau_i + e_{ij},\ i = 1, 2, 3,\ j = 1, \cdots, n_i.$ and $Y_{4j} = \mu + e_{ij},\ j = 1, \cdots, n_4.$

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}_{15 \times 4}$$

The treatment means are $\mu_4 = \mu$ and $\mu_i = \mu_4 + \tau_i$ and the τ_i are now estimable because $\mathbf{X}'\mathbf{X}$ is of full rank and hence the LSE of the model parameters, $\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'Y$ exist.

d. Without constraints:

$$\tau_i = \mu_i - \mu,\ i = 1, 2, 3, 4 \quad \text{with } \mu = \frac{1}{n}\sum_{i=1}^{t} n_i\mu_i, \text{ however, the } \tau_i \text{ are not estimable}$$

e. With the constraint $\tau_4 = 0,\ \ \mu = \mu_4$ and

$$\tau_i = \mu_i - \mu = \mu_i - \mu_4,\ i = 1, 2, 3, 4 \quad \Rightarrow \quad \tau_1 = \mu_1 - \mu_4;\ \ \tau_2 = \mu_2 - \mu_4;\ \ \tau_3 = \mu_3 - \mu_4;\ \ \tau_4 = 0;$$

P6. (10 points)

a. We have relatively homogeneous experimental units which are randomly assigned to the treatments in order to avoid any subjective assignment of treatments.

A probability sample of units should be selected from available members of each treatment population. Units are selected from within each population such that each unit has an equal chance of entering sample. Note that each population represents a separate treatment classification, and random sampling is maintained only within the population.

b. The e_{ij}s are uncorrelated with mean 0 and the same variance then the least squares estimators are Best Linear Unbiased Estimators (BLUE) of the population parameters. If we further include the condition that the e_{ij}s are independent, normally distributed then the least squares estimators are Uniformly Minimum Variance Unbiased Estimators (UMVUE). Note the difference in these two statements. BLUE's have smallest variance amongst all unbiased estimators which are linear functions of the data and UMVUE's have smallest variance amongst all unbiased estimators no matter the form of the estimator.

c. y_{ij} have independent $N(\mu_i, \sigma^2)$ distributions for all i and j.

d. Permutation and rank based procedures would be valid provided we require the condition "y_{ij}s are independently distributed with the same variance". Rank and permutation procedures would not require the normality condition.

We will discuss alternative procedures in Handout 5 for when the conditions of normality and constant variance do not hold.

P7. (15 points)

a. The model can be written as $y_j = \beta_o + \beta_1 X_{\mathbf{Bj}} + \beta_2 X_{\mathbf{Cj}} + \beta_3 X_{\mathbf{Dj}} + e_j$ where

$X_{\mathbf{Bj}}$ equals 1 if Treatment B is used and 0 otherwise

$X_{\mathbf{Cj}}$ equals 1 if Treatment C is used and 0 otherwise

$X_{\mathbf{Dj}}$ equals 1 if Treatment D is used and 0 otherwise

Let $\mu_{\mathbf{A}}$, $\mu_{\mathbf{B}}$, $\mu_{\mathbf{C}}$, $\mu_{\mathbf{D}}$ be the mean responses of the four treatment groups.

The model coefficients can be interpreted as follows:

$$\mu_{\mathbf{A}} = E[y_j] = E[\beta_o + \beta_1 X_{\mathbf{Bj}} + \beta_2 X_{\mathbf{Cj}} + \beta_3 X_{\mathbf{Dj}} + e_j] = \beta_o + \beta_1(0) + \beta_2(0) + \beta_3(0) + 0 = \beta_o$$
$$\mu_{\mathbf{B}} = E[y_j] = E[\beta_o + \beta_1 X_{\mathbf{Bj}} + \beta_2 X_{\mathbf{Cj}} + \beta_3 X_{\mathbf{Dj}} + e_j] = \beta_o + \beta_1(1) + \beta_2(0) + \beta_3(0) + 0 = \beta_o + \beta_1$$
$$\mu_{\mathbf{C}} = E[y_j] = E[\beta_o + \beta_1 X_{\mathbf{Bj}} + \beta_2 X_{\mathbf{Cj}} + \beta_3 X_{\mathbf{Dj}} + e_j] = \beta_o + \beta_1(0) + \beta_2(1) + \beta_3(0) + 0 = \beta_o + \beta_2$$
$$\mu_{\mathbf{D}} = E[y_j] = E[\beta_o + \beta_1 X_{\mathbf{Bj}} + \beta_2 X_{\mathbf{Cj}} + \beta_3 X_{\mathbf{Dj}} + e_j] = \beta_o + \beta_1(0) + \beta_2(0) + \beta_3(1) + 0 = \beta_o + \beta_3$$

Thus, we have

- β_o is $\mu_{\mathbf{A}}$
- β_1 is $\mu_{\mathbf{B}} - \mu_{\mathbf{A}}$
- β_2 is $\mu_{\mathbf{C}} - \mu_{\mathbf{A}}$
- β_3 is $\mu_{\mathbf{D}} - \mu_{\mathbf{A}}$

b. From the model, the mean response in treatment group B is $\mu_{\mathbf{B}} = \beta_o + \beta_1$.

Note that $\beta_o + \beta_1$ can be written as $\mathbf{v}'\beta$ where $\mathbf{v}' = (1, 1, 0, 0)$ and $\beta' = (\beta_o,\ \beta_1,\ \beta_2,\ \beta_3)$.

Thus, the estimator of $\mu_{\mathbf{B}} = \beta_o + \beta_1$ is $\hat{\mu}_{\mathbf{B}} = \hat{\beta}_o + \hat{\beta}_1 = 37.5 - 11.5 = 26$.

The estimated standard error of the estimate $\hat{\mu}_{\mathbf{B}}$ is the estimated standard error of $\hat{\beta}_o + \hat{\beta}_1 = \mathbf{v}'\beta$ where $\mathbf{v}' = (1, 1, 0, 0)$

$$\widehat{SE}(\hat{\mu}_{\mathbf{B}}) = \sqrt{\widehat{Var}(\mathbf{v}'\beta)} = \hat{\sigma_e}\sqrt{\mathbf{v}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{v}} = (19.45)\sqrt{.02} = 2.75.$$

The 95% C.I. is as follows: $df_{MSE} = 4(50) - 4 = 196$ and $t_{\alpha/2,df_E} = t_{.025,196} = qt(.975, 196) = 1.972$

$$(\text{point estimator}) \pm (t_{\alpha/2,df_{MSE}})[\widehat{SE}(\text{point estimator})] = 26 \pm (1.972)(2.75) = 26 \pm 5.42 = (20.58,\ 31.42)$$

c. From the model, the differences in the mean responses $\mu_B - \mu_A = \beta_1$.

Therefore, a 95% C.I. on $\mu_{\mathbf{B}} - \mu_{\mathbf{A}}$ is a 95% C.I. on β_1 which is given by

$$\hat{\beta}_1 \pm (t_{\alpha/2,df_E})[\widehat{SE}(\hat{\beta}_1)] = -11.5 \pm 1.972 \times 3.89 = (-19.17,\ -3.83)$$

where β_1 can be written as $\mathbf{v}'\beta$ where $\mathbf{v}' = (0, 1, 0, 0)$.

The estimated standard error of the estimate $\hat{\beta}_1$ is

$$\hat{\sigma_e}\sqrt{\mathbf{v}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{v}} = (19.45)\sqrt{.04} = 3.89.$$