Some Chapter 2 Notes on Completely Randomized Designs

Want to compare the weight gains of hogs on 3 diets (A,B,C)

12 hogs available for study

Can use table of random digits on p. 624 of Kuehl. Arbitrarily start at Line 6, Column 6.

ID Number	Random Number
1	29
2	69
3	24
4	14
5	90
6	78
7	59
8	62
9	39
10	67
11	53
12	73

RANDOMIZE $\longrightarrow \longrightarrow \longrightarrow \longrightarrow$

ID Number	Diet	Weight Gain	Response Label
4	A	28	y_{11}
3	A	33	y_{12}
1	A	30	y_{13}
9	A	37	y_{14}
11	B	23	y_{21}
7	B	28	y_{22}
8	B	25	y_{23}
10	B	24	y_{24}
2	C	31	y_{31}
12	C	28	y_{32}
6	C	33	y_{33}
5	C	28	y_{34}

Assumptions All observations are independent and

$$\begin{aligned} y_{11}, y_{12}, y_{13}, y_{14} &\sim N(\mu_1, \sigma^2) \\ y_{21}, y_{22}, y_{23}, y_{24} &\sim N(\mu_2, \sigma^2) \\ y_{31}, y_{32}, y_{33}, y_{34} &\sim N(\mu_3, \sigma^2) \end{aligned}$$

$$\begin{aligned} y_{ij} &= \mu_i + e_{ij} \\ i &= 1, \ldots, t \ (t = 3 \text{ in this case}) \\ j &= 1, \ldots, r_i \ (r_1 = r_2 = r_3 = 4 \text{ in this case}) \\ e_{ij} &\sim \text{ independent } N(0, \sigma^2) \end{aligned}$$

e_{ij} are experimental errors. *If there was no experimental error, the weight gains of all hogs given the same treatment would be exactly the same.*

If $\mu_1 = 33$, what is e_{13}? How about e_{14}?

We want to test

$$H_0 : \mu_1 = \mu_2 = \mu_3 \text{ versus } H_A : \mu_i \neq \mu_k \text{ for some } i, k.$$

If H_0 is true, there exists μ such that

$$y_{ij} = \mu + e_{ij} \quad i = 1, \ldots, t \quad j = 1, \ldots, r_i.$$

This is known as the reduced model.

The full model is

$$y_{ij} + \mu_i + e_{ij} \quad i = 1, \ldots, t \quad j = 1, \ldots, r_i.$$

Under the reduced model, all diets are equivalent. Diets don't affect weight gain. Under the full model, diets matter. Which model seems to fit data the best?

For the Full Model: $(y_{ij} = \mu_i + e_{ij} \;\; i = 1, \ldots, t \;\; j = 1, \ldots, r_i)$

We want estimates of the model parameters μ_1, μ_2, μ_3, and σ^2.

Choose estimates of μ_1, μ_2, and μ_3 that will make *residuals* as close to zero as possible.

experimental error $e_{ij} = y_{ij} - \mu_i$ residual $\hat{e}_{ij} = y_{ij} - \hat{\mu}_i$

Choose $\hat{\mu}_i$ so that $SSE = \sum_{i=1}^{t} \sum_{j=1}^{r_i} \hat{e}_{ij}^2 = \sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \hat{\mu}_i)^2$ is as small as possible.
Such estimates are called *least squares estimates*.

Can use calculus or some slightly tricky algebra to show that
the least squares estimate of μ_i is $\hat{\mu}_i = \bar{y}_{i\cdot} = \sum_{j=1}^{r_i} y_{ij}/r_i$.

What is the least squares estimates of μ_2 in the weight gain experiment?

Note that

$$SSE = \sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \hat{\mu}_i)^2 = \sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i\cdot})^2 = \sum_{i=1}^{t} (r_i - 1) s_i^2 \text{ where } s_i^2 = \frac{1}{r_i - 1} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i\cdot})^2$$

s_i^2 is the sample variance in the ith group.

The error variance σ^2 is estimated by

$$MSE = \frac{SSE}{N - t} = \frac{(r_1 - 1)s_1^2 + \cdots + (r_t - 1)s_t^2}{(r_1 - 1) + \cdots + (r_t - 1)} = \text{weighted average of within-group sample variances.}$$

$(N = \sum_{i=1}^{t} r_i$ =total sample size)

For the Reduced Model: $(y_{ij} = \mu + e_{ij} \;\; i = 1, \ldots, t \;\; j = 1, \ldots, r_i)$

We want to estimate μ. $\hat{e}_{ij} = y_{ij} - \hat{\mu}$ $SSE_R = \sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \hat{\mu})^2$

Least squares estimate of μ is $\hat{\mu} = \bar{y}_{\cdot\cdot} = \sum_{i=1}^{t} \sum_{j=1}^{r_i} y_{ij}/N$.

$SSE_R = \sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \hat{\mu})^2 = \sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{\cdot\cdot})^2 = SS_{\text{Total}}$

$SSE_R \geq SSE_F$ $SSE_R - SSE_F = SST = SS_{\text{Treatment}}$ =sum of squares treatment
(SST is reduction in sum of squares error due to including treatment in the model.)

$$SS_{\text{Total}} = SS_{\text{Treatment}} + SS_{\text{Error}}$$

$$\sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{t} r_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + \sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i\cdot})^2$$

Source	DF	Sum of Squares	Mean Square	F_o
Treatment	$t - 1$	$\sum_{i=1}^{t} r_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	$SST/(t - 1)$	MST/MSE
Error	$N - t$	$\sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i\cdot})^2$	$SSE/(N - t)$	
Total	$N - 1$	$\sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{\cdot\cdot})^2$		

$F_o \sim F$ with $t - 1$ and $N - t$ degrees of freedom when $H_0 : \mu_1 = \cdots = \mu_t$ is true.