CSUEB – STAT 6305 – Winter 2017 - Prof Yan Zhou

Homework 7 - Henry Lankin, Gui Larangeira

March 2, 2017

HW 7: 17.31, 17.32

17.31

17.31 Tablet hardness is one comparative measure for different formulations of the same drug product; some combinations of ingredients (in addition to the active drug) in a formulation give rise to harder tablets than do other combinations. Suppose that three batches of a formulation are randomly selected for examination. Three different 1-kg samples of tablets are randomly selected from each batch and seven tablets are randomly selected for testing from each of the 1-kg samples. The hardness readings are given here.

a) Identify the design.

This experiment uses a nested design where the major factor (batch) is a random effect with 3 treatment levels and the minor factor (sample) is a random effect with 9 total treatment levels, 3 for each level of the site factor. There are 7 replications within each level of the batch treatment.

b) Give an appropriate model with assumptions.

Nested factorial design model:

$$\boldsymbol{y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{ijk}}$$

$$i = 1, 2, 3;\ j = 1, 2, 3, ;\ k = 1, 2, \ldots, 7$$

$y_{ij}k$ – the hardness reading of the k^{th} tablet in the j^{th} sample from the i^{th} batch representing the observed response variable: 63 observations from 7 replications of the 9 batch-sample nested treatment combinations.

τ_i – the random effect due to the i^{th} batch: 3 treatment levels chosen from a continuous distribution of batches

$\beta_{j(i)}$ – the random effect due to the j^{th} sample in the i^{th} batch: 3 treatment levels chosen from a continuous distribution of samples within each of the 3 batch levels

ε_{ijk} – random error associated with each batch-sample nested combination: 63 residual errors

Assumptions:

1. The random effect due to the i^{th} batch, τ_i, is assumed to follow a normal distribution with mean 0 and variance ${\sigma_\tau}^2$.
2. The τ_i's are independent of each other.
3. The random effect due to the j^{th} sample in the i^{th} batch, $\beta_{j(i)}$, is assumed to follow a normal distribution with mean 0 and variance ${\sigma_{\beta(\tau)}}^2$.
4. The $\beta_{j(i)}$'s are independent of each other.
5. The residual effect, ε_{ijk}, is assumed to follow a normal distribution with mean 0 and common variance of ${\sigma_\varepsilon}^2$.
6. The ε_{ijk}'s are independent of each other.

c) Give the sources of variability and degrees of freedom for an AOV.

Source	DF	SS	Mean Square	EMS	Estimated variability	F Value	Pr > F
B: Batch	2	9095.523810	4547.761905	${\sigma_\varepsilon}^2 + 7{\sigma_{\beta(\tau)}}^2 + 21{\sigma_\tau}^2$	${\sigma_\tau}^2 = 214.4293$	101.63	$< .0001$
S(B): Sample(Batch)	6	268.476190	44.746032	${\sigma_\varepsilon}^2 + 7{\sigma_{\beta(\tau)}}^2$	${\sigma_{\beta(\tau)}}^2 = 2.222978$	1.53	0.1851
Error	54	1576.000000	29.185185	${\sigma_\varepsilon}^2$	${\sigma_\varepsilon}^2 = 29.185185$		

d) Perform an analysis of variance and draw conclusions about the tablet hardness data for the formulation under study. Use $\alpha = 0.05$.

- Test for significance of the major factor, the batch the tablet is from, on tablet hardness:

Hypothesis:

$$H_0: {\sigma_\tau}^2 = 0$$
$$H_a: {\sigma_\tau}^2 > 0$$

Test statistic:

$$F = \frac{MSB}{MSS(B)} = \frac{SSB/(b-1)}{SSS/b(s-1)} = \frac{9095.52/2}{268.48/6} = 101.63$$

Rejection region:

At $\alpha = 0.05$, we reject the null hypothesis for $F > F_{2,6}(0.05) = 5.14$.

Conclusion:

Since $F = 101.63 > 5.14$ and this test gives a p-vlaue < 0.0001, we reject the null hypothesis and conclude that the batch factor significantly affects the mean hardness reading of the tablet.

- Test for significance of the minor factor, the sample the tablet is from, on tablet hardness:

Hypothesis:

$$H_0: {\sigma_{\beta(\tau)}}^2 = 0$$

$$H_a: {\sigma_{\beta(\tau)}}^2 > 0$$

Test statistic:

$$F = \frac{MSS(B)}{MSE} = \frac{SSS/b(s-1)}{SSE/bs(n-1)} = \frac{268.48/6}{1576.00/54} = 1.53$$

Rejection region:

At $\alpha = 0.05$, we reject the null hypothesis for $F > F_{6,54}(0.05) = 2.27$.

Conclusion:

Since $F = 1.53 < 2.27$ and this test gives a p-vlaue $= 0.1851$, we fail to reject the null hypothesis and conclude that the sample in each batch level does not significantly affect the mean hardness reading of the tablet.

17.32

17.32 An anthropologist is interested in the impact of the usage of mind-altering drugs in religious ceremonies. She selects five underdeveloped countries for inclusion in her study. She then selects 10 tribes in each country. Finally she randomly selects 20 families from each tribe for an in-depth interview. After the interview, the anthropologist assigns a score which reflects the impact of the usage of mind-altering drugs in religious ceremonies. The researcher is interested in determining if there is a difference in the average scores across countries and the degree of variability in the index across tribes and families. In this study, there are three factors of interest to the researcher: country, tribe, and family.

a) Identify each of the factors as fixed or random; justify your answer.

- Country is a random effect because there are many possible underdeveloped countries that could have been chosen.
- Tribe is a random effect because there are many possible tribes that could have been chosen in each country.
- Family is a random effect because 20 families are randomly chosen from each tribe.

b) State whether the factors are nested or crossed; provide reasons for your answers.

- Country is the major factor: C
- Tribe is a minor factor nested in the country factor: $T(C)$
- Family is a minor factor nested in the tribe factor: $F(T, C)$

c) Provide an AOV table that includes source of variation, degrees of freedom, and expected mean squares.

- Model:

$$\boldsymbol{y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(ij)} + \varepsilon_{ijk}}$$

$$i = 1, 2, 3, 4, 5;\ j = 1, 2, .., 10, ;\ k = 1, 2, \ldots, 20$$

for a total of $i\ x\ j\ x\ k = 5(10)20 = 1000$ observations and where μ is the overall mean, τ_i the country effect, $\beta_{j(i)}$ the tribe nested in country effect and $\gamma_{k(ij)}$ the family nested in tribe, country effect.

Source	DF	SS	Mean Square	EMS
C: Country	4	SSC	MSC	$\sigma_\varepsilon{}^2 + \sigma^2_{F(T,C)} + 20\sigma^2_{T(C)} + 200\sigma_C{}^2$
T(C): Tribe(Country)	45	$SST(C)$	$MST(C)$	$\sigma_\varepsilon{}^2 + \sigma^2_{F(T,C)} + 20\sigma^2_{T(C)}$
F(T, C): Family(Tribe, Country)	950	$SSF(T,C)$	$MSF(T,C)$	$\sigma_\varepsilon{}^2 + \sigma^2_{F(T,C)}$
Error	0	SSE	MSE	$\sigma_\varepsilon{}^2$

The design suggests there is no degrees of freedom for the residual errors, and, thus, the family factor cannot be tested properly. Thus, by looking at family as a factor of interest, the design is flawed without any replications for each family. This would be remedied by using the family observations as replications within each tribe or by having at least two observations per family.

SAS code:

```
* input data with do loop;
data tablethardness_looped;
        do batch = 1 to 3;
                do sample 1 to 3;
                        do rep 1 to 7;
                                input response@@;
                                output;
                        end;
                end;
        end;
        cards;
'Response'
85
94
91
98
85
96
93
76
87
90
91
88
94
96
95
98
94
96
99
100
93
108
100
105
109
104
102
108
117
```

```
106
103
109
100
104
102
101
108
100
99
117
109
105
71
85
78
68
85
67
76
81
70
84
83
72
81
78
72
68
80
72
75
79
74
;
run;

proc print data=tablethardness_looped;
run;


* input data with stacked data;
data tablethardness;
        input Sample$ Batch$ Response;
        cards;
1       1       85
1       1       94
1       1       91
1       1       98
1       1       85
1       1       96
1       1       93
2       1       76
2       1       87
2       1       90
2       1       91
2       1       88
2       1       94
2       1       96
3       1       95
3       1       98
3       1       94
3       1       96
3       1       99
3       1       100
3       1       93
1       2       108
1       2       100
1       2       105
1       2       109
1       2       104
1       2       102
1       2       108
2       2       117
2       2       106
2       2       103
```

```
2       2       109
2       2       100
2       2       104
2       2       102
3       2       101
3       2       108
3       2       100
3       2       99
3       2       117
3       2       109
3       2       105
1       3       71
1       3       85
1       3       78
1       3       68
1       3       85
1       3       67
1       3       76
2       3       81
2       3       70
2       3       84
2       3       83
2       3       72
2       3       81
2       3       78
3       3       72
3       3       68
3       3       80
3       3       72
3       3       75
3       3       79
3       3       74
;
run;

proc print data=tablethardness;
run;

proc glm data=tablethardness;
        class batch sample;
        model response = batch sample(batch);
        random batch sample(batch) / test;
run;
quit;
```