Homework 3

Cody Frisby

January 27, 2016

1a ANOVA table:

The Group means are as follows:

$\hat{\mu}_{15} = 9.8$, $\hat{\mu}_{20} = 15.4$, $\hat{\mu}_{25} = 17.6$, $\hat{\mu}_{30} = 21.6$, $\hat{\mu}_{35} = 10.8$

And the overall mean is $15.04 = \bar{y}_{..}$ $a = 5$ $n = 5$ $N = 25$

The sum of squares for the treatments is computed as follows:

$$SS_{treatments} \;=\; n\sum(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$475.76 = 5 \times [(9.8 - 15.04)^2 + (15.4 - 15.04)^2 + ... + (10.8 - 15.04)^2]$$

The overall, total, sum of squares is as follows:

$$SS_{total} \;=\; \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

$$636.96 = (7 - 15.04)^2 + (7 - 15.04)^2 + (15 - 15.04)^2 + (11 - 15.04)^2 + ... + (11 - 15.04)^2$$

The sum of squares for within, or errors, or residuals, is computed as follows:

$$SS_{errors} \;=\; SS_{total} - SS_{treatments}$$

$$161.2 = 636.96 - 475.76$$

The degrees of freedom for treatments is $a - 1 = 4$. The degrees of freedom for errors is $N - a = 25 - 5 = 20$. And the total degrees of freedom is $N - 1 = 24$.

The mean square for the treatments is computed by taking $SS_{treatments}$ and dividing it by the degrees of freedom.

$$MS_{treatments} = \frac{SS_{treatments}}{a-1}$$

.

$$118.94 = \frac{475.76}{4}$$

MSE, or mean square errors, is computed much like mean square treatments above, only we divide SS_{errors} by the degrees of freedom $N - a$.

$$MS_{errors} = \frac{SS_{errors}}{N-a}$$

$$8.06 = \frac{161.2}{20}$$

The F statistic for our ANOVA table is computed as follows:

$$F_0 = \frac{MS_{treatments}}{MS_{errors}}$$

$$14.7568238 = \frac{118.94}{8.06}$$

Analysis of Variance Table:

Sum of Squares	Degrees of Freedom	Mean Square	F_0
475.76	4	118.94	14.757
161.2	20	8.06	
636.96	24		

We conclude that there is a significant difference between treatment groups. The p-value $= 9.1279371 \times 10^{-6}$

1d Model adequacy

The residual plots below show nothing we need to be concerned about. Here we specifically want to look at the residuals (or errors) from our model. We are specifically concerned about normality and equality of variances. The qqnorm/qqline plot shows no extreme departures from normality and the residual plot below shows no concerns about the equality of variances.
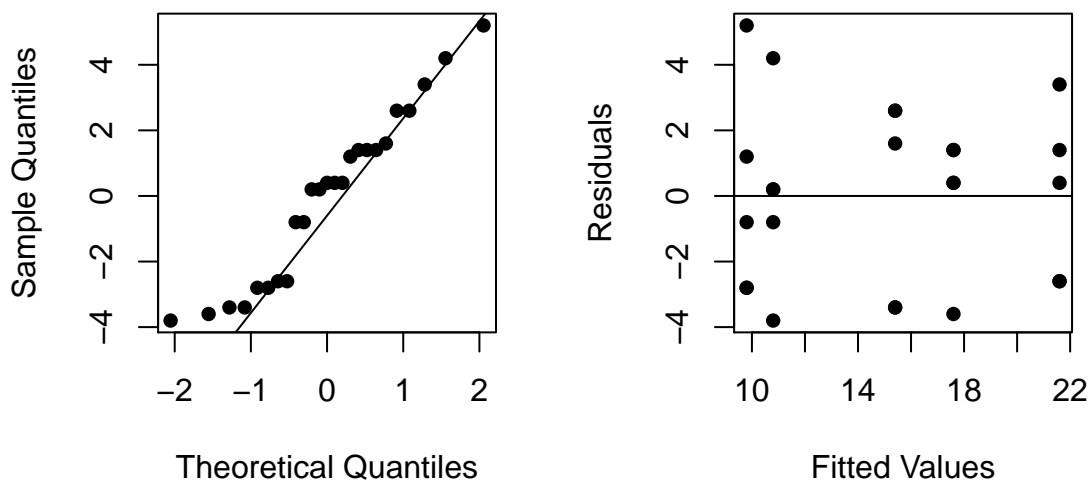
2)

A)

The total sum of squares was computed to be 280 while the treatment sum of squares was compute to be 120. This means the error (or within) sum of squares is $280 - 120 = 160$. In order to compute an ANOVA table we need to determine the group size and the size within groups. $N = 20$, $a = 5$, and $n = 5$.

Analysis of Variance Table:

Sum of Squares	Degrees of Freedom	Mean Square	F_0
120	3	53.3333333	5.3333333
160	16	10	
280	19		

B)

A statistical model for our problem could be described symbolically as

$$y_{ij} = \mu_i + \epsilon_{ij} \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., n \end{cases}$$

where y_{ij} is the ijth observation, μ_i is the mean of the ith factor level or treatment group, and ϵ_{ij} is a random error component that incorporates all other sources of variability in the experiment.

A model for our data could be expressed as

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., n \end{cases}$$

where $\mu_i = \mu + \tau_i$, $i = 1, 2, ..., a$. Here, μ is the overall mean, so for our purposes it would be the mean for all 20 of our soil treatment observations. μ_i is the group treatment means for $i = 1, 2, 3, 4$. τ_i is the effect of the soil treatment to the topsoil from the ith treatment group.

C)

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

$$H_1 : \tau_i \neq 0 \quad for\ at\ least\ one\ i$$

We test this hypothesis with the soil data by considering the value for F_0 in the above ANOVA table. If this value can be considered improbable under a given value for α then we say there is significant evidence that there is at least one treatment group that differs from the others. Our F_0 value is 5.3333333. For our data the p-value associated with this F statistic is 0.0060006. So at any practical value for α we would conclude there is at least one treatment group that has an effect on the soil moisture content.

D) Set up an orthogonal contrast

Without knowing more about the research question and the anticipated differences and similarities the researcher expects it is difficult to set up an orthogonal contrast. These are normally most useful with **preplanned comparisons**. Having said that, here's a try at one:

Contrasts	Treatments			
-	1	2	3	4
A	1/3	1/3	-1/3	-1/3
B	1	0	0	-1
C	1	-1	0	0

The coefficients must sum to zero. So, for contrast A $\frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4$ tests the hypothesis that level 1 and 2 are similar but different than levels 3 & 4, and $\frac{1}{3} + \frac{1}{3} - \frac{1}{3} - \frac{1}{3} = 0$ shows that the coefficients sum to zero.

E) Set up a 95% CI for type 2 sub-grade soil

To set up a confidence interval for type 2 group for the soil grade experiment we need to compute our t statistic. With $\alpha = 0.05$ $with\ df = 16$ our t statistic is 2.1199053. Also we need to compute $\sqrt{\frac{MS_E}{n}}$. This is equal to 1.4142136. So, our 95% confidence interval for the true mean of group2 is

$$\bar{y}_{2.} - 2.9979988 \leq \mu_2 \leq \bar{y}_{2.} + 2.9979988$$

3) Determine the effects of several teaching methods

A)

What is known? N = 30, a = 5, n = 6. Scores for each group: 120, 600, 720, 240, 420 for groups 1, 2, 3, 4, and 5 respectively. $SS_{treatments} = 340$ and $SS_{total} = 465$. We are asked to complete the ANOVA table:

Sum of Squares	Degrees of Freedom	Mean Square	F_0
340	4	85	17
125	25	5	
465	29		

The p-value associated with the F_0 statistic is equal to 3.9179604×10^{-7}.

B)

A model for our data could be expressed as

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., n \end{cases}$$

where $\mu_i = \mu + \tau_i$, $i = 1, 2, ..., a$. Here, μ is the overall mean, so for our purposes it would be the mean for all 30 students and 5 different teaching groups. μ_i is the group treatment means for $i = 1, 2, 3, 4, 5$. τ_i is the effect the teaching method has on the test scores from the ith treatment group. The hypothesis we are testing is whether or not the teaching style has an effect on test scores. It can be represented as

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

$$H_1 : \tau_i \neq 0 \quad for\ at\ least\ one\ i$$

where we assume the errors from our data to take the form

$$y_{ij} \sim N(\mu + \tau_i, \sigma^2)$$

and that the observations are mutually independent.

C)

To test the hypothesis we look at the ANOVA table above in addition to the p-value from the F statistic. All this tells us is whether or not there is at least one group that is different, or whose effect is not equal to zero. It does not tell us which group is better, worse, or statistically separate from the others. Based on the ANOVA table calculations and our p-value, we reject the null hypothesis, that the teaching method has no effect on the student's achievement scores, with a p-value = 3.9179604×10^{-7}. We conclude that teaching method has a statistically significant impact on the student's achievement scores.

D)

One set of orthogonal contrasts that would make sense simply looking at the treatment groups (not having any other knowledge about the methods) would be constructed as follows

Treatments					
-	Control	TextA w/teach	TextA w/comp	TextB w/teach	TextB w/comp
Contrasts	-4	1	1	1	1

$$H_0 : \mu_1 = \mu_2 + \mu_3 + \mu_4 + \mu_5$$

$$Contrast = -\bar{y}_{1.} + \bar{y}_{2.} + \bar{y}_{3.} + \bar{y}_{4.} + \bar{y}_{5.}$$

The set up of this contrast compares the average effect of all our factors with the control group. This might be of interest at the onset to the researcher to see if there is an effect from the Textbooks A or B and the teacher or computer when compared with the control textbook. If there is we could drill down further for more comparisons between A and B or computer and teacher.

E)

We are asked to test whether the mean score between groups 1 and 5 differ significantly at $\alpha = 0.05$. $H_0 : \mu_1 = \mu_5$, $H_a : \mu_1 \neq \mu_5$. The contrasts for our test will be $c_1 = -1$, $c_5 = 1$, $c_2 = c_3 = c_4 = 0$

$$t_0 = \frac{\sum_{i=1}^{a} c_i \bar{y}_{i.}}{\sqrt{\frac{MS_E}{n} \sum_{i=1}^{a} c_i^2}}$$

$$300 = \frac{300}{1}$$

with $t_{\alpha/2, N-a} = 2.0595386$. So, we would reject the null hypothesis at $\alpha = 0.05$. We conclude that there is a significant difference between group 1 and group 5 achievement scores.

F)

The standard error for the contrasts $\mu_1 + \mu_2 + 2\mu_3 + \mu_4 - 5\mu_5$ is computed with the following formula

$$\sqrt{\frac{MS_E}{n} \sum_{i=1}^{a} c_i^2}$$

where $MS_E = 5$ and $\sum_{i=1}^{a} c_i^2$ is equal to $(-4)^2 + 1^1 + 1^2 + 1^2 + 1^2 = 20$. So,

$$4.472136 = \sqrt{\frac{5}{5} \times 20}$$