**RSCH 6120/8120: HW 2**

**Kristen Zhang | Sept 25**

This HW uses the eclsk.csv data. You will conduct an ANOVA and use multiple comparison procedures to examine specific group differences. You should interpret any statistical tests and interpret inferences from those tests. Support your inferences by noting evidence in the results.

**1. Use an ANOVA to determine if there are significant differences in student general intelligence scores (eclsk$gen) across parent education categories (eclsk$parent.ed.cat).**

**a) Provide results.**

> anova (lm(gen~parent.ed.cat, data=eclsk))

Analysis of Variance Table

Response: gen

Df Sum Sq Mean Sq F value Pr(>F)

parent.ed.cat 2 19232 9616.2 189.64 < 2.2e-16 \*\*\*

Residuals 2432 123323 50.7

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**b) Interpret the results.**

From the results, we can tell that F value is 189.64, p-value < 0.001, which tells that there are significant differences in mean student general intelligence scores across different groups of parent education categories. We should reject the H0.

**c) List all the possible pairwise comparisons (What are all the possible group combinations where the omnibus ANOVA could identify differences?)**

There are three groups of parent education categories, A (HS or less), B (College Experiences), and C (Graduate Schooling). So there are three comparisons: A vs. B, A vs. C, B vs. C.

**2. Select a multiple comparison procedure to determine which if any pairwise comparisons from Question 1 are significant.**

**a) Provide code and results.**

> pairwise.t.test(eclsk$gen, eclsk$parent.ed.cat, p.adj = "bonf")

Pairwise comparisons using t tests with pooled SD

data: eclsk$gen and eclsk$parent.ed.cat

A - HS or less B - College Experience

B - College Experience < 2e-16 -

C - Graduate Schooling < 2e-16 2.6e-14

P value adjustment method: bonferroni

**b) Interpret the results.**

From the results above, we can tell that there’s a difference between all three comparisons, as all of the p-values are below 0.5.

**3. Select another multiple comparison procedure and again determine which groups have significant differences.**

**a) Provide code and results.**

> pairwise.t.test(eclsk$gen, eclsk$parent.ed.cat, p.adj = "holm")

Pairwise comparisons using t tests with pooled SD

data: eclsk$gen and eclsk$parent.ed.cat

A - HS or less B - College Experience

B - College Experience < 2e-16 -

C - Graduate Schooling < 2e-16 8.6e-15

P value adjustment method: holm

**b) Interpret the results.**

All of the results are below 0.5, therefore we reject the null hypothesis and we believe that there’s a significant difference in the three combinations of comparisons.

**c) Compare results with Question 2 results.**

Question 2 showed the results of the post-hoc comparison, and Question 3 showed the results of planned comparison, which led to difference comparison results between B/C (2.6e-14 vs. 8.6e-15). I think the planned comparison is more accurate because they showed the results based on prior hypotheses.

**4. Use an ANCOVA to determine if there are significant differences in student general intelligence scores (eclsk$gen) across parent education categories (eclsk$parent.ed.cat) while controlling for parental income (eclsk$income).**

**a) Provide results.**

> ancova.output <- anova(lm(gen~income+parent.ed.cat, data=eclsk))

> ancova.output

Analysis of Variance Table

Response: gen

Df Sum Sq Mean Sq F value Pr(>F)

income 1 15781 15780.6 325.853 < 2.2e-16 \*\*\*

parent.ed.cat 2 9044 4522.1 93.376 < 2.2e-16 \*\*\*

Residuals 2431 117730 48.4

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**b) Interpret the results.**

The results showed that, even though we controlled the parent income, there are still significant differences in mean student general intelligence scores across different groups of parent education categories. Therefore, we can reject the null hypothesis.

**5. Conduct a factorial ANOVA with the eclsk data. The outcome of interest is reading scores (eclsk$read) and the two factors of interest are parent education category (eclsk$parent.ed.cat) and student sex (eclsk$female).**

**a) Provide code and results**

> factorial.anova.output <- anova(lm(read~parent.ed.cat+female+parent.ed.cat:female, data=eclsk))

> factorial.anova.output

Analysis of Variance Table

Response: read

Df Sum Sq Mean Sq F value Pr(>F)

parent.ed.cat 2 41810 20904.9 111.7201 < 2.2e-16 \*\*\*

female 1 4347 4347.2 23.2326 1.524e-06 \*\*\*

parent.ed.cat:female 2 17 8.5 0.0455 0.9555

Residuals 2429 454511 187.1

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

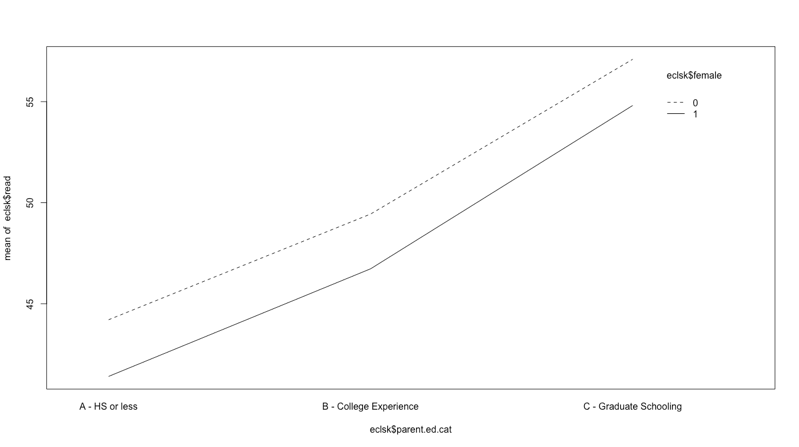
**b) Interpret the results**

From the results above, we can tell that the reading scores are significantly different across different groups of parent education (F= 111.72, *p* < .001) and genders (F = 23.23, *p* < .001). However, there’s no significant interaction effect (*p* >.05). This means that gender will not influence parent education’s impact on reading scores.

**c) Create a mean plot to illustrate the interaction or lack of interaction**

> interaction.plot(x.factor = eclsk$parent.ed.cat, trace.factor = eclsk$female,

+ response = eclsk$read, fun = mean)

As is shown in the plot, as parent education increases, the reading scores increase. Therefore, we can say that the main effect of parent education is significant; However, there’s no significant interaction effect between gender and parent education, because the lines of two different genders are almost parallel.