**Type your answers in the appropriate fields; please make answer fields larger as needed. Submit this assignment on Canvas by 11:59 PM on Tuesday, Nov 28th 2023. This assignment will be graded by Simran.**

Please use the Age Religion Health data set from lab.

**1. ANCOVA: Controlling for Continuous Variable**

1. Perform a one-way ANOVA to assess whether there are differences in overall health (Overallhealth\_04) based on education status (the school.group variable we created in lab). Fill in the results of the analysis in the table below. [1 Points]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *SS* | *DF* | *F* | *p* |
| School Group | 19.0 | 2 | 13.51 | <0.001 |
| Residuals | 574.1 | 817 | - | - |
| Total | 593.1 | 819 | - | - |

1. Assess whether the results of the one-way ANOVA performed in part 4A are still significant after controlling for how often participants have visited the doctor (SeeDr90Days\_04). Fill in the results in the table below. Use Type I Sum of Squares, which is the default for using summary() around an aov object. [1 Points]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *SS* | *DF* | *F* | *p* |
| Doctor Visits | 16.7 | 1 | 24.36 | <0.001 |
| School Group | 18.0 | 2 | 13.12 | <0.001 |
| Residuals | 558.5 | 816 | - | - |
| Total | 593.2 | 819 | - | - |

1. Without performing any post hoc tests, what are TWO conclusions that could be drawn based only on these analyses (i.e., 4A and 4B)? You do not need to write a full interpretation, report any statistics, or comment on any means or the direction of associations. Hint: There are at least three conclusions that could be drawn. [1 Point]
2. School group is a significant predictor of overall health.
3. Even after controlling for the number of visits to the doctor, there were still differences in overall health between school groups.
4. Write up an interpretation of your results, referring to the lab for an example [1 point].

In this problem, we investigated whether there were differences between people with high school, undergraduate, or graduate degrees in overall health. An initial ANOVA revealed that there were statistically significant differences between the groups (F(2, 817) = 13.51, p < .001). To assess whether this difference remained significant when controlling for how often participants have visited the doctor, a follow-up ANCOVA was performed with overall health as the dependent variable, school group as the independent variable, and the number of visits to the doctor as a covariate. The analysis showed that the differences between the groups in overall health remained significant (F(2, 816) = 13.12, p < .001) even when number of doctor visits was controlled for. Doctor visits was a significant covariate (F(1, 816) = 22.88, p < .001).

1. Conduct the orthogonal contrasts mentioned below, and fill in the table. Be sure to still control for how often participants visit the doctor! [1 points]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *SS* | *DF* | *F* | *p* |
| Doctor Visits | 16.7 | 1 | 24.36 | <0.001 |
| School Group | 18.0 | 2 | 13.12 | <0.001 |
| High School vs. Rest | 16.4 | 1 | 23.94 | <0.001 |
| Undergraduate vs. Graduate | 1.6 | 1 | 2.31 | 0.129 |
| Residuals | 558.5 | 816 | - | - |
| Total | 593.2 | 819 | - | - |

1. What can you conclude from the orthogonal contrasts? In other words, which contrasts were significant? [1 point]

The orthogonal contrasts showed that, after controlling for the doctor visits, people with high school degrees have higher overall health than people with undergraduate and graduate degrees combined. The second contrast revealed that undergraduate and graduate groups have similar overall health scores.

**2. ANCOVA – Test for Moderation**

1. Perform an ANCOVA to assess whether the association between how often participants visit the doctor (SeeDr90Days\_04) and overall health (Overallhealth\_04) is moderated by heart health status (HeartHealth\_04). Use SeeDr90Days\_04 as the dependent variable. Use Type II Sum of Squares (hint: use the Anova() function). [1 Point]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *SS* | *DF* | *F* | *p* |
| Overall Health | 4060 | 1 | 19.34 | <0.001 |
| Heart Health | 17 | 1 | 0.080 | 0.777 |
| Interaction | 904 | 1 | 4.31 | 0.038 |
| Residuals | 172992 | 824 | - | - |
| Total | 177973 | 827 | - | - |

1. Perform follow up simple effect analyses (using linear regression, or the lm() function) to assess whether overall health predicts number of doctor visits at each level of heart health, and fill in the results of the analysis in the table below. [1 Points]

|  |  |  |
| --- | --- | --- |
|  | *b* | *p* |
| Poor | 4.09 | <0.001 |
| Good | 1.57 | 0.057 |

1. Create a graph showing the relation between overall health (on the x axis) and number of doctor visits (on the y axis) for each of the 2 groups of heart health. Make sure to adjust the axis scales so that the data on the graphs can be clearly seen. Hint: You can easily create panels for different group with the facet\_wrap() function with ggplot2. [1 Point]

A graph of poor and good

Description automatically generated

1. Write up an interpretation of your results, mentioning both the ANOVA results (as relevant) and the results of the simple effects analyses. [1 Points]

In this problem we investigated whether heart health status moderated the association between the frequency of doctor visits and overall health. To test this hypothesis, a ANCOVA model was performed, with Doctor visits as the dependent variable and with Overall Health, Heart health status, and the interaction between these variables as independent variables. The analysis revealed a significant interaction between Overall Health and Heart health status(F(1, 824) = 4.31, p = .04). To assess how the association between frequency of doctor visits and overall health differed between heart health status, follow up simple effect analyses were performed using simple regression, with doctor visits as the dependent variable and overall health as the independent variable, and separate analyses being performed for each heart health status. The analyses revealed a significant association between doctor visits and overall health in the group with Poor heart health (b = 4.09, p < .001), but not in the group with Good heart health status (b = 1.57, p = .06).

**Extra Credit Question**

How do the results in 1B (in particular, the main effect of number of doctor visits) change if you used Type II SS? Explain why. [1 point]

The main effect of school group does not change because in both analyses this effect is being controlled by the number of doctor visits. On the other hand, the main effect of the number of doctor visits has a lower Type II sum of squares than Type I because in the first case we are partialling out the variance explained by the school group.

Syntax for Question 1

## A

```{r}

ovrhlt\_aov <- aov(Overallhealth\_04 ~ school.group, data = lab.data)

summary.aov(ovrhlt\_aov)

```

## B

```{r}

ovrhlt\_aov2 <- aov(Overallhealth\_04 ~ SeeDr90Days\_04 + school.group , data = lab.data)

Anova(ovrhlt\_aov, type= 2)

```

## C

```{r}

lab.data$school.group = factor(lab.data$school.group, levels = c("HighSchool", "Undergraduate", "Graduate"))

contrasts1 = cbind(

c1 = c(0, 1, -1),

c2 = c(1, -0.5, -0.5)

)

contrasts(lab.data$school.group) = contrasts1

split\_list = list(

school.group = list("Undergrad vs Grad" = 1,

"HS vs Rest" = 2)

)

summary.aov(aov(Overallhealth\_04 ~ SeeDr90Days\_04 + school.group , data = lab.data), split = split\_list)

```

Syntax for Question 2

## A

```{r}

Anova(aov(SeeDr90Days\_04~Overallhealth\_04 \* HeartHealth\_04, data = lab.data), type = 2)

```

## B

```{r }

summary(lm(SeeDr90Days\_04 ~ Overallhealth\_04, data = lab.data %>% filter(HeartHealth\_04 == "Good")))

summary(lm(SeeDr90Days\_04 ~ Overallhealth\_04, data = lab.data %>% filter(HeartHealth\_04 == "Poor")))

```

## C

```{r }

lab.data |>

dplyr::filter(!is.na(HeartHealth\_04)) |>

ggplot( aes(x = Overallhealth\_04, y = SeeDr90Days\_04, fill = HeartHealth\_04)) +

geom\_smooth(method = "lm", fullrange = T, color = "black") +

geom\_point(shape = 21, alpha = .4) +

theme\_bw() +

facet\_wrap(~HeartHealth\_04) +

ylim(0, 20)

```