Homework 2

Peyton Hall

01/29/2025

Load Necessary Libraries

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(ggplot2)
library(readxl)
```

Question 1

```r
# The following data come from a study that examines the efficacy of saliva
# cotinine as an indicator for exposure to tobacco smoke. In one part of the
# study, seven subjects were each required to smoke a single cigarette. Samples
# of saliva were taken from all individuals 12 and 24 hours after smoking the
# cigarette and cotinine levels were recorded.

# Cotinine levels (nmol/l)
# Individuals, After 12 hours, After 24 hours
# 1,           73,             24
# 2,           58,             27
# 3,           67,             49
# 4,           93,             59
# 5,           33,             0
# 6,           18,             11
# 7,           147,            43

# Perform an appropriate test to see if the mean cotinine levels after 12 hours
# is significantly higher than the mean cotinine level after 24 hours with 0.05
# as the significance level.
```

```r
# Step 1: Formulate the null and alternative hypotheses using symbols
# Null Hypothesis (H_0): The mean cotinine levels after 12 hours are equal to
#                        the mean cotinine levels after 24 hour.
# _12 = _24
# Alternative Hypothesis (H_a): The mean cotinine levels after 12 hours are
#                               higher than the mean cotinine levels after 24
#                               hours.
# _12 > _24

# Step 2: Choose the _____ tailed ____________test
#         Choose the one-tailed paired t-test.

# Step 3: Find the test statistic and p-value from R
after_12_hours <- c(73, 58, 67, 93, 33, 18, 147)
after_24_hours <- c(24, 27, 49, 59, 0, 11, 43)
t.test(after_12_hours, after_24_hours, alternative = "greater", paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  after_12_hours and after_24_hours
## t = 3.3228, df = 6, p-value = 0.007974
## alternative hypothesis: true mean difference is greater than 0
## 95 percent confidence interval:
##  16.37073      Inf
## sample estimates:
## mean difference
##        39.42857
```

```r
# Step 4: Make a decision based to the null hypothesis; and explain the decision
# in the context
```

t = 3.3228; p-value = 0.007974 Reject H0; there is evidence to support that the mean cotinine levels after 12 hours are significantly higher than the mean cotinine levels after 24 hours. This suggests that the cotinine concentration in saliva decreases significantly over time after smoking a cigarette.

Question 2

```r
# The Bayley Scales of Infant Development yield scores on the Psychomotor
# Development Index (PDI). PDI can be used to assess a child's level of
# functioning at approximately one year of age. As part of a study investigating
# the development status of children who had undergone reparative heart surgery
# during the first three months of life, the Bayley Scales were administered to
# a sample of one-year-old infants born with congenital heart disease. The
# children had been randomized to one of two different treatment groups, known
# as "circulatory arrest" (denoted as treatment group 0) and "low-flow bypass"
# (denoted as treatment group 1). The groups differed in the specific way in
# which the reparative surgery was performed. Unlike circulatory arrest,
# low-flow bypass maintains continuous circulation through the brain; although
# it is felt to be preferable by some physicians, it also has its own associated
# risk of brain injury. The data is on d2l (PDI score). Assume the two
# treatments have unequal variances. At the 0.05 level of significance, test the
# claim that the mean PDI score at one year of age for the circulatory arrest
```

```r
# treatment group is significantly different from that for the low-flow group.

PDI_data <- read_excel("~/Desktop/STAT 301/Week 2/PDI data.xlsx")
PDI_data
```

```
## # A tibble: 143 x 2
##      PDI Treatment
##    <dbl> <chr>
##  1    80 0
##  2    98 0
##  3    98 0
##  4   111 0
##  5    82 0
##  6    86 0
##  7   122 0
##  8    78 0
##  9    92 0
## 10    86 0
## # i 133 more rows
```

```r
# Step 1: Formulate the null and alternative hypotheses using symbols
# Null Hypothesis (H_0): The mean PDI score at one year of age for the
#                        circulatory arrest treatment group (TG0) is equal to
#                        that for the low-flow group (TG1).
# _TG0 = _TG1
# Alternative Hypothesis (H_a): The mean PDI score at one year of age for the
#                               circulatory arrest treatment group (TG0) is not
#                               equal to that for the low-flow group (TG1).
# _TG0   _TG1

# Step 2: Choose the ______tailed ___________test
#         Choose the two-tailed t-test.

# Step 3: Find the test statistic and p-value from R

pdi_tg0 <- PDI_data$PDI[PDI_data$Treatment == 0]  # Data for circulatory arrest
pdi_tg1 <- PDI_data$PDI[PDI_data$Treatment == 1]  # Data for low-flow bypass
# perform the Welch t-test for unequal variances
t.test(pdi_tg0, pdi_tg1, alternative = "two.sided", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  pdi_tg0 and pdi_tg1
## t = -2.1715, df = 139.56, p-value = 0.03158
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.8723423  -0.5094326
## sample estimates:
## mean of x mean of y
##  91.91781  97.60870
```

```r
# Step 4: Make a decision to H0, and explain the decision in terms of the
# original claim.

# Which group has a higher PDI score? Low-flow bypass group (TG1).
```

t = -2.1715; p-value = 0.03158 Reject H0; there is evidence to support that the mean PDI score at one year of age for the circulatory arrest treatment group is significantly different from that for the low-flow bypass group.

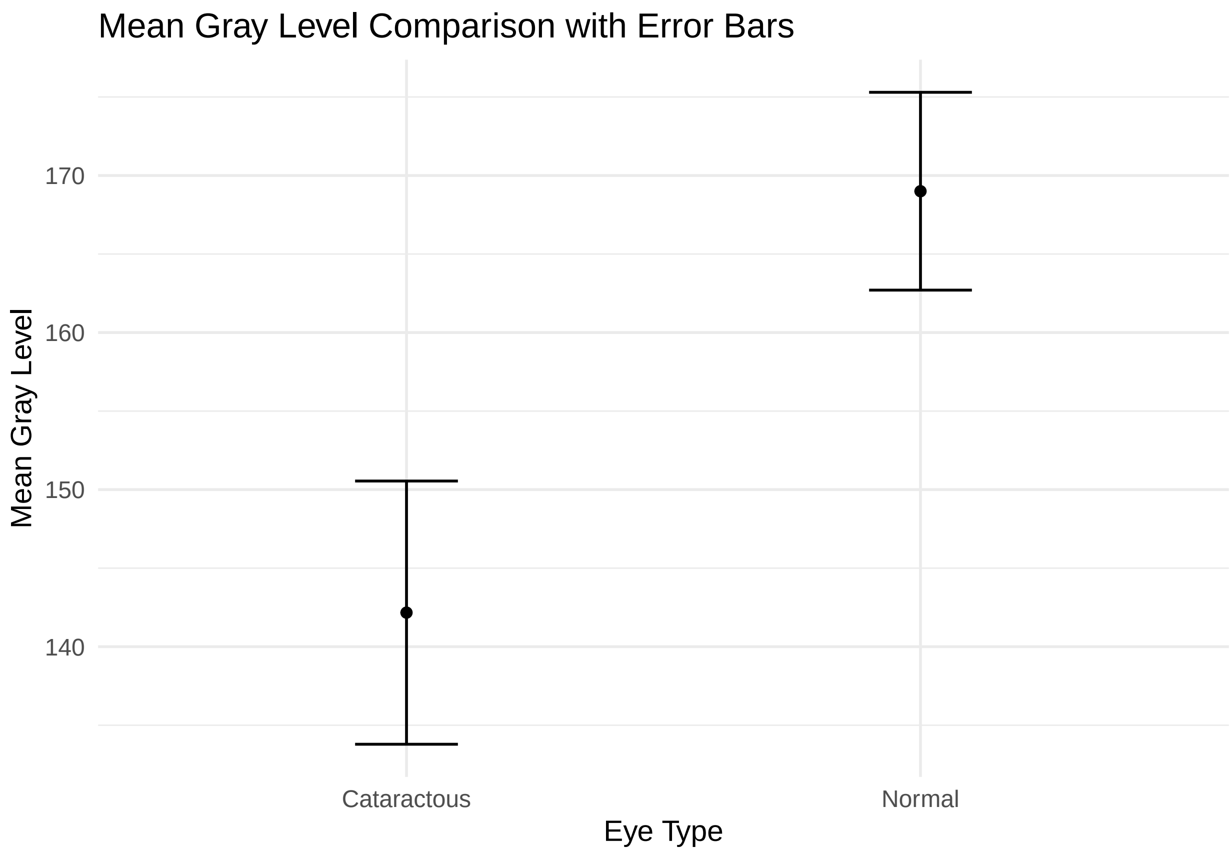
Question 3

```r
# A camera has been developed to detect the presence of cataract more
# accurately. Using the camera, the gray level of each point in the lens of a
# human eye can be characterized into 256 gradations. To test the camera,
# photographs were taken of six randomly selected normal eyes and six randomly
# selected cataractous eyes from a different group of people. The gray level of
# each eye was computed in the lens. The data are given in the following table.
# Use the 0.05 significance level to test the claim that the mean gray level
# from cataractous eyes is significantly lower than the mean gray level from
# normal eyes.

# Cataractous gray level, Normal gray level
# 161,                    158
# 140,                    182
# 136,                    185
# 161,                    145
# 106,                    167
# 149,                    177

# a) Generate a point graph to plot the mean difference between cataractous eyes
# and normal eyes with error bars. Which one has a higher mean gray level?
eye_data <- data.frame(
  Cataractous = c(161, 140, 136, 161, 106, 149),
  Normal = c(158, 182, 185, 145, 167, 177)
)

eye_data$Difference = eye_data$Normal - eye_data$Cataractous
summary_stats <- eye_data %>%
  summarise(Mean_Difference = mean(Difference), SD = sd(Difference))
summary_stats
```

```
##   Mean_Difference       SD
## 1        26.83333 30.38037
```

```r
means <- colMeans(eye_data[, c("Cataractous", "Normal")])
errors <- apply(eye_data[, c("Cataractous", "Normal")], 2, sd) / sqrt(nrow(eye_data))

error_bars <- data.frame(
  Group = c("Cataractous", "Normal"),
  Mean = as.numeric(means),
  Error = as.numeric(errors)
)
```

```r
ggplot(error_bars, aes(x = Group, y = Mean, ymin = Mean - Error, ymax = Mean + Error)) +
  geom_point() +
  geom_errorbar(width = 0.2) +
  labs(title = "Mean Gray Level Comparison with Error Bars",
       x = "Eye Type",
       y = "Mean Gray Level") +
  theme_minimal()
```




```r
# b) Test the hypothesis in the question with the steps below:

# Step 1: Formulate the null and alternative hypotheses using symbols
# Null Hypothesis (H0): The mean gray level from cataractous eyes is equal to
#                       the mean gray level from normal eyes.
# _cataractous = _normal
# Alternative Hypothesis (H_a): The mean gray level from cataractous eyes is
#                               less than the mean gray level from normal eyes.
# _cataractous < _normal

# Step 2: Choose the ______tailed ___________test
#         Choose the one-tailed t-test

# Step 3: Find the test statistic and p-value from R
t.test(eye_data$Cataractous, eye_data$Normal, alternative = "less")
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  eye_data$Cataractous and eye_data$Normal
## t = -2.5597, df = 9.2821, p-value = 0.015
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##    -Inf -7.683
## sample estimates:
## mean of x mean of y
##  142.1667  169.0000
```

```
# Step 4: Make a decision to H0, and explain the decision to address the
# original cl
```

t = -2.5597; p-value = 0.015 Reject H0; there is evidence to support that the mean gray level from cataractous eyes is significantly lower than the mean gray level from normal eyes.

Question 4

```
# Data below lists the times required for randomly selected eight flights to
# taxi out for takeoff and the corresponding times required to taxi in after
# landing for the same flight. All eight flights are American Airlines from New
# York to Los Angeles. Use the significance level of 0.01 to test the claim that
# the average time to taxi in is different from the average time to taxi out.

# Flight,     1,  2,  3,  4,  5,  6,  7,  8
# Taxi-out, 30, 19, 12, 19, 18, 22, 19, 13
# Taxi-in,  12, 13,  8, 21, 17, 15, 26, 12

# Step 1: State the null and alternative hypotheses using the correct symbols
# Null Hypothesis (H_0): The mean taxi-in time ( _in) is equal to the mean
#                        taxi-out time ( _out).
# _in = _out
# Alternative Hypothesis (H_a): The mean taxi-in time ( _in) is not equal to the
#                               mean taxi-out time ( _out).
# _in   _out

# Step 2: Choose the ____tailed ________________ test
#         Choose the two-tailed paired t-test.

# Step 3: Find the test statistic and p-value in R
taxi_out <- c(30, 19, 12, 19, 18, 22, 19, 13)
taxi_in <- c(12, 13, 8, 21, 17, 15, 26, 12)
t.test(taxi_in, taxi_out, paired = TRUE, alternative = "two.sided")
```

```
##
##  Paired t-test
##
## data:  taxi_in and taxi_out
## t = -1.3401, df = 7, p-value = 0.2221
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -9.675894  2.675894
## sample estimates:
```

```
## mean difference
##            -3.5
```

```r
mean(taxi_out)
```

```
## [1] 19
```

```r
mean(taxi_in)
```

```
## [1] 15.5
```

```r
# Step 4: Make a decision to the null hypothesis; and explain the decision to
# address the original claim.
```

t = -1.3401; p-value = 0.2221 Fail to reject H0; there is no evidence to support that the average time to taxi in is different from the average time to taxi out. Based on their means (taxi out being 19 and taxi in being 15.5), there is a difference, yet that difference is not statistically significant due to the significance level (= 0.01).