Assignment_10

Zhaowen Fan
&
Rafael Ignacio Gonzalez Chong

UT Austin GSI

(1) Summarize the data by whether children participated in the meal preparation or not. Use an appropriately labelled table to show the results. Also include a graphical presentation that shows the distribution of calories for participants vs. non-participants. Describe the shape of each distribution and comment on the similarity (or lack thereof) between the distributions in each population.

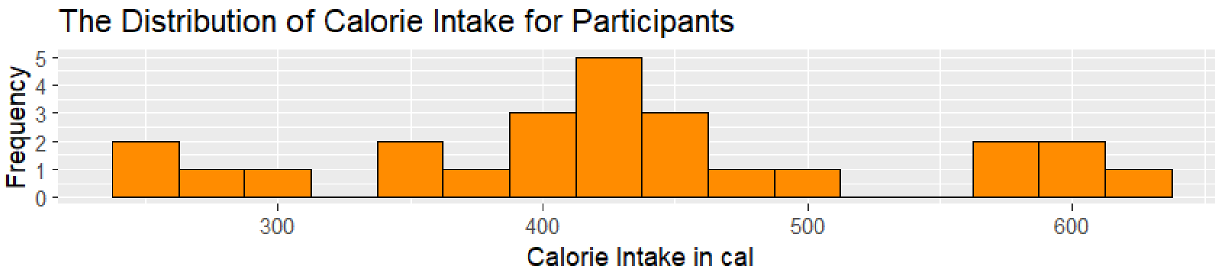
>>> The code:

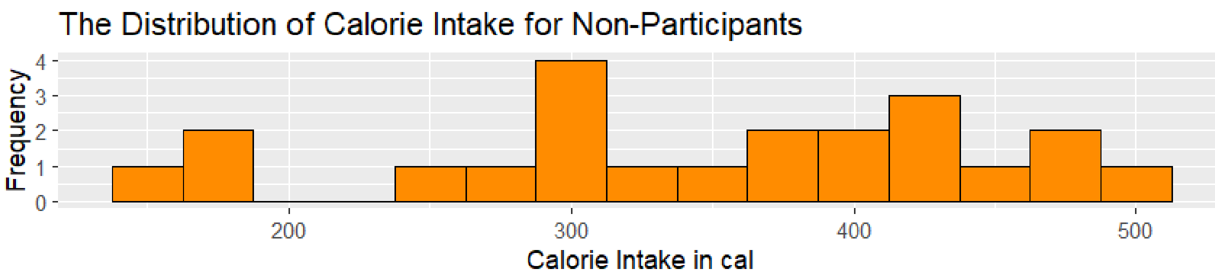
```
participant <- data %>%
  filter(Participant == "1")
non_participant <- data %>%
  filter(Participant == "0")

aggregate(CalorieIntake ~ Participant, data, summary)

p1 <- ggplot(participant, aes(x = CalorieIntake)) +
  geom_histogram(binwidth = 25, fill = "darkorange", color = "black") +
  labs(
    title = "The Distribution of Calorie Intake for Participants",
    x = "Calorie Intake in cal",
    y = "Frequency"
  )

p2 <- ggplot(non_participant, aes(x = CalorieIntake)) +
  geom_histogram(binwidth = 25, fill = "darkorange", color = "black") +
  labs(
    title = "The Distribution of Calorie Intake for Non-Participants",
    x = "Calorie Intake in cal",
    y = "Frequency"
  )

grid.arrange(p1, p2, ncol = 1)
```

>>> The output:

```
 Participant CalorieIntake.Min. CalorieIntake.1st Qu. CalorieIntake.Median
CalorieIntake.Mean
1           0           139.6900              290.3975              361.0200
346.7991
2           1           249.8600              368.5100              428.7400
431.3996
  CalorieIntake.3rd Qu. CalorieIntake.Max.
1              422.5675           503.4600
2              477.9600           635.2100
```

>>> The graph:





According to the histogram presented above, we can see that the distribution of calorie intake for participants is higher in average than non-participants, the maximum is higher than non-participant and the minimum is lower than non-participants. In addition, the mean and the median of participants are higher than non-participants.

(2) Does the mean calorie consumption for those who participated in the meal preparation differ from 425? Formally test at the level using the 5 steps outlined in the module.

>>> The code:

```
# Step 1: State the hypothesis:
# H0: mu = 425 (calorie intake for participants in the meal preparation is 425)
# H1: mu != 425 (calorie intake for participants in the meal preparation isn't 425)

# Step 2: Choose the significance level:
# alpha = 0.05

# Step 3: Compute the test statistic:
# Perform one-sample t-test:
t.test(data$CalorieIntake[data$Participant == 1], mu = 425)

# Step 4: Decision rule:
# Compare the p_value with the alpha
```

Step 5: Conclusion:
Failed to reject the null hypothesis!
print(
 "There is insufficient evidence to conclude that the mean calorie intake differs from 425."
)

>>> The output:

```
> t.test(data$CalorieIntake[data$Participant == 1], mu = 425)

      One Sample t-test

data:  data$CalorieIntake[data$Participant == 1]
t = 0.30272, df = 24, p-value = 0.7647
alternative hypothesis: true mean is not equal to 425
95 percent confidence interval:
 387.7683 475.0309
sample estimates:
mean of x
 431.3996

[1] "There is insufficient evidence to conclude that the mean calorie intake differs from 425."
```

>>> The conclusion:

There is insufficient evidence to conclude that the mean calorie consumption for those who participated in the meal preparation differ from 425.

(3) Calculate a 90% confidence interval for the mean calorie intake for participants in the meal preparation. Interpret the confidence interval.

>>> The code:

```
sample_mean = mean(participant$CalorieIntake)
sample_sd = sd(participant$CalorieIntake)
sample_size = nrow(participant)
df <- sample_size - 1
t_critical <- qt(0.95, df)
margin_error <- t_critical * (sample_sd / sqrt(sample_size))

CI_lower <- sample_mean - margin_error
CI_upper <- sample_mean + margin_error
```

c(CI_lower, CI_upper)

cat(
  "We are 90% confident that the true mean calorie intake for participants in the",
  "meal preparation falls between",
  CI_lower, "and", CI_upper, "calories."
)

>>> The output:

```
> c(CI_lower, CI_upper)
[1] 395.2311 467.5681
```

>>> The interpretation from the output:

```
We are 90% confident that the true mean calorie intake for participants in the
meal preparation falls between 395.2311 and 467.5681 calories.
```

(4) Formally test whether or not participants consumed more calories than non-participants at the level using the 5 steps outlined in the module.

(5) Are the assumptions of the test used in (4) met? How do you know?

>>> The code:

Step 1: State the hypothesis:
H0: mu1 = mu2 (the mean of calorie intake for both categories are the same)
H0: mu1 > mu2 (the mean of participant is bigger than non-participant)

Step 2: Choose the significance level:
alpha = 0.05

# Step 3: Compute the test statistic:
t.test(participant, non_participant, alternative = "greater", var.equal = FALSE)

Step 4: Decision rule:
Compare the p_value with alpha

Conclusion:
Failed to reject the null hypothesis!
cat(
  "At the 0.05 significance level, we failed to reject the null hypothesis.",

```
  "The p-value was greater than 0.05, indicating that the observed difference",
  "in calorie intake between participants and non-participants is not",
  "statistically significant. Thus, we do not have sufficient evidence to",
  "conclude that participants consumed more calories than non-participants."
)
```

>>> The output:

```
At the 0.05 significance level, we failed to reject the null hypothesis. The
p-value was greater than 0.05, indicating that the observed difference in
calorie intake between participants and non-participants is not statistically
significant. Thus, we do not have sufficient evidence to conclude that
participants consumed more calories than non-participants.
```

>>> Conclusion:

The assumption in task 4 does not meet! The p_value is 0.99 and the alpha is 0.05, and it is obvious that 0.99 > 0.05. Therefore, we failed to reject the null hypothesis. Thus, we do not have sufficient evidence to conclude that participants consumed more calories than non-participants.