MET CS 555 Assignment 2 – 20 points

SUBMISSION REQUIREMENTS: **Please submit a single document (word or PDF) for submission. Your submission should contain a summary of your results (and answers to questions asked on the homework) as well as your R code used to generate your results (please append your R code to the end of your submission).**

**An experiment was conducted to determine the effect of children participating in a given meal preparation on calorie intake for that meal. Data are recorded below. Save the data to a format that can be read into R. Read the data in for analysis. Use R to calculate the quantities and generate the visual summaries requested below. You will lose points if you are not utilizing R.**

**Import Data: R code:**

setwd("C:/Users/XuWei/OneDrive/BU - Boston University/04 - 2022 Spring Semester/MET CS 555 A2 Data Analysis and Visualization with R/Assignment 2")

data <- read.csv("Date Set for Assignment 2.csv", header = TRUE)

head(data)

participants <- data$CalorieIntakeforparticipants

nonParticipants <- data$CalorieIntakeforNon.participants

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. **(2 points)**

**Summary：Participated**

**Min. 1st Qu. Median Mean 3rd Qu. Max.**

**211.0 298.4 424.9 410.1 456.3 635.2**

**Summary: non-Participated**

**Min. 1st Qu. Median Mean 3rd Qu. Max. NA's**

**139.7 296.4 374.7 374.1 445.6 688.8 3**

**Histogram Grphic:**

图表, 直方图

描述已自动生成

**As we can see from the histogram graph, both the distribution of participated and non-participated are not symmetric. And for the participants, it follows the positively skewed distribution. For non-participants, it follows normal distribution.**

**R code:**

**summary(participants)**

**summary(nonParticipants)**

**par(mfrow = c(1, 2))**

**hist(participants)**

**hist(nonParticipants)**

1. Does the mean calorie consumption for those who participated in the meal preparation differ from 425? **Formally test at the alpha = 0.05 level using the 5 steps** outlined in the module. **(6 points)**

**5 Steps:**

1. **Set up the hypotheses and select the alpha level**
   1. **H0: μ= 425**
   2. **H1: μ ≠ 425**
   3. **alpha = 0.05**
2. **Select the appropriate test statistic**
   1. **t = -0.61394**
3. **State the decision rule**
   1. **Reject H0 if p-value < alpha**
4. **Compute the test static and the associated p-value.**
   1. **p-value = 0.545**
5. **State your conclusion**
   1. **Since p-value > alpha, fail to reject H0**
   2. **This does not mean calorie consumption for those who participated in the meal preparation differ from 425.**

**R code:**

**t.test(participants, mu=425)**

**R output:**

**One Sample t-test**

**data: participants**

**t = -0.61394, df = 24, p-value = 0.545**

**alternative hypothesis: true mean is not equal to 425**

**95 percent confidence interval:**

**359.9212 460.2380**

**sample estimates:**

**mean of x**

**410.0796**

1. Calculate a **90% confidence interval** for the mean calorie intake for participants in the meal preparation. Interpret the confidence interval. **(4 points)**
   1. **The 90% confidence interval for the mean calorie intake for participants in the meal preparation is between 358.5004 and 451.6588.**

**R code:**

**t.test(participants, conf.level = 0.9)**

**R output:**

**One Sample t-test**

**data: participants**

**t = 16.874, df = 24, p-value = 8.15e-15**

**alternative hypothesis: true mean is not equal to 0**

**90 percent confidence interval:**

**368.5004 451.6588**

**sample estimates:**

**mean of x**

**410.0796**

1. Formally test whether or not participants consumed more calories than non-participants at the alpha = 0.05 level **using the 5 steps** outlined in the module. **(6 points)**

**5 steps:**

1. **Set up the hypotheses and select the alpha level**
   1. **H0: μ1 = μ2**
   2. **H1: μ1 > μ2**
   3. **alpha = 0.05**
2. **Select the appropriate test static**
   1. **t = 0.9636**
3. **State the decision rule**
   1. **Reject H0 if p-value < alpha**
4. **Compute the test statistic and the associated p-value**
   1. **p-value = 0.1703**
5. **Conclusion**
   1. **Since p-value > alpha, fail to reject H0**
   2. **We have no significant evidence concluded that participants consumed more calories than non-participants at the alpha = 0.05 level.**

**R code:**

**t.test(participants, nonParticipants, alternative = "greater")**

**R output:**

**Welch Two Sample t-test**

**data: participants and nonParticipants**

**t = 0.9636, df = 42.901, p-value = 0.1703**

**alternative hypothesis: true difference in means is greater than 0**

**95 percent confidence interval:**

**-26.81338 Inf**

**sample estimates:**

**mean of x mean of y**

**410.0796 374.0718**

(5) Are the assumptions of the test used in (4) met? How do you know? **(2 points)**

**R code:**

**diff <- participants - nonParticipants**

**hist(diff)**

图表, 直方图

描述已自动生成

**I think it met the assumption of the test. Since the graphic of difference of calorie intake of participants and calorie intake of non-participants follow the normal distribution.**

Data Set for Assignment 2

**Calorie Intake for participants**

|  |
| --- |
| 230.16 |
| 210.99 |
| 288.73 |
| 590.28 |
| 582.59 |
| 635.21 |
| 249.86 |
| 441.66 |
| 572.43 |
| 357.78 |
| 396.79 |
| 298.38 |
| 282.99 |
| 368.51 |
| 388.59 |
| 256.32 |
| 408.82 |
| 424.94 |
| 477.96 |
| 428.74 |
| 432.52 |
| 428.27 |
| 596.79 |
| 456.30 |
| 446.38 |

**Calorie intake for non-participants**

|  |
| --- |
| 614.61 |
| 503.46 |
| 425.22 |
| 688.77 |
| 184.00 |
| 299.73 |
| 350.65 |
| 394.94 |
| 261.55 |
| 295.28 |
| 139.69 |
| 462.78 |
| 179.59 |
| 301.75 |
| 436.58 |
| 371.39 |
| 469.02 |
| 378.09 |
| 287.31 |
| 448.55 |
| 332.64 |
| 403.98 |