MET CS 555 Assignment 2 – 20 points Spring, 2018

**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 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.**

(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.

**Labelled table generated from code:**

**setwd("/Users/anthony.valencia/met/cs555/assignments/02/")**

**participants <- read.csv("participants")**

**participants <- participants$CalorieIntakeForparticipants**

**non.participants <- read.csv("non-participants")**

**non.participants <- non.participants$CalorieIntakeForNon.participants**

**max.len = max(length(participants), length(non.participants))**

**participants = c(participants, rep(NA, max.len - length(participants)))**

**non.participants = c(non.participants, rep(NA, max.len - length(non.participants)))**

**df <- data.frame(participants, non.participants)**

**names(df) <- c("participants", "non.participants")**

**df**

**participants non.participants**

**1 435.16 414.61**

**2 338.99 503.46**

**3 488.73 425.22**

**4 590.28 288.77**

**5 582.59 184.00**

**6 635.21 299.73**

**7 249.86 350.65**

**8 441.66 394.94**

**9 572.43 261.55**

**10 357.78 295.28**

**11 396.79 139.69**

**12 298.38 462.78**

**13 282.99 179.59**

**14 368.51 301.75**

**15 388.59 436.58**

**16 256.32 371.39**

**17 408.82 469.02**

**18 424.94 378.09**

**19 477.96 287.31**

**20 428.74 448.55**

**21 432.52 332.64**

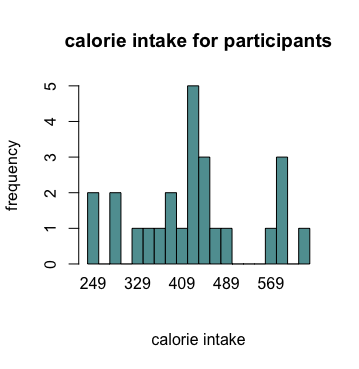
**22 428.27 403.98**

**23 596.79 NA**

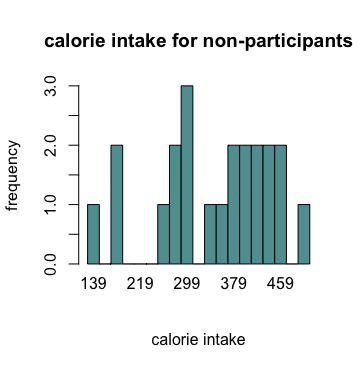
**24 456.30 NA**

**25 446.38 NA**

**Representation of participants:**



**Representation of non-participants:**



**Both calorie intakes appear somewhat normally distributed. That for participants is more narrowly distributed around 415 calories, while that for non-participants has a broader distribution centered around 300.**

(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 following code was used:**

**sample.mean.participants <- mean(participants)**

**sample.sd.participants <- sd(participants)**

**n.participants <- length(participants)**

**t <- (sample.mean - 425)/(sample.sd/sqrt(n))**

**to draw the conclusions:**

**critical value for df = 25 - 1 = 24 and p = 0.05 is 1.711**

**reject H0 if |t|≥1.711**

**|t| = 0.3027212, so calorie consumption does not differ**

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

**The following code:**

**sample.mean.participants + (1.711 \* (sample.sd.participants/sqrt(n.participants))) sample.mean.participants - (1.711 \* (sample.sd.participants/sqrt(n.participants)))**

**sample.mean**

**suggests that the 90% confidence interval (395.2286, 467.5706) with a mean of 431.3996 calories for participants in the study.**

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

**The following code:**

**sample.mean.non.participants <- mean(non.participants)**

**sample.sd.non.participants <- sd(non.participants)**

**n.non.participants <- length(non.participants)**

**s.non.participants.sqd <- sample.sd.non.participants^2**

**s.participants.sqd <- sample.sd.participants^2**

**t <- (sample.mean.participants - sample.mean.non.participants)/(sqrt((s.participants.sqd/n.participants) + (s.non.participants.sqd/n.non.participants)))**

**suggests:**

**t = 2.824836**

**with a critical value for df = 25 - 1 = 24 and p = 0.05 is 1.711**

**where t = 2.824836 >= 1.711, so we reject H0 (participants consumed more calories than non-participants).**

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

**The assumptions used in 4 are the following:**

**1. samples must be independent (not infueniing eaih other) and randomly selected from the two distinct populations of interest**

**the samples are assumed to be randomly selected fro mtwo populations of interest, but attmitedly, this is assumed.**

**2. the variable of interest must be measured in the same way in each of the populations**

**calorie consumption is assumed to be measured in the same way.**

**3. the parameter of interest should be normally distributed (or at least have similar shapes and without outliers)**

**Per the images shown in question 1 above, the parameter of interest for each sample population appears to be normally distributed.**

**Calorie Intake for participants**

|  |
| --- |
| 435.16 |
| 338.99 |
| 488.73 |
| 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**

|  |
| --- |
| 414.61 |
| 503.46 |
| 425.22 |
| 288.77 |
| 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 |