

Stat 451

Fall 2020

Homework 8

Due Thur. Nov 19, by midnight on Moodle as an Rmd file (not an html file). No late HW is accepted for this assignment. There is no grad problem on this assignment.

The file MM-All.csv contains the gross (not net) weights (in grams) of a sample of packages of M&M's I bought in 2016 and a second sample from 2017. These were all labeled 49.3 grams, net weight. The weight of the packaging (called the tare weight) is 1.3 grams, so subtract that amount to get the net weights. Don't change the csv file: create a new variable in R this way:

```
mm <- read_csv("MM-All.csv")  
mm <- mutate(mm, NetWt=GrossWeight-1.3)
```

Assume these are simple random samples from shipments of M&M's the store received in the respective years. We're interested in whether the mean net weight of packages in the 2017 shipment is different than in the 2016 shipment and how big that difference might be. Note that the claimed net weight is not part of this analysis; you're simply comparing the two sets of packages.

- a) Create side-by-side boxplots of the net weights with proper labels and describe what you find (like back in Chapter 4), focusing on the comparison of the two samples.
- b) Do the assumptions of t procedures appear satisfied?
- c) Report the group means, standard deviations and sample sizes in a table (see the summaryBy command in 10 d) on p. 6 of "Some relevant R commands" in Moodle).
- d) Report the 99% confidence interval for the difference in means and interpret in the context of the problem. Use the `t.test` command. You can extract the endpoints of the confidence interval by saving the result of the `t.test` command, say in `tt`. Then the endpoints of the confidence interval are `tt$conf.int[1]` and `tt$conf.int[2]`.
- e) Report the P-value (from `t.test`) for a two-sided test of the hypothesis that the two population means are equal and interpret in the context of the problem. Do not use reject/don't reject or significant/not significant terminology. Talk about strength of evidence. You can extract the P-value from the saved `t.test` command with `tt$p.value`.