Week 6 Day 2
Stat140-04

Part I: Cherry Blossom Race

Is the typical US runner getting faster or slower over time? We consider this question in the context of the Cherry Blossom Race, which is a 10-mile race in Washington, DC each spring.

The average time for all runners who finished the Cherry Blossom Race in 2006 was 93.29 minutes (93 minutes and about 17 seconds). We want to determine using data from 100 participants in the 2017 Cherry Blossom Race whether runners in this race are getting faster or slower, versus the other possibility that there has been no change.

Step 1: Set up the hypotheses

What are appropriate hypotheses for this context?

Step 2: Check the conditions

The data come from a simple random sample of all participants, so the observations are independent. However, should we be worried about the normality condition? See the following histogram of the differences and evaluate if we can move forward. If you forget about the conditions for t-distribution, check out materials on Week 5 Day 3.
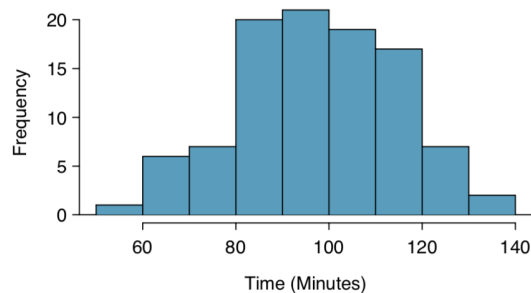

Figure 1: A histogram of time for the sample Cherry Blossom Race data.

- Independence condition

- Nearly normal condition

Step 3: Compute the test statistics

The sample mean and sample standard deviation of the sample of 100 runners from the 2017 Cherry Blossom Race are 97.32 and 16.98 minutes, respectively. Recall that the sample size is 100 and the average run time in 2006 was 93.29 minutes. Compute the test statistics.

Step 4: Color the p-value

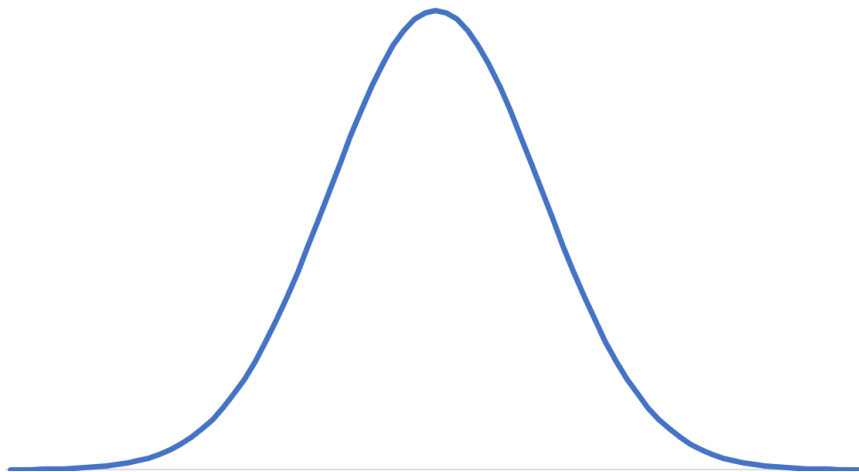In the following t-model curve, mark the center and color the p-value.

Figure 2: A t-model curve

Part II: `tail.p` function

Once you computed a z or t statistic by hand using the formula

$$z = \frac{statistic - null}{SE}$$

and then compare this statistic to the standard normal distribution or t distribution with the command `pnorm` or `pt` respectively.

Using `pnorm` for normal distribution

For example, to find the two tailed p-value for a z statistic of 2.4, you would type

`2 * pnorm(2.4, lower.tail = FALSE)`

In general, `p-value = pnorm(z, lower.tail = )`.

- Left-Tailed: `p-value = pnorm(z, lower.tail=TRUE)`
- Right-Tailed: `p-value = pnorm(z, lower.tail=FALSE)`
- Two-Tailed: `p-value = 2 * pnorm( abs(z), lower.tail=FALSE)`

Using `pt` for t distribution

To find the upper tailed p-value for a t statistic of 3 using the t distribution with degrees of freedom equal to 35, you would type

`pt(3, df = 35, lower.tail = FALSE)`

In general, `p-value = pt(t, df = , lower.tail = )`.

- Left-Tailed Tests: `p-value = pt(t, df = n-1, lower.tail=TRUE)`
- Right-Tailed Tests: `p-value = pt(t, df = n-1, lower.tail=FALSE)`
- Two-Tailed Tests: `p-value = 2 * pt( abs(t), df = n-1, lower.tail=FALSE)`

1. Using the normal distribution, find the lower-tailed p-value for $z = -1.4$.

2. Using the t distribution with df=48, find the two tailed p-value for a $t = 2.2$.

3. Compute the p-value for part I.