Descriptive & Inferential Techniques for Studies Involving One Quantitative Variable

Dr Austin R Brown

Kennesaw State University

Introduction

- So far, we have learned about the research process in general in addition to some concepts in probability and hypothesis testing.

- Now, how do we take those general ideas and apply them to real research questions?
  - As we have discussed, the specific tools and techniques we use to answer research questions depend largely in part on the type of variable or variables we are working with!

- Let's first consider an example where we're working with a single quantitative variable coming from a single population.

Introduction

- Suppose I am interested in TikTok usage among KSU undergraduate students. To understand how much students in this population use TikTok, I stood outside of the Student Center during lunchtime hours and asked 500 students how much they estimate they use TikTok per day (in hours). After conversations with people in that demographic, I hypothesize the mean amount of daily time spent using TikTok will be greater than 3 hours per day.

Introduction

- Remember, the research process in general follows a systematic series of steps:

1. Specify the Research Question
2. Specify the Research Hypothesis
3. Specify the Statistical Hypotheses
4. Collect Data
5. Perform Evidence Gathering Techniques
    a. Descriptive (data visualizations and quantitative summaries)
    b. Inferential (hypothesis tests and effect sizes)
6. Draw Contextual Conclusions

Specify the Research Question

1. **Population of Interest:** KSU Undergraduate Students

2. **Variable of Interest:** Daily Time (in hours) Spent on TikTok (quantitative variable)

3. **Characteristic to be Analyzed:** Mean Amount of Daily Time (in hours) Spent on TikTok

4. **General Analytical Method:** Comparison

Specify the Research Hypothesis

- We hypothesize the mean amount of daily time spent using TikTok will be greater than 3 hours per day.

Specify the Statistical Hypotheses

- Remember, the alternative hypothesis is the same as the research hypothesis:

$$H_A : \mu > 3$$

- And the null hypothesis is just the opposite:

$$H_0 : \mu \leq 3$$

Evidence Gathering Techniques: Quantitative Summaries

- To begin to determine if our data provide support for the research hypothesis, it is a good idea to perform some descriptive analysis beginning with calculating appropriate/relevant quantitative summaries of our data.

- When we're working with a single quantitative variable, some pieces of relevant information we'd like to know about our data might be:
  1. What is a central/middle/typical value? (Measures of Central Tendency)
  2. How similar are the observations to one another? (Measures of Variability)
  3. How does one value relate to all of the other observations? (Measures of Position)

- Let's start with the Measures of Central Tendency!

Evidence Gathering Techniques: Quantitative Summaries

- As mentioned, the Measures of Central Tendency let us know what a central or common value is among the data collected.
- While there are several ways of measuring a central or common value, the two most prominently used measures are the **sample mean** and **sample median**.

Evidence Gathering Techniques: Quantitative Summaries

- We have all heard of a mean or average, right? We calculate it using the below formula:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Less technically, we add up all of our values and divide by the total number of values observed (n).

- But how do we interpret this value?

Evidence Gathering Techniques: Quantitative Summaries

- Imagine if I put all 500 TikTok times on little slips of paper and then put them into a hat.
- I ask you to randomly pull a single slip of paper out of the hat and keep it enclosed in your palm. I then ask you to guess what you think the value on the slip of paper is.
- The sample mean is your best guess as to what value is on the slip of paper.
  - How do we calculate this using R?

Evidence Gathering Techniques: Quantitative Summaries

```r
## Step 1: Read in the Data ##

tiktok <- readxl::read_xlsx("TikTok Usage.xlsx")

## Step 2a: Calculate the Sample Mean ##

mean(tiktok$`Daily Time`)
```

```
[1] 3.56534
```

Evidence Gathering Techniques: Quantitative Summaries

- So our best guess as to what a randomly selected TikTok time would be is 3.57 hours.

- What about the sample median? The sample median, also known as the 50th percentile, is the value where half (or 50%) of the observed values are greater than it and half (or 50%) of the observed values are less than it. Using R:

```
## Step 2b: Calculate Sample Median ##

median(tiktok$`Daily Time`)
```

```
[1] 3.59
```

Evidence Gathering Techniques: Quantitative Summaries

- So now our question is: do these measures of central tendency provide support for the research hypothesis?
- Since both values are greater than 3, which is what we hypothesized, we would conclude that these two descriptive values do provide some support for the research hypothesis.
- The degree of support we can determine both on a contextual basis, as well as using a measure of position, which we will discuss in a little bit!

Evidence Gathering Techniques: Quantitative Summaries

- The second piece of information we typically want to know about a quantitative variable is how similar all of the values are to one another.

- We can think of this like throwing 10 darts at a dart board. If all of the darts land close together on the board, we'd conclude our throws were similar.
  - If all the darts land far apart (or even off the board!!), we'd conclude our throws were different.

- In essence, what we're measuring in this case is variability! So how do we quantify variability when we're working with a quantitative variable? We have a couple of methods available to us!

Evidence Gathering Techniques: Quantitative Summaries

- The first measure is the **sample range**, which you've likely heard of. Mathematically:

$$\text{Range} = \text{Maximum Value} - \text{Minimum Value}$$

- In R:

```
max(tiktok$`Daily Time`) - min(tiktok$`Daily Time`)
```

```
[1] 5.58
```

Descriptive Evidence Gathering Techniques: Quantitative Summaries

- But there are problems with relying on the range as the sole measure of variability. Why?

1. It is an inefficient use of data (only looking at two values!)
2. It is not a standardized value (big values of the range or small values of the range don't necessarily imply more or less variability)
3. It is sensitive to outliers (makes sense! It uses the two values most likely to be outliers!)

- So what else is out there?

Evidence Gathering Techniques: Quantitative Summaries

- Instead, we typically use a measure called **sample standard deviation**. What is sample standard deviation? Mathematically:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Okay this may look a little intimidating. So what does it mean contextually?

Evidence Gathering Techniques: Quantitative Summaries

- Remember the slips of paper in the hat? Suppose you randomly draw one again. This time, I ask you how far away from the sample mean you'd guess the value in your hand is (in either the positive or negative direction).

- That's what standard deviation is: the expected distance an observation is away from its mean, in either the positive or negative direction.

Evidence Gathering Techniques: Quantitative Summaries

- The reasons why we like standard deviation are:

1. $s > 0$, which means that bigger values imply more variability
2. It is in the original units being measured which helps us interpret the value more easily.

- How do we calculate this in R?

```
## Step 3: Calculate Sample Standard Deviation ##

sd(tiktok$`Daily Time`)
```

```
[1] 0.9808935
```

Evidence Gathering Techniques: Quantitative Summaries

- The last question we typically have as it relates to quantitative summaries of a quantitative variable is: how does a specific value of our variable relate to all the others?
- In other words, what is its position?
- Here we use a couple of measures:
  1. Z-Scores
  2. Quartiles (special cases of percentiles)

Evidence Gathering Techniques: Quantitative Summaries

- A z-score is the number of standard deviations a particular value is away from its mean. Mathematically:

$$Z_i = \frac{x_i - \bar{x}}{s}$$

- A z-score helps us know if our value, x_i, is typical (meaning $-2 \leq z_i \leq 2$) or atypical (meaning either $z_i < -2$ or $z_i > 2$).
  - "Typical" values are relatively close to their mean. "Atypical" values are relatively far away from their mean.

- Let's consider an example. Is a person who spends 2 hours per day considered typical or atypical? Let's calculate the z-score for 2 hours in R and find out!

Evidence Gathering Techniques: Quantitative Summaries

```
(2 - mean(tiktok$`Daily Time`))/sd(tiktok$`Daily Time`)
```

```
[1] -1.595831
```

- Since $z = -1.60$, we would conclude that someone who spends 2 hours per day on TikTok is a typical user, at least in the context of this sample.

- What about somebody who uses TikTok for 6 hours per day?

Evidence Gathering Techniques: Quantitative Summaries

```
(6 - mean(tiktok$`Daily Time`))/sd(tiktok$`Daily Time`)
```

```
[1] 2.482084
```

- Since $z = 2.48$, we would conclude that somebody who spends 6 hours per day on TikTok is an atypical user, at least in the context of this sample.

Evidence Gathering Techniques: Quantitative Summaries

- The other measure of position we are generally interested in are the quartiles, of which there are three:

1. Q1 or the 25th percentile
2. Q2 or the 50th percentile
3. Q3 or the 75th percentile

- Remember, percentiles are the proportion of values less than our value of interest. Let's calculate these (and all of the other measures we've discussed so far besides z-scores) in R using the `rstatix` package.

Evidence Gathering Techniques: Quantitative Summaries

```
install.packages('rstatix')
tiktok |>
  rstatix::get_summary_stats(`Daily Time`,type='full') |>
  dplyr::select(mean,median,sd,q1,q3)
```

```
# A tibble: 1 x 5
   mean median    sd    q1    q3
  <dbl>  <dbl> <dbl> <dbl> <dbl>
1  3.56   3.59 0.981  2.83  4.27
```

Evidence Gathering Techniques: Quantitative Summaries

- Now the question is: do Q1 and Q3 provide support for the research hypothesis?
- To make this determination, let's draw a line and label Q1, Q3, and our cutpoint specified in the research hypothesis (and statistical hypotheses) of 3 hours:

Evidence Gathering Techniques: Quantitative Summaries