MA206 Lesson 8 - Generalization

What is μ?
μ indicates the population mean, it is a parameter.

What is $\bar{x}$ and how do we calculate it using R?
$\bar{x}$ is the sample mean, the statistic observed from quantitative data.
We find it using the command mean(data)

What is σ?
σ indicates the population standard deviation, it is a parameter.

What is s and how do we calculate it using R?
s is the sample standard deviation, the statistic observed from quantitative data.
We find it using the comand sd(data)

Given sufficiently large samples of a population, what does the Central Limit Theorem state the standard deviation of the mean ($\bar{x}$) will converge to?
$\text{SD}(\bar{x}) = \frac{\sigma}{\sqrt{n}}$, where we can estimate σ with s

Review: How is the standardized statistic for a **single proportion** calculated?
z $= \frac{\hat{p} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$

How is the standardized statistic for a **single mean** calculated?
t $= \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

What are the validity conditions for a single mean quantitative variable to use a theory-based approach for strength of evidence?
The distribution must be symmetric **OR** must have at least 20 observations and not be heavily skewed

What is an alternative measure for centrality if a distribution is heavily skewed?
The Median

What is the R code to calculate a p-value from a t-distribution for a single mean quantitative variable?
$H_a : \mu >$ 　 1 - pt(t, n-1)
$H_a : \mu <$ 　 pt(t, n-1)
$H_a : \mu \neq$ 　 2*(1 - pt(abs(t), n-1))

Load the provided ACFT1.csv Dataset. This constitutes data collected on or about 2019 (pre-Covid) from randomly selected cadets in 1st Regiment on physical fitness and includes APFT, IOCT, and ACFT performance. Researchers were curious to know how cadets performed physically on various tests, specifically if they were above average. We will conduct analysis on all three tests.

library(tidyverse)

ACFT <- read_csv("ACFT1.csv")

(Alternatively, as this dataset is loaded online as a .csv file, we can run the following line of code).

```
ACFT <- read_csv("https://raw.githubusercontent.com/rslasater82/MA206Datasets/main/ACFT.csv")
```

We can begin exploring our available data to see what variables are captured.

head(ACFT)
colnames(ACFT)

1. What are the observational units in this study?
Each individual cadet

2. What are some categorical and quantitative variables of interest? List and classify some of interest.
Some variation, but a categorical could be "Sex", quantitative could include "age, IOCT_Score, AFCT_score, etc.

APFT

The APFT is scored with 3 events on a 100 point scale, where 60 is passing in each of the three events (pushup, situp, and 2 mile run). If a Soldier scores at least 100 points on each event, they can push past 300 to an "extended scale," but if any event falls beneath the 100 point threshold then the "additional" points are lost. It has been suggested that the Corps of Cadets had an average APFT score of 270, and we want to assess if this is true.

Step 1: Ask a Research Question

A.3) What is the research question?
Is the Corps of Cadets average APFT score equal to 270 or not?

A.4) Based on these questions, what is the population of interest?
The population of interest is the Corps of Cadets

Step 2: Design a Study and Collect Data

A.5) What is our null and alternate hypothesis, stated in both words and symbols?
$H_0 : \mu = 270$. The true mean APFT score for the Corps of Cadets is 270.

$H_a : \mu \neq 270$. The true mean APFT score for the Corps of Cadets is not 270.

A.6) What is our variable of interest? Is it categorical or quantitative?
Our variable of interest is APFT_score, which is a quantitative variable.

A quantitative variable will conduct a single-mean test with the t statistic.

Step 3: Explore the Data

(Helpful R code, courtesy of the TidyVerse Tutorial. Know where to find this for your own reference.)

```
ACFT %>%
  ggplot(aes(x = APFT_score)) +
  geom_histogram()+
  labs(x = "Overall Score", y="Count", title="Cadet APFT Scores")

ACFT %>%
   summarize(
       median = median(APFT_score),
       mean = mean(APFT_score),
       s = sd(APFT_score),
       n = n()
   )
```

A.8) Use R to create a histogram of the APFT scores and describe the shape, center, variability, and any unusual observations.

The shape appears to have two peaks split at the 300 cutoff point, which makes sense due to the grading rules. It has a mean at 304 and median at 296, indicating some right skew. Our standard deviation is 29.1, with values ranging from about 225 up to 375.

A.9) Calculate the mean, median, standard deviation, and sample size of the APFT scores. Include the proper notation as appropriate.

From the code above, we find that

mean $= \bar{x} = 304$

median $= 296$

standard deviation $= s = 29.1$

sample size $= n = 293$

Step 4: Draw Inferences Beyond the Data

A.10) Have we met our validity conditions to use theoretical methods? Why or why not?

Yes, for quantitative data we have more than 20 observations and the data is not strongly skewed.

$t = \frac{304-270}{\frac{29.1}{\sqrt{293}}} = 19.99953$

$2 * (1 - pt(abs(19.99953), 292)) = 0$

A.13) Comment on the strength of evidence as it applies to our null and alternate hypotheses.

With a p-value of (computationally) 0, we have very strong evidence that the true mean of corps of cadets APFT scores is not equal to 270.

Step 5: Formulate Conclusions

A.14) Do you feel comfortable generalizing the results of your analysis to all Army cadets (USMA, ROTC, G2G)? Explain your reasoning.

No, those observational units did not have an equal chance of being selected and so we cannot generalize these results.

A.15) Do you feel comfortable generalizing the results of your analysis to the Corps of Cadets? Explain your reasoning.

No, observational units from different regiments did not have an equal chance of being selected and so we cannot generalize these results to populations outside of 1st REG.

A.16) What population could you generalize your results to? Explain your reasoning.

We can generalize these results to the 1st Regiment population of cadets, as those cadets had an equal probability of being sampled for this study.

A.17) How confident are you to say that we have proven the alternate hypothesis?

We never say we proved or disproved a hypothesis, only that we have very strong evidence against the null hypothesis.

Step 6: Look Back and Ahead

A.18) Suggest how you might redesign the experiment to allow you to draw a more broad conclusion.

To generalize to the USMA Corps of Cadets, we would randomly sample from the entire population of the corps of cadets instead of just one regiment. To generalize to all cadets, we would need to be able to randomly sample from the Corps of Cadets, ROTC, and G2G populations with equal probability per cadet.

A.19) Suppose your research question had an alternate hypothesis for a one-sided test. That is, do USMA cadets, on average, perform better than 270? Report your new alternate hypothesis, p-value and the significance of your findings. Is this surprising?

$H_a : \mu > 270; \quad 1 - pt(19.99953, 292)) = 0$

This is not surprising, it should be half the two-sided test, and $\frac{1}{2}$ of 0 is 0.

A.20) Suppose you wanted to assess the data for only males or only females. Repeat your original ($\neq$) analysis with the proper subsets and report on your findings.

MALES: $t = \frac{301-270}{\frac{28.6}{\sqrt{234}}} = 16.58073$, p-value ≈ 0

FEMALES: $t = \frac{313-270}{\frac{29.4}{\sqrt{59}}} = 11.23433$, p-value $= 2.22e^{-16}$

IOCT

The IOCT is a timed event and times can be converted to a 1000 point scale for calculation into the Physical Program Score (Cumulative) where 900 points is the cutoff for an A, 800 for a B, and so on with a minimum passing score of 680. It has been suggested that DPE curves their scoring to keep the average at 800, but we want to assess if the current average is different than 800.

Step 1: Ask a Research Question

B.3) What is the research question?
Is the Corps of Cadets average IOCT score equal to 800 or not?

B.4) Based on these questions, what is the population of interest?
The population of interest is the Corps of Cadets

Step 2: Design a Study and Collect Data

B.5) What is our null and alternate hypothesis, stated in both words and symbols?
$H_0 : \mu = 800$. The true mean IOCT score for the Corps of Cadets is 800.
$H_a : \mu \neq 800$. The true mean IOCT score for the Corps of Cadets is not 800.

B.6) What is our variable of interest? Is it categorical or quantitative?
Our variable of interest is IOCT_Score, which is a quantitative variable.

B.7) Based on the classification in **6**, what theory-based test will we conduct? (hint: z or t)
A quantitative variable will conduct a single-mean test with the t statistic.

Step 3: Explore the Data

(Helpful R code, courtesy of the TidyVerse Tutorial. Know where to find this for your own reference.)

```
ACFT %>%
  ggplot(aes(x = IOCT_Score)) +
  geom_histogram()+
  labs(x = "Overall Score", y="Count", title="Cadet IOCT Scores")

ACFT %>%
    summarize(
        median = median(IOCT_Score),
        mean = mean(IOCT_Score),
        s = sd(IOCT_Score),
        n = n()
    )
```

B.8) Use R to create a histogram of the IOCT scores and describe the shape, center, variability, and any unusual observations.
The shape appears to be evenly split between 750 and 950, with smaller tails from 680 to 1000. There is a larger peak at 780ish and centered about 800 or 850.

B.9) Calculate the mean, median, standard deviation, and sample size of the IOCT scores. Include the proper notation as appropriate.

From the code above, we find that mean $= \bar{x} = 829$

median $= 824$

standard deviation $= s = 83.9$

sample size $= n = 293$

Step 4: Draw Inferences Beyond the Data

B.10) Have we met our validity conditions to use theoretical methods? Why or why not?

Yes, for quantitative data we have more than 20 observations and the data is not strongly skewed.

B.11) Assume we met our validity conditions. Using theory, calculate the appropriate standardized statistic.

$t = \frac{824-800}{\frac{83.9}{\sqrt{293}}} = 4.89647$

B.12) Calculate the appropriate p-value using the theoretical method and standardized statistic from **11**.

$2 * (1 - pt(abs(4.89647), 292)) = 1.617708e^{-6}$

B.13) Comment on the strength of evidence as it applies to our null and alternate hypotheses.

With a p-value of 0.000006, we have very strong evidence that the true mean of corps of cadets IOCT scores is not equal to 800.

Step 5: Formulate Conclusions

B.14) Do you feel comfortable generalizing the results of your analysis to the Corps of Cadets? Explain your reasoning.

No, observational units from different regiments did not have an equal chance of being selected and so we cannot generalize these results to populations outside of 1st REG.

B.15) What population could you generalize your results to? Explain your reasoning.

We can generalize these results to the 1st Regiment population of cadets, as those cadets had an equal probability of being sampled for this study.

B.16) How confident are you to say that we have proven the alternate hypothesis?

We never say we proved or disproved a hypothesis, only that we have strong evidence against the null hypothesis.

Step 6: Look Back and Ahead

B.17) Suggest how you might redesign the experiment to allow you to draw a more broad conclusion.

To generalize to the corps of cadets, we would randomly sample from the entire population of the corps of cadets instead of just one regiment.

$H_a : \mu > 800; \quad 1 - pt(4.89647, 292)) = 8.088539e^{-7}$

This is not surprising, it should be half the two-sided test, and $\frac{1}{2}$ of .0000016 is 0.0000008

MALES: $t = \frac{824 - 800}{\frac{85.1}{\sqrt{234}}} = 4.314094$, p-value $= 2.37e^{-5}$

FEMALES: $t = \frac{847 - 800}{\frac{77.1}{\sqrt{59}}} = 4.682411$, p-value $= 1.75e^{-5}$

ACFT

The ACFT is scored with 6 events on a 100 point scale, where 60 is passing in each of the six events. If the arbitrary "average" score is 80 in each event (480 points), is the Corps of Cadets "average" at the ACFT?

Step 1: Ask a Research Question

C.3) What is the research question?
Is the Corps of Cadets average ACFT score equal to 480 or not?

C.4) Based on these questions, what is the population of interest?
The population of interest is the Corps of Cadets

Step 2: Design a Study and Collect Data

C.5) What is our null and alternate hypothesis, stated in both words and symbols?
$H_0 : \mu = 480$. The true mean ACFT score for the Corps of Cadets is 480.
$H_a : \mu \neq 480$. The true mean ACFT score for the Corps of Cadets is not 480.

C.6) What is our variable of interest? Is it categorical or quantitative?
Our variable of interest is ACFT_score, which is a quantitative variable.

C.7) Based on the classification in **6**, what theory-based test will we conduct? (hint: z or t)
A quantitative variable will conduct a single-mean test with the t statistic.

Step 3: Explore the Data

(Helpful R code, courtesy of the TidyVerse Tutorial. Know where to find this for your own reference.)

```
ACFT %>%
  ggplot(aes(x = ACFT_score)) +
  geom_histogram()+
  labs(x = "Overall Score", y="Count", title="Cadet ACFT Scores")

ACFT %>%
    summarize(
        median = median(ACFT_score),
        mean = mean(APFT_score),
        s = sd(APFT_score),
        n = n()
    )
```

C.8) Use R to create a histogram of the ACFT scores and describe the shape, center, variability, and any unusual observations.
The shape is left skewed with a peak at about 500, centered between 450 and 500. Values range from nearly 600 down to 250 with no obvious outliers.

C.9) Calculate the mean, median, standard deviation, and sample size of the ACFT scores. Include the proper notation as appropriate.

From the code above, we find that mean $= \mu = 487$

median $= 505$

standard deviation $= s = 64.7$

sample size $= n = 293$

Step 4: Draw Inferences Beyond the Data

C.10) Have we met our validity conditions to use theoretical methods? Why or why not?

Yes, for quantitative data we have more than 20 observations and the data is not **strongly** skewed.

C.11) Assume we met our validity conditions. Using theory, calculate the appropriate standardized statistic.

$t = \frac{487-480}{\frac{64.7}{\sqrt{293}}} = 1.851943$

C.12) Calculate the appropriate p-value using the theoretical method and standardized statistic from **11**.

$2 * (1 - pt(abs(1.851943), 292)) = 0.06504327$

C.13) Comment on the strength of evidence as it applies to our null and alternate hypotheses.

With a p-value of 0.065, we have moderate evidence that the true mean of corps of cadets ACFT scores is not equal to 480.

Step 5: Formulate Conclusions

C.14) Do you feel comfortable generalizing the results of your analysis to all Army cadets (USMA, ROTC, G2G)? Explain your reasoning.

No, those observational units did not have an equal chance of being selected and so we cannot generalize these results.

C.15) Do you feel comfortable generalizing the results of your analysis to the Corps of Cadets? Explain your reasoning.

No, observational units from different regiments did not have an equal chance of being selected and so we cannot generalize these results to populations outside of 1st REG.

C.16) What population could you generalize your results to? Explain your reasoning.

We can generalize these results to the 1st Regiment population of cadets, as those cadets had an equal probability of being sampled for this study.

C.17) How confident are you to say that we have proven the alternate hypothesis?

We never say we proved or disproved a hypothesis, only that we have moderate evidence against the null hypothesis.

Step 6: Look Back and Ahead

C.18) Suggest how you might redesign the experiment to allow you to draw a more broad conclusion.

To generalize to the corps of cadets, we would randomly sample from the entire population of the corps of cadets instead of just one regiment. To generalize to all cadets, we would need to be able to randomly sample from the Corps of Cadets, ROTC, and G2G populations with equal probability per cadet.

C.19) Suppose your research question had an alternate hypothesis for a one-sided test. Report your new new alternate hypothesis, p-value and the significance of your findings. Is this surprising?

$H_a : \mu > 480; \quad 1 - pt(1.851943, 292)) = 0.03252164$

This is not surprising, it should be half the two-sided test, and $\frac{1}{2}$ of 0.065 is 0.0325

C.20) Suppose you wanted to assess the data for only males or only females. Repeat your original ($\neq$) analysis with the proper subsets and report on your findings.

MALES: $t = \frac{512-480}{\frac{33.9}{\sqrt{234}}} = 14.4397$, p-value ≈ 0

FEMALES: $t = \frac{386-480}{\frac{58.9}{\sqrt{59}}} = -12.25853$, p-value ≈ 0

C.21) What do the signs of the t statistic tell you that the p-value alone does not in this case?

Here, it tells us two things. One, the evidence is strong that the male mean is larger than the null hypothesis (positive t) while the evidence is strong that the female mean is smaller than the null hypothesis (negative t). Furthermore, despite both p-values being computationally 0, we can see that the distance of the male scores is farther away, in terms of standard deviations, than the female scores.