Class 16

DATA1220-55, Fall 2024

Sarah E. Grabinski

2024-10-04

Recap: The Central Limit Theorem

- Properties of the normal distribution let us calculate the probability of observing a given value or range of values
- ***The Central Limit Theorem***: The probability distribution of the sample means from multiple samples of the same size n from the same population approximates a normal distribution as n increases (i.e. the *sampling distribution*)
- The ***sampling distribution*** provides point estimates and confidence intervals for population parameters

Recap: Point Estimates & Confidence Intervals

- A ***point estimate*** describes the ***location*** of an estimate or distribution
- A ***confidence interval*** describes the ***scale*** or ***precision*** of an estimate or distribution
- The ***confidence threshold*** or ***confidence level*** describes our uncertainty regarding the ***accuracy*** of our estimates

Recap: Z-Scores & Confidence Intervals

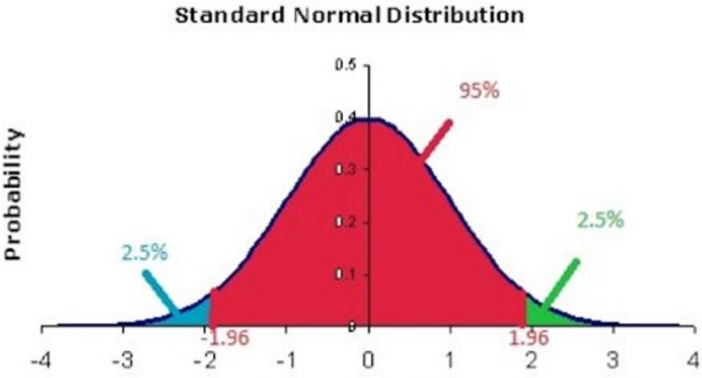


Figure 1: We use Z-Scores from the standard normal distribution to calculate the boundaries of our confidence interval.

Recap: Assumptions

1. A random process follows a known distribution which we can use to model that process and draw inferences about our population.

Recap: Assumptions

1. A random process follows a known distribution which we can use to model that process and draw inferences about our population.
2. Your data is ***reliable***: your sample statistics are ***reliable*** estimations of the sample population's distribution.

Recap: Assumptions

1. A random process follows a known distribution which we can use to model that process and draw inferences about our population.
2. Your data is ***reliable***: your sample statistics are ***reliable*** estimations of the sample population's distribution.
3. Your data is ***valid***: your sampling distribution, based on your sample statistics, is a ***valid*** estimation of the study population distribution.

Recap: Assumptions

1. A random process follows a known distribution which we can use to model that process and draw inferences about our population.
2. Your data is ***reliable***: your sample statistics are ***reliable*** estimations of the sample population's distribution.
3. Your data is ***valid***: your sampling distribution, based on your sample statistics, is a ***valid*** estimation of the study population distribution.
4. Your data is ***generalizable***: your sampling distribution for your study population is ***generalizable*** as the sampling distribution for your target population

Statistical Inference and Hypothesis Testing

- We use sample statistics to describe sample populations and estimate the parameters of the study population's sampling distribution
- We also describe the variability of our measure and quantify our uncertainty regarding our estimate
- We use the overlap between theoretical distributions to decide whether any differences are meaningful

Example: Hypothesis Testing (Proportion)

Deshaun Watson, the quarterback for the Cleveland Browns, has made 176 pass attempts in the 2024 NFL season, 106 of which were completed (60.2%). How do we know if that's good or not?

Example: Hypothesis Testing (Proportion)

Deshaun Watson, the quarterback for the Cleveland Browns, has made 176 pass attempts in the 2024 NFL season, 106 of which were completed (60.2%). How do we know if that's good or not?

Among the 38 quarterbacks in the NFL who have attempted at least 20 passes, the average completion rate is 65.3%.

Hypothesis Testing Framework

- $\mathbf{H_0}$: The "Null" Hypothesis
  - Represents a position of skepticism, *nothing* is happening here
  - "There is *not* an association between process A and B"
- $\mathbf{H_A}$: The "Alternative" Hypothesis
  - The complement of H_0, *something* is happening here
  - "There *is* an association between process A and B"

Conducting a hypothesis test

- Begin by *assuming* H_0 is the "true" state

Conducting a hypothesis test

- Begin by *assuming* H_0 is the "true" state
- Calculate *the probability that you would see results as extreme or more extreme* than what you saw in your study, assuming the distribution under H_0

Conducting a hypothesis test

- Begin by *assuming* H_0 is the "true" state
- Calculate *the probability that you would see results as extreme or more extreme* than what you saw in your study, assuming the distribution under H_0
- The lower the probability, the less likely it is that we would see these results if H_0 was the "true" state of our population

Conducting a hypothesis test

- Begin by *assuming* H_0 is the "true" state
- Calculate *the probability that you would see results as extreme or more extreme* than what you saw in your study, assuming the distribution under H_0
- The lower the probability, the less likely it is that we would see these results if H_0 was the "true" state of our population
- If the probability is sufficiently low, we *reject* $\mathbf{H_0}$ and *accept* $\mathbf{H_A}$

Example: Hypothesis Testing (Proportion)

Research question: Is Deshaun Watson's pass completion rate below average for the 2024 NFL season so far?

- H_0: Deshaun Watson has an average pass completion rate for the 2024 NFL season ($\hat{p}_{\mathrm{DW}} \approx p_{\mathrm{NFL}}$)
- H_A: Deshaun Watson has a below-average pass completion rate for the 2024 NFL season ($\hat{p}_{\mathrm{DW}} \neq p_{\mathrm{NFL}}$)

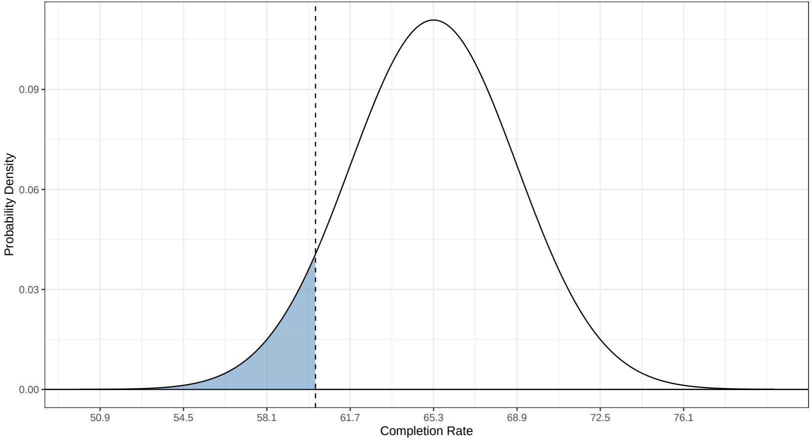
Example: Hypothesis Testing (Proportion)

We can use the NFL's average (65.3%) plus the sample size ($n = 176$ passes) to construct a sampling distribution $\hat{p} \sim N(p = 65.3, SE = 3.6)$ for the average NFL quarterback's pass completion rate.

$$
\begin{aligned}
SE_p &= \sqrt{\frac{p(1-p)}{n}} \\
&= \sqrt{\frac{0.653(1-0.653)}{176}} \\
&= 0.036
\end{aligned}
$$

Example: Hypothesis Testing (Proportion)

Assuming Deshaun Watson is an average NFL quarterback, what is the probability he would have a completion rate of 60.2% *or less* over the last 176 passes?



Example: Hypothesis Testing (Proportion)

The probability that Deshaun Watson would have a completion rate of 0.602 or worse, *assuming he is an average NFL quarterback*, is 0.078.

```
pnorm(0.602, mean = 0.653, sd = 0.036)
```

```
[1] 0.0782902
```

Example: Hypothesis Testing (Proportion)

The probability that Deshaun Watson would have a completion rate of 0.602 or worse, *assuming he is an average NFL quarterback*, is 0.078.

```
pnorm(0.602, mean = 0.653, sd = 0.036)
```

```
[1] 0.0782902
```

Is this situation unlikely enough that we can reject our null hypothesis H_0?

Significance Level/Threshold

- α is also called the ***significance level***
- The probability below which you will reject the null hypothesis
- Predetermined before doing hypothesis test (often $p < 0.05$)
- Also the probability of rejecting the null hypothesis when H_0 is true (i.e. ***Type I Error*** or ***false positive rate***)

Decision Errors

In a perfect world, we will only reject H_0 when H_A is true, and we will always fail to reject H_0 when H_0 is true.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_A true		✓

Decision Errors

When we reject H_0 when H_0 is “true” (i.e. *false positive*), it is called a ***Type I Error***.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	Type 1 Error
	H_A true		✓

Decision Errors

When we fail to reject H_0 when H_A is "true" (i.e. *false negative*), it is called a ***Type II Error***.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	Type 1 Error
	H_A true	Type 2 Error	✓

Decision Errors

- $\alpha = P(\text{FalsePositive})$
- $\beta = P(\text{FalseNegative})$

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	α
	H_A true	β	✓

Decision Errors

- $1 - \alpha = \text{ConfidenceLevel}$
- $1 - \beta = \text{Power}$

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$1 - \alpha$	α
	H_A true	β	$1 - \beta$

Practice: Decision Errors

The justice system can be thought of like a hypothesis test. We *assume* the defendant is innocent (H_0), until we have sufficient evidence to reject the null hypothesis and accept the alternate hypothesis (H_A) that the defendant is guilty.

Practice: Decision Errors

In a criminal trial, what type of error is committed when the jury finds the defendant guilty, when they were innocent?

What are the hypotheses?

Practice: Decision Errors

In a criminal trial, what type of error is committed when the jury finds the defendant guilty, when they were innocent?

What are the hypotheses?

H_0: Defendant is not guilty

H_A: Defendant is guilty

Practice: Decision Errors

In a criminal trial, what type of error is committed when the jury finds the defendant guilty, when they were innocent?

What are the hypotheses?

H_0: Defendant is not guilty

H_A: Defendant is guilty

Type I Error

Practice: Decision Errors

In a criminal trial, what type of error is committed when the jury finds the defendant not guilty, when they were guilty?

What are the hypotheses?

Practice: Decision Errors

In a criminal trial, what type of error is committed when the jury finds the defendant not guilty, when they were guilty?

What are the hypotheses?

H_0: Defendant is not guilty

H_A: Defendant is guilty

Practice: Decision Errors

In a criminal trial, what type of error is committed when the jury finds the defendant not guilty, when they were guilty?

What are the hypotheses?

H_0: Defendant is not guilty

H_A: Defendant is guilty

Type II Error

Decision Error Trade-Offs

- Often, reducing the false positive rate increases the false negative rate (and vice versa)
- There are different costs to false negatives and false positives

Decision Error Trade-Offs

- Often, reducing the false positive rate increases the false negative rate (and vice versa)
- There are different costs to false negatives and false positives

We'd rather let a guilty person go free than an innocent person go to jail (false negative preferable to false positive)

Decision Error Trade-Offs

- Often, reducing the false positive rate increases the false negative rate (and vice versa)
- There are different costs to false negatives and false positives

We'd rather let a guilty person go free than an innocent person go to jail (false negative preferable to false positive)

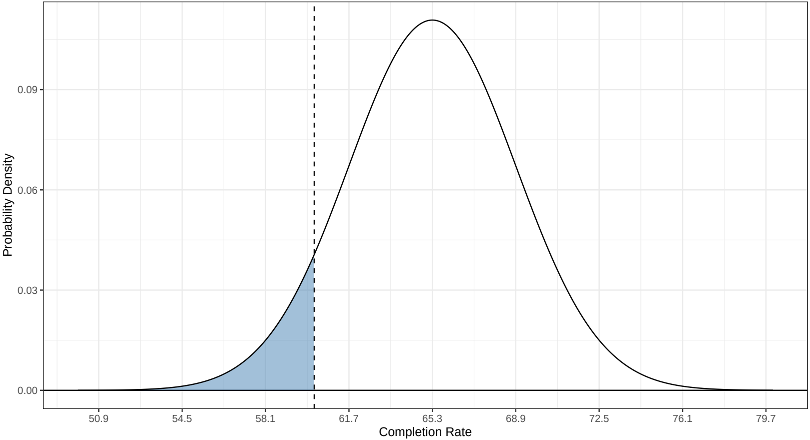
Metal detectors are overly sensitive so weapons aren't missed during scans (false positive preferable to false negative)

Test Statistics

- When conducting a hypothesis test, you calculate a ***test statistic*** to assess how "extreme" your observed result is compared to the reference distribution
- For normal sampling distributions (means & proportions), the test statistic is the $Z - Score$
- Remember: $Z = \frac{\bar{x}-\mu}{\sigma}$

Example: Hypothesis Testing (Proportion)

Assuming Deshaun Watson is an average NFL quarterback, what is the probability he would have a completion rate of 60.2% *or less* over the last 176 passes?



Example: Hypothesis Testing (Proportion)

Assuming $\hat{p} \sim N(p = 65.3, SE = 3.6)$...

$$\begin{aligned} Z &= \frac{\hat{p} - p}{SE} \\ &= \frac{60.2 - 65.3}{3.6} \\ &= -1.42 \end{aligned}$$

Example: Hypothesis Testing (Proportion)

Assuming $\hat{p} \sim N(p = 65.3, SE = 3.6)$...

$$\begin{aligned} Z &= \frac{\hat{p} - p}{SE} \\ &= \frac{60.2 - 65.3}{3.6} \\ &= -1.42 \end{aligned}$$

Is 1.42 standard errors below the hypothesized completion rate unusual enough to reject H_0?

P-Values

A ***p-value*** is the probability of a theoretical sample having a test statistic equal to or more extreme than the one you observed, *assuming that the reference distribution is "true"*.

P-Values

A ***p-value*** is the probability of a theoretical sample having a test statistic equal to or more extreme than the one you observed, *assuming that the reference distribution is "true"*.

- If the p-value is *low*, you were very *unlikely* to have observed the data that you did if the null hypothesis were true.

P-Values

A ***p-value*** is the probability of a theoretical sample having a test statistic equal to or more extreme than the one you observed, *assuming that the reference distribution is "true"*.

- If the p-value is *low*, you were very *unlikely* to have observed the data that you did if the null hypothesis were true.
- If the p-value is *high*, you were very *likely* to have observed the data that you did if the null hypothesis were true.

P-Values

- If the p-value is below our pre-determined ***significance level*** α, we *reject* H_0 and *accept* the alternative hypothesis H_A.

P-Values

- If the p-value is below our pre-determined ***significance level*** α, we *reject* H_0 and *accept* the alternative hypothesis H_A.
- *NEVER ACCEPT* H_0!! Always *fail to reject* H_0.

P-Values

- If the p-value is below our pre-determined ***significance level*** α, we *reject* H_0 and *accept* the alternative hypothesis H_A.
- *NEVER ACCEPT* H_0!! Always *fail to reject* H_0.
- The p-value is ***NOT*** the probability that H_0 is true or that the data were produced by chance alone

One-Sided Hypothesis Tests

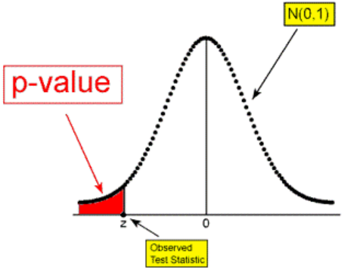


Figure 2: The probability of a test statistic *less than* the one you observed.

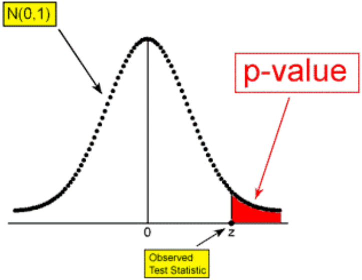


Figure 3: The probability of a test statistic *greater than* the one you observed.

Example

If our test statistic is $Z = -1.42$, what is the p-value for $Z \leq -1.42$?

```
pnorm(-1.42, mean = 0, sd = 1)
```

```
[1] 0.07780384
```

Example

If our test statistic is $Z = -1.42$, what is the p-value for $Z \leq -1.42$?

```
pnorm(-1.42, mean = 0, sd = 1)
```

```
[1] 0.07780384
```

If $\alpha = 0.05$, then $p > \alpha$ and we fail to reject the null hypothesis $H_0 : \hat{p} = 0.653$. There is insufficient evidence that Deshaun Watson has a below-average completion rate.

Two-Sided Hypothesis Tests

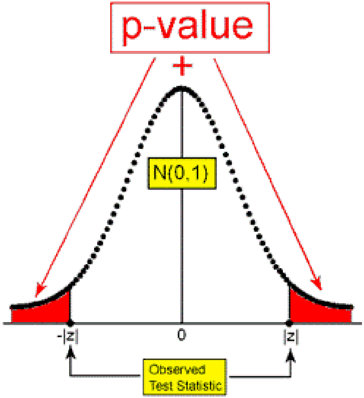


Figure 4: The probability of a test statistic more extreme (greater OR lesser than) the one you observed.

Example: Hypothesis Testing (Proportion)

If our test statistic is $Z = -1.42$, what is the p-value for the null hypothesis $\widehat{p} = 65.3$ $(|Z| \geq 1.42)$?

```
pnorm(-1.42, mean = 0, sd = 1) * 2
```

```
[1] 0.1556077
```

Example: Hypothesis Testing (Proportion)

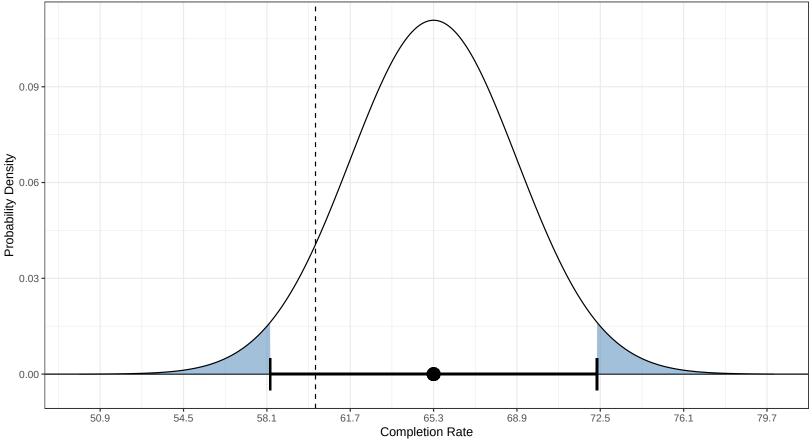
If our test statistic is $Z = -1.42$, what is the p-value for the null hypothesis $\hat{p} = 65.3$ ($|Z| \geq 1.42$)?

```
pnorm(-1.42, mean = 0, sd = 1) * 2
```

```
[1] 0.1556077
```

If $\alpha = 0.05$, then $p > \alpha$ and we fail to reject the null hypothesis $H_0 : \hat{p} = 0.653$. There is insufficient evidence that Deshaun Watson has a below-average completion rate.

Example: Hypothesis Testing (Proportion)



Checking Assumptions

- Were the observations *independent* of each other?

Checking Assumptions

- Were the observations *independent* of each other?
- Do the observations come from *identical distributions*?

Checking Assumptions

- Were the observations *independent* of each other?
- Do the observations come from *identical distributions*?
- Do we have a *sufficient sample size* to assume normality?

Checking Assumptions

- Were the observations *independent* of each other?
- Do the observations come from *identical distributions*?
- Do we have a *sufficient sample size* to assume normality?
- Is a normal distribution the *appropriate model* for this kind of data?

Example: Hypothesis Testing (Mean)

Deshaun Watson is averaging 4.84 yards per pass attempt in the 2024 NFL season. How do we know if that's good or not?

Example: Hypothesis Testing (Mean)

Deshaun Watson is averaging 4.84 yards per pass attempt in the 2024 NFL season. How do we know if that's good or not?

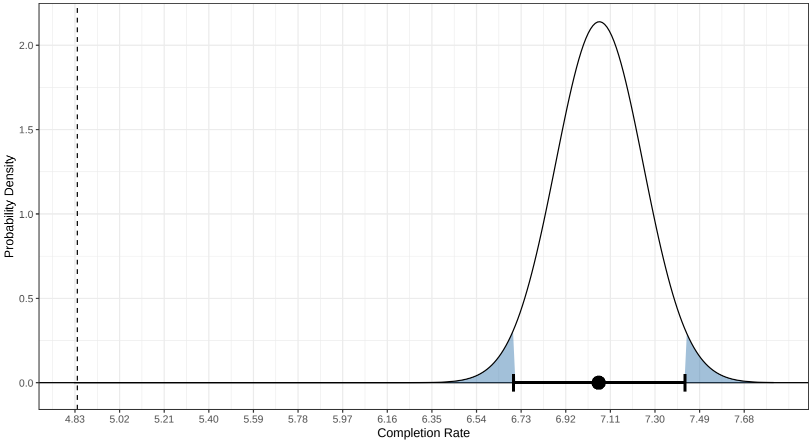
Among the 38 quarterbacks in the NFL who have attempted at least 20 passes, the average yards per pass attempt approximates the distribution $\bar{x} \sim N(\mu = 7.06, SE = 0.19)$.

Example: Hypothesis Testing (Mean)

Research question: Is Deshaun Watson's average yards per pass attempt average for the 2024 NFL season so far?

- H_0: Deshaun Watson has an average yards per pass attempt for the 2024 NFL season ($\bar{x}_{\text{DW}} \approx \mu_{\text{NFL}}$)
- H_A: Deshaun Watson does not have an average yards per pass attempt for the 2024 NFL season ($\bar{x}_{\text{DW}} \neq \mu_{\text{NFL}}$)

Example: Hypothesis Testing (Mean)



Example: Hypothesis Testing (Mean)

$$
\begin{aligned}
Z &= \frac{\bar{x} - \mu}{SE} \\
&= \frac{4.84 - 7.06}{0.19} \\
&= -11.68
\end{aligned}
$$

Example: Hypothesis Testing (Mean)

$$
\begin{aligned}
Z &= \frac{\bar{x} - \mu}{SE} \\
&= \frac{4.84 - 7.06}{0.19} \\
&= -11.68
\end{aligned}
$$

Deshaun Watson's average is 11.68 SD below the hypothesized mean. Is that unusual?

Example: Hypothesis Testing (Mean)

If our test statistic is $Z = -11.68$, what is the p-value for $|Z| \geq 11.68$?

```
pnorm(-11.68, mean = 7.06, sd = 0.19) * 2
```

```
[1] 0
```

Example: Hypothesis Testing (Mean)

If our test statistic is $Z = -11.68$, what is the p-value for $|Z| \geq 11.68$?

```
pnorm(-11.68, mean = 7.06, sd = 0.19) * 2
```

```
[1] 0
```

The probability that Deshaun Watson's average yards would be 4.84 *assuming he is an average quarterback* is $p < 0.05$.

Example: Hypothesis Testing (Mean)

- The p-value is extremely low, so we *reject* H_0 and *accept* H_A that Deshaun Watson does not have an average yards per pass attempt.
- This data provides convincing evidence that Deshaun Watson is performing below average.