Week 9: The Process of Statistical Tests

Jessie Yeung

STA 220

Winter 2024

Overview

- We will be discussing statistical testing (hypothesis testing)
- Topics for this week
  - Basics of Hypothesis Testing
  - Tests about proportions
  - Tests about means
- This content corresponds with Module 8

Motivating Examples

What is a “statistical test”?

- You’ve collected data for your study

- Using the data, you want to answer the question: *Does a particular result exist in this data?*

- **Statistical Test**: A technique that uses statistical theory to make judgement calls about results or features in that exist our data

- A statistical test can either:
  - Provide evidence that a particular result is present OR
  - Tell you that there is no evidence of that result

Motivating Example 1

- John believes that he is bad at basketball
  - This would be considered our *hypothesis*
- Over the next year, the following things happen
  1. John makes it onto his school's varsity basketball team
  2. He makes the starting lineup
  3. He averages 30 points per game, which is high
  - These events would be considered our *data*
- The events that occurred (the data) do not support the initial hypothesis. The data contradicts the hypothesis
- We would say that the data is evidence against the hypothesis

General Structure of a Hypothesis Test

- This basketball example goes through the general framework of a statistical test
- You have a hypothesis.
- Then, you collect some data.
- You check whether the data supports the hypothesis. We either reject the hypothesis or we don’t reject the hypothesis.

Motivating Example 2

- You are provided with a coin, and you are told that the coin is fair.
  - The hypothesis is that the coin is fair
- You then collect some data by flipping the coin several times. Here are the results after flipping it 100 times:
  - Number of Heads: 81
  - Number of Tails: 19
- We know that if the coin was fair, the probability of flipping this many Heads is very small. We can even calculate this probability.
- Since this data is unlikely for a fair coin, we would say that the data provides evidence against the hypothesis. We would want to reject the hypothesis that the coin is fair

Motivating Example 2

- In the example above, it seems obvious that a fair coin is very unlikely to provide us with 81 Heads in 100 flips.
- But what if the data was instead:
  - Number of Heads: 70
  - Number of Tails: 30
- Or, what if it was:
  - Number of Heads: 60
  - Number of Tails: 40
- At what point does the data make us doubt that the coin is fair?
  - We use probability and statistics to make these calls!

Hypothesis/Statistical Tests

hypothesis

- The goal of a statistical test is that you have a claim made by someone and you want to know if you can collect data that supports this claim or not.
- We can make a conclusion on whether your data supports the claim by calculating the probability that you get data that is as extreme or more extreme than the data you received, if the claim were actually true.
- Just like confidence intervals, sampling distributions are the foundation of statistical tests.
- Statistical testing is also referred to as hypothesis testing

Process of Statistical Tests

Steps of a Statistical Test

- The procedure for running a statistical test is the same every time
- Steps:
    1. Determine your null and alternative hypotheses. Assume the null is true.
    2. Collect the data and calculate your test statistic
    3. Calculate your p-value (i.e., the probability of getting data that is as extreme or more extreme as the data you have, assuming the null is true)
    4. Make a conclusion based on the p-value
- We will break down each of these steps one at a time.

1. Hypothesis

- To run a statistical test, you need to define two hypotheses:
- **Null hypothesis** (H_0): this is what you are trying to disprove using your data
  - the null will always be the claim that is being made
  - it will always represent that a parameter (i.e. μ or p) is equal to some value
- **Alternative hypothesis** (H_A): this is what you hope to show
  - can be written in a number of ways, but will always represent something opposite to the null
  - can be that parameter is less than (<), bigger than (>), or just not equal to ($\neq$)

1. Hypothesis

$\cancel{H_0: \hat{p} = 0.5}$

$\cancel{H_A: \hat{p} \neq 0.5}$

- Example: Suppose we are interested in a population proportion p, which represents the proportion of coin flips that lands on Heads

$$H_0: p = 0.5$$
$$H_A: p \neq 0.5$$

- The null hypothesis is that the coin is fair. The alternative hypothesis is that the coin is not fair.
- The **null value** is 0.5
- The null and alternative hypotheses are opposites.
- Hypotheses are always statements about true population parameters. They are never statements about sample statistics $(\hat{p}, \bar{x})$

Meaning of Hypotheses

- We are always trying to make inference about a population parameter
- These parameters represent the centre of the distribution of individual observations
  - but also the centre of the sampling distribution
- So the null hypothesis is basically making a statement about where the centre of the distribution actually is
- The alternative hypothesis represents some other centre of the distribution

One-sided or Two-sided Test?

- When we decide on what our null and alternative hypotheses should be, this tells us whether we should run a one-sided or two-sided test.
- The direction of the alternative hypothesis tells us which one to use, since the null is always “=”
- **One-sided test**: if our alternative is “<” or “>”, it is a one-sided test.
  - should only be used if there is an obvious “direction” implied in the question
- **Two-sided test**: if our alternative is “≠”, it is a two-sided test.
  - most common option
  - just means that we don’t know where the true parameter should be, just that it isn’t equal to the claim.

Example: Parking Fees

A city builds a new parking structure in the central business district. The city plans to pay for the structure through parking fees. During a 44 day period, the daily parking fees collected were on average \$126 with a standard deviation of \$15. The consultant who advised the city thinks the parking structure fees will generate an average of \$130 of revenue for the city.

$\bar{x}$

n

s

What are the null and alternative hypotheses?

$H_0: \mu = 130$

$H_A: \mu \neq 130$

Example: Cholesterol Level

A researcher tests whether mean cholesterol level among those who eat frozen pizza exceeds 200 mg/dL. What are the null and alternative hypotheses?

$H_0: \mu = 200$

$H_A: \mu > 200$

$H_0: \mu \leq 200$

$H_A: \mu > 200$

Thought process when finding hypotheses

- To first find the null hypothesis:
  - What is the population parameter of interest? μ or p
  - The population parameter is being compared to a null value. What is this value?
- To then find the alternative hypothesis:
  - Are we interested in comparing the parameter to the null value, regardless of whether it is higher or lower? $\neq$
  - Or, are we interested in whether the parameter is above the null value? $>$
  - Or, are we interested in whether the parameter is below the null value? $<$

2. Compute Test Statistic

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

- Now that we have a hypothesis for the true centre of the distribution, we need to collect data to see if that's a reasonable guess.
- We know from sampling distributions that the larger my sample is, the closer my sample statistic will be to the population parameter.
- With confidence intervals, we measured how far the sample statistic is from the parameter by using the z-score
  - that's exactly what we will do here too!
- So we will need to be working with the sampling distribution again.

Test Statistic for Proportions

- We know the sampling distribution for a sample proportion

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

- We also know that we compute a z-score by subtracting the mean and dividing by the standard deviation of the distribution
- So the test statistic for a test involving proportions is

$$Z\text{-}Statistic = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

- Of course, like with confidence intervals, we don't actually know p
- But we are working under the assumption that the null is true, so just replace p with the null value we are testing.

Example: Insurance

A politician claims that in one in five accidents, a teenager was behind the wheel. An insurance company checks police records on 582 accidents selected at random and notes that teenagers were behind the wheel for 91 of them.

a) What are the hypotheses?

b) What is the test statistic?

a)

$H_0: p = 0.2$

$H_A: p \neq 0.2$

b) $$z\text{-statistic} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

$\hat{p} = 91/582$

$n = 582$

$$= \frac{(91/582) - 0.2}{\sqrt{\frac{0.2 \times 0.8}{582}}} = -2.632$$

Test Statistic for Means

- We also know the sampling distribution for a sample mean is

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Once again, we can determine how far a sample value is from the proposed μ by computing a z-score $\mathrm{Z} = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}$ $\sim N(0,1)$
- As with confidence intervals, we don't know the true value for σ, so we can estimate them with the sample value s
- However, because we have to estimate σ, we cannot use N(0, 1) for the same reason as with confidence intervals
- Instead, we have $\textit{T-Statistic} = \frac{\bar{X}-\mu}{\sqrt{s^2/n}} \sim t_{n-1}$

Example: Parking Fees

A city builds a new parking structure in the central business district. The city plans to pay for the structure through parking fees. During a 44 day period, the daily parking fees collected were on average \$126 with a standard deviation of \$15. The consultant who advised the city thinks the parking structure fees will generate an average of \$130 of revenue for the city.

$\bar{x}$

n

s

What is the test statistic?

$H_0: \mu = 130$

$H_A: \mu \neq 130$

$n = 44$

$\bar{x} = 126, \quad s = 15$

$$\text{T-Statistic} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{126 - 130}{15/\sqrt{44}} = -1.769$$

Summary of Test Statistics

- The test statistic formula will change depending on the parameter of interest.
- For proportions, $Z\text{-}Statistic = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$
- For means, $T\text{-}Statistic = \frac{\bar{X}-\mu}{\sqrt{s^2/n}} \sim t_{n-1}$
- The test statistic is only useful when it comes from a known distribution.
- We previously assumed that the null hypothesis is true. We calculate the test statistic using our data and the assumption that the null is true. Since we know the distribution of the test statistic, we can determine whether it is unusual to see the data that we have

3. Compute the p-value

- **P-values** are probabilities, so must have a value between 0 and 1
  - they represent how likely it is to get a test statistic that is even farther away from the centre of the distribution as the one we calculated.
  - the key to this is that we are finding this probability assuming that the null hypotheses (which we are trying to prove is wrong) is actually right.
  - e.g. if the politician's claim is true, what are the chances that we could observed as few accidents caused by teenagers than we actually saw
  - so like confidence intervals, the p-value reflects the idea of getting different values for different samples.

Computing a p-value

- How we go about finding the p-value will depend on whether we are running a one-sided or two-sided test.
- We want the probability that we could get a more extreme test statistic than ours

	H_A: parameter < value	H_A: parameter > value	H_A: parameter $\neq$ value
Proportion	P(Z < test value)	P(Z > test value)	2 x P(Z > |test value|)
Mean	P(T_{n-1} < test value)	P(T_{n-1} > test value)	2 x P(T_{n-1} > |test value|)

Example: Insurance

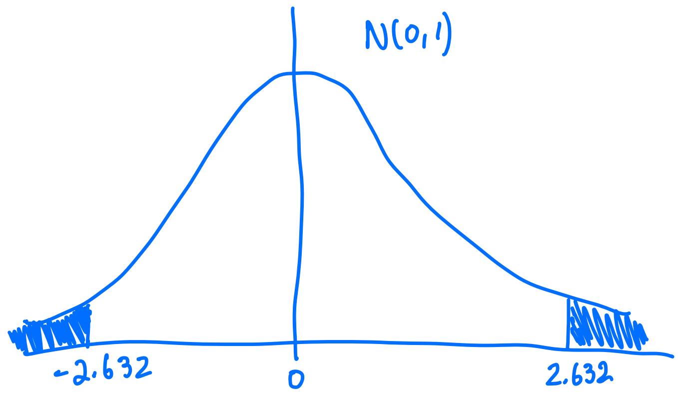
A politician claims that in one in five accidents, a teenager was behind the wheel. An insurance company checks police records on 582 accidents selected at random and notes that teenagers were behind the wheel for 91 of them.

What is the p-value?

$H_0: p = 0.2$

$H_A: p \neq 0.2$

z - statistic = −2.632 which comes from $N(0, 1)$



p-value is the probability that I would get a result as extreme or more extreme than what I observed

p-value = Area of shaded region

$= P(Z < -2.63) + P(Z > 2.63)$

$= 0.0086$

To find p-value:

1. Use distribution table: $2 \times 0.0043 = 0.0086$

2. use R:

   pnorm(q=-2.632) × 2

   ⇒ 0.0086

Example: Parking Fees

A city builds a new parking structure in the central business district. The city plans to pay for the structure through parking fees. During a 44 day period, the daily parking fees collected were on average \$126 with a standard deviation of \$15. The consultant who advised the city thinks the parking structure fees will generate an average of \$130 of revenue for the city.

What is the p-value?

$H_0: \mu = 130$

$H_A: \mu \neq 130$

T-statistic $= -1.769$ which we compare to T_{43}

$$p\text{-value} = P(T_{43} < -1.769) + P(T_{43} > 1.769)$$

Using R:

```
pt(q=-1.769, df=43) * 2
```

$\Rightarrow 0.084$

Example: Parking Fees

$\bar{x} = 126$

How would our p-value change if instead we were testing the alternative that average revenue would be less than \$130?

$H_0: \mu = 130$

$H_A: \mu < 130$

T-Statistic = -1.769 (unchanged)

p-value = $P(T_{43} < -1.769)$

Using R:

pt(q = -1.769, df = 43)

$\Rightarrow$ 0.042

4. Making a Conclusion

- The key to making the right conclusion is understanding what the p-value means
- The p-value represents the chances that we could have collected data that is as extreme or more extreme, assuming that the null hypothesis is true.
- If we get a p-value that is very small, then this means that our test statistic/data is far enough away from the centre of the distribution
  - so if it is very unlikely we could have collected something this far away, then it implies our data is strong enough evidence to suggest that the centre of the distribution is not where originally claimed.
- In summary, a small p-value shows that the data and the null hypothesis are not consistent

How small is small enough?

- We need small p-values in order to reject the null hypothesis and say that the centre of the distribution is not where we thought.
- Below is a table with guidelines for how small we ideally are looking for. Remember that these represent arbitrary cutoffs for what is small enough
    - therefore we cannot say with absolute certainty that the null is wrong
    - we can only say that the strength of the evidence indicates it is likely to be wrong

P-Value	Strength of evidence against the null hypothesis
P-value < 0.001	Very strong
0.001 < P-value < 0.01	Strong
0.01 < P-value < 0.05	Moderate
0.05 < P-value < 0.1	Weak
P-value > 0.1	None

Rejecting the Null Hypothesis

- In this course, unless stated otherwise, we will be using the convention that:
  - If the p-value < 0.05, then <u>we reject the null hypothesis</u>.
  - If the p-value is > 0.05, then <u>we fail to reject the null hypothesis</u>
- Caveat 1: Even though the threshold of 0.05 commonly used, it is completely arbitrary.
  - It is important to recognize this and not feel conflicted when your p-value = 0.051.
- Caveat 2: In real life, it is much more complex than deciding to reject or not to reject.
  - Statistical testing isn't intended to make conclusions. It is intended to present the evidence (strong evidence against null, weak evidence against the null, etc.)
  - In practice, statisticians provide the evidence. Whether we indeed want to reject claims requires consulting with subject matter experts

Example: Insurance

A politician claims that in one in five accidents, a teenager was behind the wheel. An insurance company checks police records on 582 accidents selected at random and notes that teenagers were behind the wheel for 91 of them.

What conclusion can we make?

$H_0: p = 0.2$

$H_A: p \neq 0.2$

p-value = 0.0086

Since p-value < 0.05, we reject the null hypothesis that 20% of accidents have a teenager behind the wheel.

We have evidence against the null hypothesis that ...

Example: Parking Fees

A city builds a new parking structure in the central business district. The city plans to pay for the structure through parking fees. During a 44 day period, the daily parking fees collected were on average \$126 with a standard deviation of \$15. The consultant who advised the city thinks the parking structure fees will generate an average of \$130 of revenue for the city.

What conclusion can we make?

$H_0: \mu = 130$

$H_A: \mu \neq 130$

p-value = 0.084

Since p-value > 0.05, we fail to reject the null hypothesis

Exercise

In 1960, census results indicated that the mean age at which Canadian women first married was 22.6 years. It is widely suspected that young people are waiting longer to get married. We wish to determine if the mean age of first marriage has increased. In a sample of 40 women, we have an average age of 27.2 years with a standard deviation of 5.3 years. Conduct an appropriate hypothesis test.

n

$\bar{x}$

s

$H_0: \mu = 22.6$

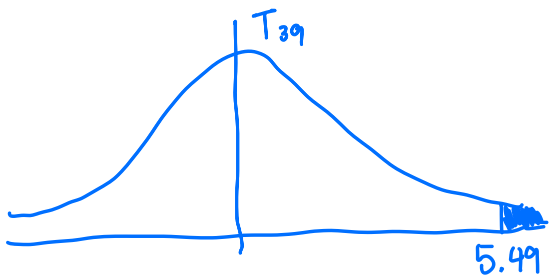
$H_A: \mu > 22.6$

Test Statistic:

$n = 40$ $\bar{x} = 27.2$ $s = 5.3$

T-Statistic $= \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{27.2 - 22.6}{5.3/\sqrt{40}} = 5.49$

p-value :



p-value $= P(T_{39} > 5.49) < 0.0001$

Using R: 1- pt(q=5.49, df=39)

Conclusion:

We have very strong evidence against the null hypothesis.

We reject the null hypothesis that the age of first marriage is 22.6.

We reject the null hypothesis, if favour of the alt. hypothesis that ...

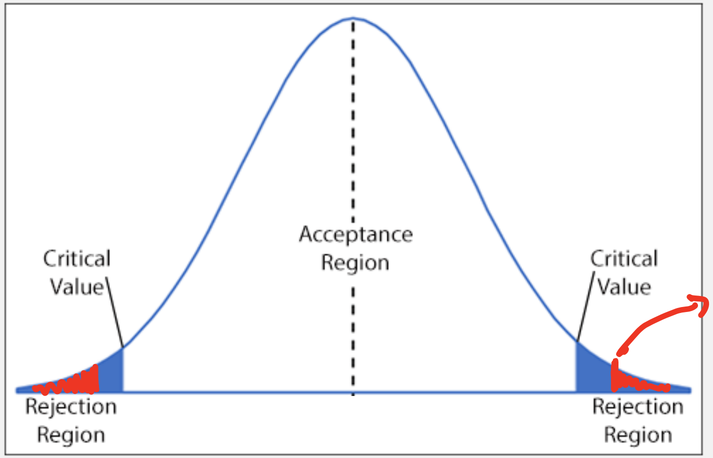
Comparing Critical Values instead of Computing p-values

Instead of calculating the p-value

- Instead of calculating the p-value and comparing it to a threshold, there is a *shortcut*
- Instead of calculating the p-value and comparing it to a threshold, we can just compare the test statistic to a critical value

Rejection and Acceptance Regions for Two-Sided Tests

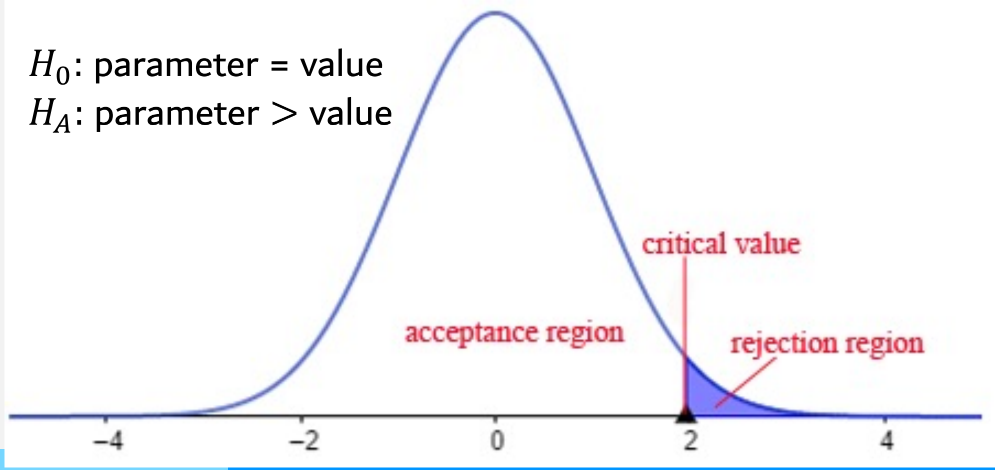
- To use this shortcut with a threshold of 0.05:
  - Find the critical value such that the area of both tails together is 0.05. That means the area of each tail should be 0.025.
  - <u>Rejection Region</u>: If the test statistic is less than (- critical value) or greater than (+ critical value), then the p-value will be less than 0.05.
  - <u>Acceptance Region</u>: If the test statistic is between (- critical value) and (+ critical value), then the p-value will be greater than 0.05.



Test Stat

Rejection and Acceptance Regions for One-Sided Tests

- To use this shortcut with a threshold of 0.05:
  - Find the critical value such that the area of the rejection tail is 0.05. The direction of the tail depends on the direction of H_A
  - Rejection Region: If the test statistic is within the rejection tail, then the p-value will be less than 0.05.
  - Acceptance Region: If the test statistic is not with the rejection tail, then the p-value will be greater than 0.05.



Example

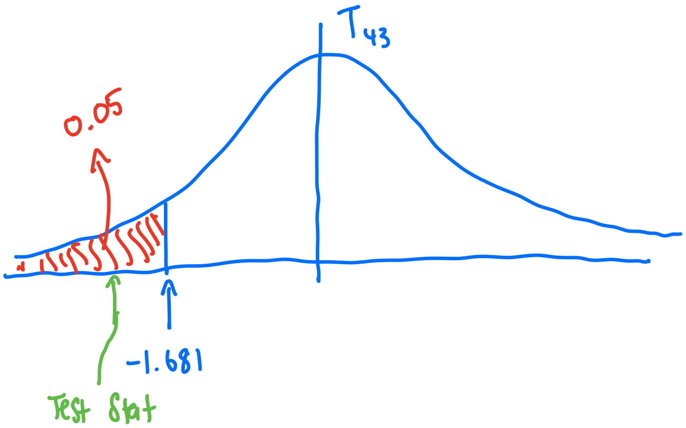
- Let's re-do the example from Slide 29

$$H_0: \mu = 130$$
$$H_A: \mu < 130$$

Test statistic = -1.769, which we compare to T_{43}

- Previously, we calculated the p-value as follows:
  - P-value $= P(T_{43} < -1.769) = 0.042$
  - Then we reject H_0 since the p-value is less than 0.05
- Instead, let's use the critical value to bypass finding the p-value

Using the critical value:



Using R:

qt(p = 0.05, df = 43)

⇒ -1.681

Test Stat = $\underline{-1.769}$

Since test statistic < -1.681, we would reject the null.

Practice Problems

- In this lecture, we covered all of Module 8: The Process of Statistical Tests
- Practice problems are posted

Upcoming

- Usual quiz is due Sunday
- Assignment is posted. It is due March 24 on Crowdmark.