

P-values

Donna Dietz

American University

dietz@american.edu

STAT 202 - Spring 2020

We have been using p-values in this course for awhile without fully explaining what they mean. It's finally time to explain it!

This song by Mark Lesser is sung to “Row Row Row Your Boat”:

What P-Value Means

It is key to know

What p-value means –

It's the chance

(with the null)

you obtain

data that's

At least that extreme!

What P-Value Means

It's the chance you would be at least that far away from the hypothesized mean due to sampling error alone, presuming the null hypothesis is right.

The p - *value* is a conditional probability!

Intervals: The plus and the minus

We've been using the z – scores and t – scores to create confidence intervals. This is generally wise. By creating an interval, we are emphasizing that the sample mean (\bar{x}) or sample proportion (\hat{p}) are never going to be precisely the population mean (μ) or proportion (p).

There is wisdom in making up the rules before you play the game, and then playing through to completion.

If you have time, please Google *p-hacking* later.

This is one of my favorite articles that explains something you should not do!

https://www.graphpad.com/guides/prism/7/statistics/index.htm?stat_why_you_shouldnt_recompute_p_v.htm

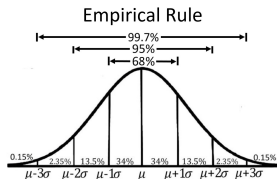
However, when you simply report that you did or did not reject a hypothesis, this doesn't tell anyone how overwhelming the evidence was, only that it met your specific threshold for significance!

So, instead of creating an interval, you can report the *p* – *value*.

Example

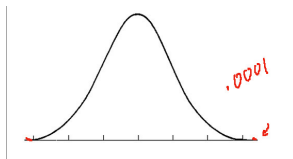
First, let's recall how we've been using these p – values when we see them in StatCrunch.

For our Empirical Rule, we start in the middle of the bell curve and work our way out:



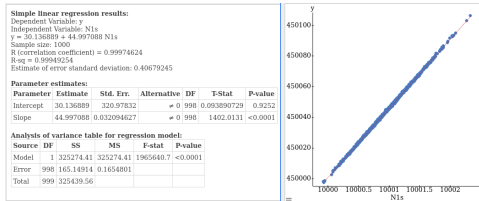
However...

However, when we talk about p – values, we are talking about the area in the tails!



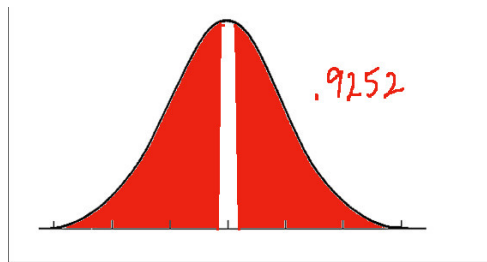
This is a sketch of what you should imagine when you see $p < 0.0001$.

Example



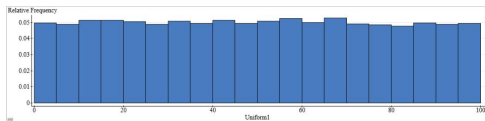
Here, we see there is a high significance to R or the slope of the line. This is why the p – value is too low to print. However, the intercept is nothing to brag about.

A tail that's not a tail



This is what you should imagine when you see a high p – value. This value, which is very close to one, fills the two so-called tails, nearly entirely! This is not unlikely. In fact, I'm actually suspicious if the p – value is too high. When it's precisely equal to 1, it's probably designed to be that way!

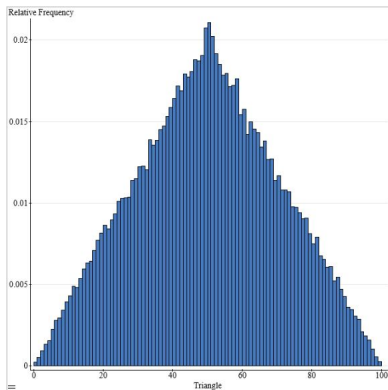
A uniform distribution



This is a histogram of a StatCrunch random draw of values between 0 and 100. Just looking at this, roughly how likely do you think it is that a random value between 0 and 100 is 90 or above?

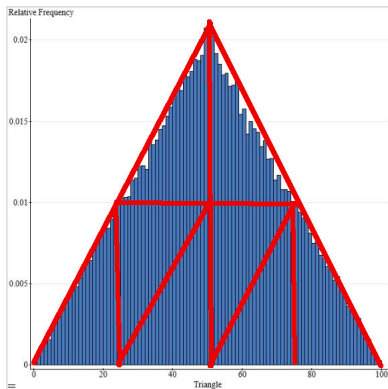
For uniform distributions, this is easy to calculate, because the distribution is made to do this very easily! We can also do something like this with triangles.

A triangle distribution



You can make a triangle distribution by sampling twice from a uniform distribution and adding the two draws. So, to create a triangle distribution from 0 to 100, you can draw twice from a uniform distribution from 0 to 50.

A triangle distribution



Let's verify that this area is about 1 or 100%. Each of these 8 smaller triangles has a height of 0.01 and a width of 25. That means each one has an area of

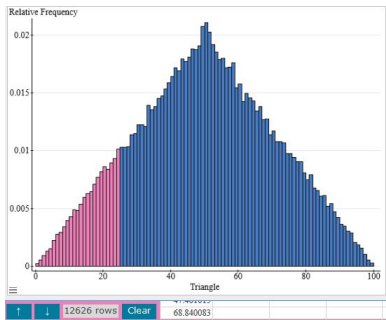
$$\frac{1}{2} \cdot 0.01 \cdot 25 = 0.125$$

And also note that

$$\frac{1}{8} = 0.125.$$

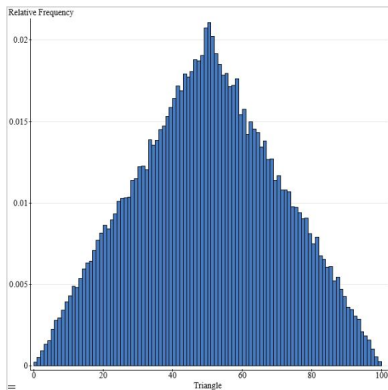
So this means that the total is 1 or 100%.

A p – value



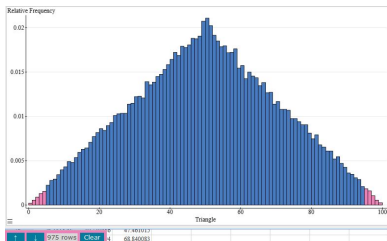
So, for this distribution, the probability of being at or below 25 is theoretically 12.5%. In this case, 12,626 out of 100,000 rows worked, or about 12.6%, which is very close.

A p – value



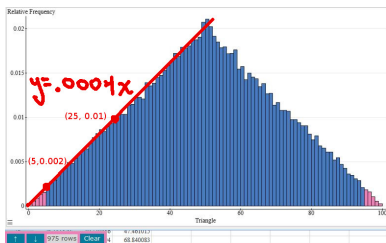
So, for this triangle, you could also ask something like, what is the probability of being farther than 45 away from 50? In other words, what's the probability of **not** being between 5 and 95?

Two far triangles



So, what's the probability that you land way out in one of the two tiny corners?

Two far triangles



We can figure out the equation of the line on the left side of the triangle, and then use that to figure out the area of the small left region, then double that!

$$\frac{1}{2} \cdot 5 \cdot 0.002 = .005$$

$$0.005 \cdot 2 = 0.01 = 1\%.$$

This is the same idea behind all of our p – values. For any distribution, we can determine how likely we are to randomly wind up in one of the two tails.

When people run 1-tail tests, they only test the probability of landing in one tail, but we've already discussed the issues with that.

What does the p – value tell us?

p – value

The p – value can only tell you how *unusual* your findings are **if** your Null Hypothesis is true. It is a conditional probability.

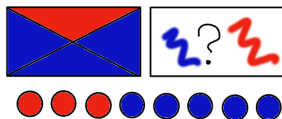
What we wish we could know, but we sadly can never know, is the likelihood that the Null Hypothesis is true (or false).

Example

I have two infinite boxes of balls, both are full of red and blue balls. The one box contains 25% red balls. The other one, I have NO clue about!

I randomly pull 8 balls from one of the boxes. Three are red. What's the chance that I pulled the balls out of the first box?

Answer: There is no way to tell!



Example

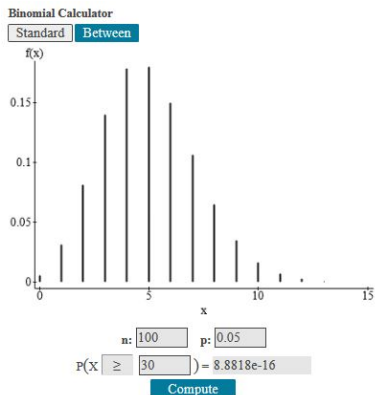
I have two infinite boxes of balls, both are full of red and blue balls. The one box contains 5% red balls. The other one, I have NO clue about!

I randomly pull 100 balls from one of the boxes. Thirty are red. What's the chance that I pulled the balls out of the first box?

Answer: I cannot assign a probability to this. However, at this point, I can say it's unlikely to be the first box. If I'd pulled from the first box, the chance of drawing 30 or more red balls would be something like 0.000000000000000088818.

Did you just use a one-tail test?

Yes, I just did that! But, I told you not to do it! Why? The distribution is quite skewed, and there is nothing analogous to drawing -20 red balls.



StatCrunch gave me that answer, by the way!

Why did I say “or more”?

When you are calculating this conditional probability (this special one we call the *p* – *value*), you need to always include everything more extreme than the value you are looking at.

Another example

I have two infinite boxes of balls, both are full of red and blue balls. The one box contains 50% red balls. The other one, I have NO clue about!

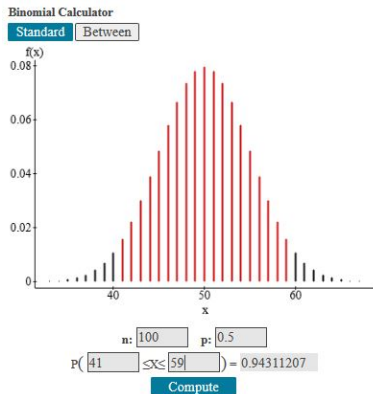
I randomly pull 100 balls from one of the boxes. Forty are red. Now, what's the chance that I pulled the balls out of the first box?

Answer: There is **still** no way to tell!

The plot thickens.

However, at this point, I can say it could be unlikely to be the first box. If I'd pulled from the first box, the chance of drawing 40 or more red balls **or** 60 or more red balls would be $1 - 0.94311 = 0.05689$ or about 5.7%. This is a **conditional** probability! We still really don't know which box we pulled the balls from! With *p* – *value* near 0.05, we should be thinking “one in 20”, because we can get this type of result fairly often. We can't rule out the possibility that this just randomly happened due to sampling noise. And let's not forget: We still have no idea what's in the other box!

Look at it this way...



Calculating p – values

Finding p – values

- Determine if you will use t or z Tables
 - For proportions, use z
 - For normal populations or samples over 1000, use z
 - When in doubt, use t (with $df = n - 1$)
- Calculate $\frac{\bar{x} - \mu_0}{s/\sqrt{n}}$.
 - Use σ if available, otherwise use s .
 - μ_0 is the hypothesized value
 - For proportions use $\sqrt{p(1-p)/n}$ in the denominator.
- Find the area in one tail using that cutoff
- Double that area

Both ways...

Solve the hypothesis question below using a confidence interval and again using a p - *value*.

The formula for the confidence interval is:

$$\bar{x} - t^* \cdot s/\sqrt{n} < \mu < \bar{x} + t^* \cdot s/\sqrt{n}$$

For a sample size of 34 with sample mean of 7.3 and sample standard deviation of 1.2, with a hypothesized mean of 8, find the 99% confidence interval. Give the null and alternative hypotheses, make your decision, rephrase it. Then, give the p - *value* version of this question and confirm/deny your decision that you gave.

Answer:

$$H_0 : \mu = 8 \quad H_A : \mu \neq 8$$

Since $df = 33$ and we want a 99% confidence level, using the paper table, we use $df = 30$ to get a cutoff of $t^* = 2.750$.

$$\text{Interval: } 7.3 - 2.75 \cdot 1.2/\sqrt{35} < \mu < 7.3 + 2.75 \cdot 1.2/\sqrt{35}$$

$$6.742 < \mu < 7.858$$

We reject H_0 , meaning that we believe it is unlikely that the population mean is 8. Since we've rejected H_0 we can actually say that we believe the population mean is below 8.

$$H_0 : \mu = 8 \quad H_A : \mu \neq 8$$

Since $df = 33$ and we want a 99% confidence level, using the paper table, we use $df = 30$.

Find the t – *value* associated with the sample mean.

$$t = \frac{7.3 - 8}{1.2/\sqrt{34}} \approx -3.45$$

According to the table, in the $df = 30$ row, this lands us with an upper tail probability between 0.001 and 0.0005, so for both tails, that would be a p value somewhere between 0.001 and 0.002.

StatCrunch gives a more accurate p-value (Stat > Calculators > T)

$$2 \cdot 0.00077649 \approx 0.0016$$

Natural Pairings

For data which are naturally paired, you can often subtract them and then treat them as just a single set of data.

Consider the question of whether a campus book store charges more for books than Amazon does.

These values are from David Diez (stats.libretexts.org).

For 73 randomly selected courses with texts, $\bar{x} = 12.76$ and $s = 14.26$. Note that the data are actually skewed right. For each book, the differences were given as *bookstore minus Amazon*.

With H_0 being no difference between prices, calculate the t – score as

$$\frac{12.76}{14.26/\sqrt{73}} = 7.645.$$

I don't even have to look this one up. It's highly significant.
Reject H_0 . Amazon is cheaper on average than the campus store.

Non-paired data

For data which are not paired, there is a different formula to use. If either one has a sample size under 1000, subtract one from the smaller sample (or population) and that will be the *df* you use by hand with the paper table. If both samples are over 1000, you can use *z – scores*.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Non paired data

A prof who teaches two sections of a statistics course wishes to see if the final exam scores are statistically different based on their averages. (If we'd had 3 or more sections, we'd have to use ANOVA to do this, but for just two, we may opt for non-paired t-tests.)

For section 1, $n = 24$, $\bar{x} = 80$ and $s = 15$. For section 2, $n = 16$, $\bar{x} = 85$, and $s = 10$. Do the test at a 95% level.

For section 1, $n = 24$, $\bar{x} = 80$ and $s = 15$. For section 2, $n = 16$, $\bar{x} = 85$, and $s = 10$. Do the test at a 95% level.

Since $df = 15$, the cutoff value is $t^* = 2.131$. Let's find the t - value now.

$$t = \frac{80 - 85}{\sqrt{\frac{15^2}{24} + \frac{10^2}{16}}} \approx -1.265$$

This value is within plus or minus of the cutoff value so the difference is not significant. We fail to reject the null hypothesis, and we cannot conclude there is any difference between the sections, even though section 2 would prefer to brag about their achievements!

3 MEMORY QUESTIONS

home/dietz/pCloudDrive/A: X +

STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html ☆

Google Canvas Cups EduUnempPovPopCo... MATH221_Text Mail JAM

STAT 202 Memory Questions

Combined Sets ▾

To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.

Click all correct answers, then click submit:

You have two boxes of red and blue balls. Box one has 25 red and 75 blue balls. You don't know what is in the other box. You draw 8 balls from one of the boxes at random and 3 balls are red. What box did you likely pull from?

You probably took from box two.

You took from box one for sure.

This question can't be answered.

You probably but not certainly took from box one.

SUBMIT

Browser address bar: /home/dietz/pCloudDrive/A: X +

Browser tabs: /STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html ☆

Browser extensions: Google, Canvas, Cups, EduUnempPovPopCo..., MATH221_Text, Mail, JAM

STAT 202 Memory Questions

Combined Sets ▾

To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.

Click all correct answers, then click submit:

You have two boxes of red and blue balls. Box one has 25 red and 75 blue balls. You don't know what is in the other box. You draw 8 balls from one of the boxes at random and 3 balls are red. What box did you likely pull from?

You probably took from box two.

You took from box one for sure.

This question can't be answered.

You probably but not certainly took from box one.

SUBMIT

Browser address bar: /home/dietz/pCloudDrive/A: X +

Browser tabs: /STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html ☆

Browser extensions: Google, Canvas, Cups, EduUnempPovPopCo..., MATH221_Text, Mail, JAM

STAT 202 Memory Questions

Combined Sets ▾

To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.

Click all correct answers, then click submit:

You have two boxes of red and blue balls. Box one has 10 red and 75 blue balls. You don't know what is in the other box. You draw 8 balls from one of the boxes at random and all 8 balls are red. What box did you likely pull from?

You pulled from box one.

Nothing can be said here.

You were unlikely to have pulled from box one.

You pulled from box two for certain.

SUBMIT

Browser address bar: /home/dietz/pCloudDrive/A: X +

Browser tabs: /s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html ☆

Browser extensions: Google, Canvas, Cups, EduUnempPovPopCo..., MATH221_Text, Mail, JAM

STAT 202 Memory Questions

Combined Sets ▾

To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.

Click all correct answers, then click submit:

You have two boxes of red and blue balls. Box one has 10 red and 75 blue balls. You don't know what is in the other box. You draw 8 balls from one of the boxes at random and all 8 balls are red. What box did you likely pull from?

You pulled from box one.

Nothing can be said here.

You were unlikely to have pulled from box one.

You pulled from box two for certain.

SUBMIT

Browser address bar: /home/dietz/pCloudDrive/A: X +

Browser tabs: /STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html ☆

Browser extensions: Google, Canvas, Cups, EduUnempPovPopCo..., MATH221_Text, Mail, JAM

STAT 202 Memory Questions

Combined Sets ▾

To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.

Click all correct answers, then click submit:

What is the difference between paired data and non-paired data?

If the two sets don't have a natural pairing, then we call it 'unpaired'.

I have no clue.

If you have a natural pairing between two sets of data you should subtract them and treat the data as if it is just one set of data.

There is no difference. It's just two ways to say the same thing.

SUBMIT

Browser address bar: /home/dietz/pCloudDrive/A: X +

Browser tabs: /STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html

Browser extensions: Google, Canvas, Cups, EduUnempPovPopCo..., MATH221_Text, Mail, JAM

STAT 202 Memory Questions

Combined Sets ▾

To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.

Click all correct answers, then click submit:

What is the difference between paired data and non-paired data?

If the two sets don't have a natural pairing, then we call it 'unpaired'.

I have no clue.

If you have a natural pairing between two sets of data you should subtract them and treat the data as if it is just one set of data.

There is no difference. It's just two ways to say the same thing.

SUBMIT