Exploration librairie swirl : P-value

> swirl()

Welcome to swirl! Please sign in. If you've been here before, use the same name as you did then. If you are new, call yourself something unique.

What shall I call you? jlbellier

Please choose a course, or type 0 to exit swirl.

1: Statistical Inference
2: Take me to the swirl course repository!

Selection: 1

Please choose a lesson, or type 0 to return to course menu.

1: Introduction	*2: Probability1*	*3: Probability2*
4: ConditionalProbability	*5: Expectations*	*6: Variance*
7: CommonDistros	*8: Asymptotics*	*9: T Confidence Intervals*
10: Hypothesis Testing	*11: P Values*	*12: Power*
13: Multiple Testing	*14: Resampling*	

Selection: 11

Attempting to load lesson dependencies...

Package 'ggplot2' loaded correctly!

0%

P_Values. (Slides for this and other Data Science courses may be found at github https://github.com/DataScienceSpecialization/courses/. If you care to use them, they must be downloaded as a zip file and viewed locally. This lesson corresponds to 06_Statistical_Inference/10_pValues.)

==   2%

In this lesson, as the name suggests, we'll discuss p-values which have nothing to do with urological testing. Instead they are the most common measure of statistical significance.

====   5%

However, because they're popular they're used a lot, and often they're misused or misinterpreted. In this lecture we'll focus on how to generate them and interpret them correctly.

======   7%

The question motivating p-values is this. Given that we have some null hypothesis concerning our data (for example, its mean), how unusual or extreme is the sample value we

get from our data? Is our test statistic consistent with our hypothesis? So there are, implicitly, three steps we have to take to answer these types of questions.

========= **10%**

What do you think the first step is?

1: Create a null hypothesis
2: Consult your crystal ball
3: Compare the test statistic to a Z or t quantile
4: Calculate a test statistic from the data

Selection: **1**

Great job!

========== **12%**

So we have to begin with a null hypothesis which is a reasoned guess at some distribution of a data summary (a statistic). Recall from the last lesson that the null hypothesis H_0 is a baseline against which we'll measure an alternative hypothesis using the actual observed data.

============ **14%**

So you propose a null hypothesis. What's the next step?

1: Go back to the crystal ball
2: Calculate a test statistic from the given data
3: Compare the test statistic to a Z or t score
4: Reject H_0

Selection: **2**

That's the answer I was looking for.

=============== **17%**

Now you have a proposed statistic (from your reasoned hypothesis) and a test statistic computed from your gathered data. What's the final step?

1: Go back to the crystal ball
2: Reject H_0
3: Compare the test statistic to the hypothetical distribution
4: Calculate a test statistic from the given data

Selection: **3**

Nice work!

================== **19%**

Your comparison tells you how "extreme" the test value is toward the alternative hypothesis. The p-value is the probability under the null hypothesis of obtaining evidence as or more extreme than your test statistic (obtained from your observed data) in the direction of the alternative hypothesis.

==================== **21%**

So if the p-value (probability of seeing your test statistic) is small, then one of two things happens. EITHER H_0 is true and you have observed a rare event (in this unusual test statistic) OR H_0 is false. Let's go through an example.

Suppose that you get a t statistic of 2.5 with 15 df testing H_0, (that mu = mu_0) versus an alternative H_a (that mu > mu_0). We want to find the probability of getting a t statistic as large as 2.5.

R can help us! We can use the R function pt, the distribution function of the t distribution. This function returns one of two probabilities, EITHER the probability of X > q (if lower.tail is FALSE) OR X <= q (if lower.tail is TRUE), where q is a quantile argument. Here we'll set q=2.5, df=15, lower.tail=FALSE since H_a says that mu>mu_0. We have to gauge the extremity in the direction of H_a. Run this now.

```
> pt(q=2.5, df=15,lower.tail = FALSE)
[1] 0.0122529
```

Your dedication is inspiring!

This result tells us that, if H_0 were true, we would see this large a test statistic with probability 1% which is rather a small probability.

What should we do?

1: Reject H_0
2: Consult the crystal ball
3: Fail to reject H_0

Selection: 1

You nailed it! Good job!

Another way to think about a p-value is as an attained significance level. This is a fancy way of saying that the p-value is the smallest value of alpha at which you will reject the null hypothesis.

Recall the example from our last lesson in which we computed a test statistic of 2. Our H_0 said that mu_0 = 30 and the alternative H_a that mu > 30. Assume we used a Z test (normal distribution). We rejected the one sided test when alpha was set to 0.05.
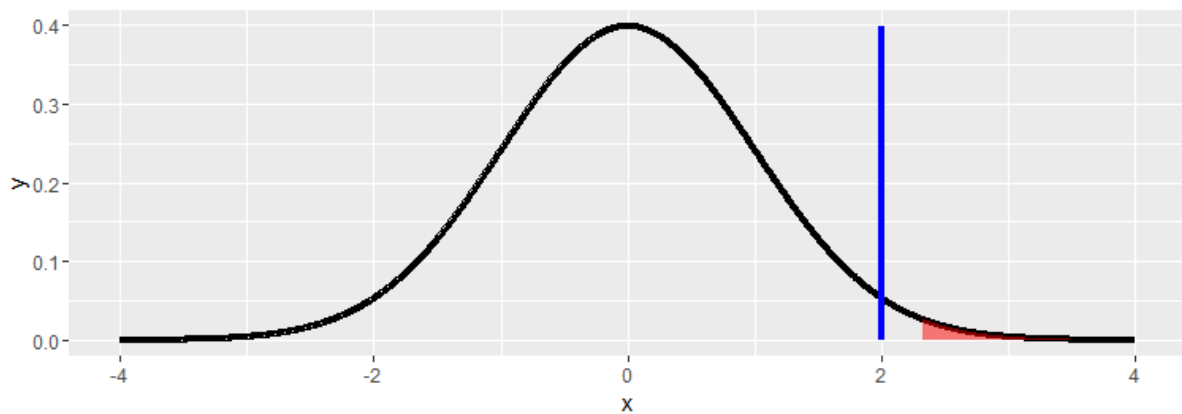
Why did we reject? Find the quantile associated with this test, that's the place to start. Use qnorm at the 95th percentile.

```
> qnorm(0.95)
[1] 1.644854
```

Your dedication is inspiring!

We rejected H_0 because our data (the test statistic actually) favored H_a. The test statistic 2 (shown by the vertical blue line) falls in the shaded portion of this figure because it exceeds the quantile. As you know, the shaded portion represents 5% of the area under the curve.

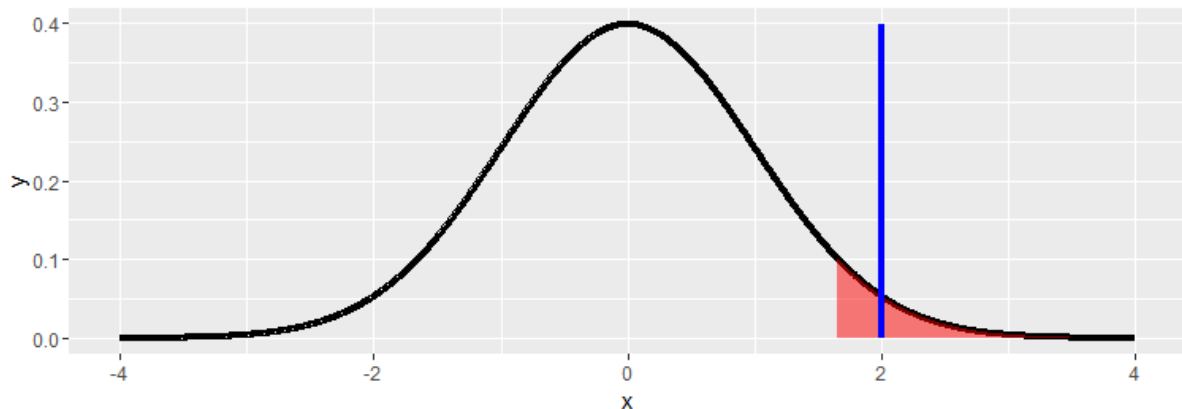*Now try the 99th percentile to see if we would still reject H_0.*

> *qnorm(0.99)*

[1] 2.326348

Keep up the great work!

Would we reject H_0 if alpha were .01?

1: Yes
2: No

Selection: 2

You are amazing!

Again, a picture's worth a thousand words, right? The vertical line at the test statistic 2 is not in the region of rejection.

So our data (the test statistic) tells us what the attained significance level is. We

use the R function pnorm to give us this number. With the default values, specifically lower.tail=TRUE, this gives us the probability that a random draw from the distribution is less than or equal to the argument. Try it now with the test statistic value 2. Use the default values for all the other arguments.

> *pnorm(q=2)*
[1] 0.9772499

Keep working like that and you'll get there!
```
    ================================================            52%
```
Just as we thought, somewhere between .95 (where we rejected) and .99 (where we failed to reject). That's reassuring.
```
    ================================================            55%
```
Now let's find the p value associated with this example. As before, we'll use pnorm. But this time we'll set the lower.tail argument to FALSE. This gives us the probability of X exceeding the test statistic, that is, the area under the curve to the right of test statistic. Try it now with the test statistic value 2.

> *pnorm(q=2,lower.tail = FALSE)*
[1] 0.02275013

Excellent work!
```
 ======================================================         57%
```
This tells us that the attained level of significance is about 2%.
```
    ========================================================    60%
```
By reporting a p-value, instead of an alpha level and whether or not you reject H_0, reviewers of your work can hypothesis test at any alpha level they choose. The general rule is that if the p-value is less than the specified alpha you reject the null hypothesis and if it's greater you fail to reject.
```
    =========================================================   62%
```
For a two sided hypothesis test, you have to double the smaller of the two one-sided p values. We'll see an example of this shortly. Most software assumes a two-sided test and automatically doubles the p value.
```
    ==========================================================  64%
```
Now for the two-sided test. Recall the binomial example from the last lesson - the family with 8 children, 7 of whom are girls. You want to test H_0, that p=.5, where p is the probability of a girl (like a fair coin flip). H_a is that p is not equal to .5. It's either greater or less than .5.
```
    =========================================================== 67%
```
This is a two-sided test. First we find the probability of having at least i girls, for i running from 0 to 8. We have a vector of these probabilities, mybin. Look at it now.

> *mybin*
[1] 1.00000000 0.99609375 0.96484375 0.85546875 0.63671875 0.36328125 0.14453125
[8] 0.03515625 0.00390625

That's a job well done!

The second last value shows us that the probability of having at least 7 girls (out of 8 children) is .035, assuming that genders are equally likely (p=.5). You can verify this with the R function pbinom, with the arguments 6, size=8, prob=.5, and lower.tail=FALSE. (This last yields the probability that X>6.) Try this now.

> **pbinom(6,8, prob=0.5,lower.tail=FALSE)**
[1] 0.03515625

Great job!
We see a probability of about .03. Should we reject or fail to reject H_0 if alpha = .05?

1: Fail to reject
2: Reject

Selection: **2**

Keep up the great work!
We see a probability of about .03. Should we reject or fail to reject H_0 if alpha = .04?

1: Fail to reject
2: Reject

Selection: **2**

All that practice is paying off!
We see a probability of about .03. Should we reject or fail to reject H_0 if alpha = .03?

1: Reject
2: Fail to reject

Selection: **2**

That's correct!
For the other side of the test we want the probability that X<=7, again out of a sample of size 8 with probability .5. Again, we use pbinom, this time with an argument of 7 and lower.tail=TRUE. Try this now.

> **pbinom(q=7,size=8, prob=0.5, lower.tail = TRUE)**
[1] 0.9960938

That's a job well done!

So it's pretty likely (probability .996) that out of 8 children you'll have at most 7 girls. The p value of this two sided test is 2*the smaller of the two one-sided values. In this case the lower value is .035, so 2*.035 is the p-value for this two-sided test.

==**83%**

Now a final example using a Poisson distribution. Remember that this is discrete and it involves counts or rates of counts. The example from the slides involves rates of infections in a hospital.

== **86%**

Suppose that the hospital has an infection rate of 10 infections per 100 person/days at risk. This is a rate of 0.1. Assume that an infection rate of 0.05 is the benchmark. This is our alpha level, recognize it? With this model, could the observed rate (.1) be larger than the benchmark 0.05 by chance or does it indicate a problem?

== **88%**

In other words, H_0 says that lambda = 0.05 so $lambda_0 * 100 = 5$, and H_a says that lambda > 0.05. Is H_0 true and our observed rate (.1) is just a fluke OR should we reject H_0 ?

==**90%**

As before, R has the handy function ppois, which returns probabilities for Poisson distributions. We want the probability of seeing at least 9 infections using a lambda value of 5 and lower.tail=FALSE. As when we used pbinom we have to use 9 as the argument since we're looking for a probability of a value greater than the argument. Try this now.

> **ppois(q=9, lambda=5, lower.tail=FALSE)**
[1] 0.03182806

All that hard work is paying off!
== **93%**

We see a probability of about .03. Should we reject or fail to reject H_0? (Remember those helpful pictures with shaded areas. Smaller areas mean smaller probabilities and vice versa.)

1: Reject
2: Fail to reject

Selection: **1**

All that hard work is paying off!
== **95%**

So we reject the infection rate hypothesized by H_0 since the data favors H_a, indicating that the rate is much higher.

== **98%**

Congrats! You finished this lesson. We hope you p-valued it.

== **100%**