**P-Values**

**FAQ**

1. **What is NULL hypothesis? Is this different every time? Is NULL what *we want?***

**ANSWER**: NULL hypothesis states that there is no association. Basically that “You have found nothing significant”. Nice to hear, isn’t it ☺ For sure, you will love that someone who is telling you this. Yeah, this is why statisticians are not the most popular and liked people in the world!

So do we want to support or reject NULL?

For sure, **we hope to reject NULL hypothesis** – since only then we “have something” – we show that there is an association….what we want to show is often called **ALTERNATIVE HYPOTHESIS**

Thus the smaller p-value the better is the news for us. Our discovery is unlikely due to random chance! The larger the p-value the more grim are the news. Close to 1, means it is almost for sure just random. NULL hypothesis holds and we do “have nothing”

Examples of NULL hypotheses:

1. There is no association between tunnels and traffic volumes – they all the same
2. There is no association between happiness of immigrants and natives
3. There is no association between capital gains and zodiac signs
4. There is no association between age and happiness

Examples of ALTERNATIVE HYPOTHESES

1. Traffic in Holland Tunnel is higher than traffic in Lincoln tunnel
2. Immigrants are happier than Natives
3. Capital gains of born under Virgo sign are higher than under any other sign
4. Older people are happier than younger ones
5. **Is p-value a probability of NULL being true? or is it a probability of ALTERNATIVE hypothesis to be false?**

**ANSWER: None of the above.** P-value is just a probability of getting the result as extreme or more extreme as we got UNDER ASSUMPTION than NULL hypothesis is true. Some call is “conditional probability”

Basically p-value is response to a question from audience

**“Hey, isn’t it all due to random chance?”**

p-value is your answer: No, only 5% chance that we will get what we got due to random chance! Thus what we got is \*too extreme\* to be random, too far away from zero difference of means…there we reject NULL hypothesis and accept Alternative Hypothesis.

1. **What about this permutation test. If 950 out of 1000 permutations support your alternative hypothesis doesn’t it say that there is 95% chance that alternative hypothesis is true?**

**ANSWER:**  Nope. Remember these 950 permutations generate “junk data” where all associations are destroyed (say between traffic volumes an tunnels, or happiness and immigrant status. They are not \*evidence\* for anything other than randomness under NULL hypothesis. In fact they kind of mean the opposite to that 95% chance that our alternative hypothesis is correct. They mean that there is 95% we can get the observed result (Immigrants happier than Natives etc), by RANDOM CHANCE

1. **Why is p-value different every time we run permutation test?**

**ANSWER:** Permutation test provides us only with approximation of p-value, since permutations are random we will get random results, thus different every time. But if we run them many times (depends on the size of our sample, for large samples it better be large number of times) the p-value will converge

1. **Can we apply permutation test for NULL hypothesis stating that difference of means is equal to K, different than zero?**

**ANSWER**: No. permutation test we use, always assumes NULL as no association, thus difference of means being zero. Our permutations always generate results around mean of zero (since we scramble data randomly). So in our “junk instances” we always get nearly zero difference between traffic in Lincoln and Holland, happiness between natives and immigrants, capital gains between Virgo and Aries etc.

We can apply z-test to reject hypothesis that difference of means is equal to K different than zero

Please run the code (from the class to get a feel for permutation test for two distributions of Cat1 and Cat2.

Change the mean and sd in rnorm distributions

Val1<-rnorm(10,mean=25, sd=10)

Val2<-rnorm(10,mean=30, sd=10)

and see how it affects p-value! Very revealing.....

install.packages("devtools")

devtools::install\_github("devanshagr/PermutationTestSecond")

Cat1<-rep("GroupA",10)

Cat2<-rep("GroupB",10)

Cat<-c(Cat1,Cat2)

Cat

Val1<-rnorm(10,mean=25, sd=10)

Val2<-rnorm(10,mean=30, sd=10)

Val<-c(Val1,Val2)

Val

d<-data.frame(Cat,Val)

mean(d[d$Cat=='GroupA',2])-mean(d[d$Cat=='GroupB',2])

PermutationTestSecond::Permutation(d, "Cat", "Val",10000, "GroupA", "GroupB")

1. **Example – like the one from clicker and lecture slides but in more detail**

We have used example of average happiness of immigrants vs natives in lecture notes as well as in the clicker tests. We were testing hypothesis that immigrants are more happy than natives (so called alternative hypothesis) against NULL hypothesis that there is no difference between these groups.

In two of the clicker questions we used as example the observed data where

***Mean happiness of immigrants – Mean happiness of natives = 1.1 (8.6-7.5)***

In one clicker question we had 950 permutations out of 1000 total with the following result:

*Mean happiness of immigrants – Mean happiness of natives < 1*

Remember that difference of means converges to the Bell curve centered around 0 (permutation tests scramble data in the observed sample randomly between two “labels” – immigrants and natives completely breaking any association between happiness and immigrant/native status). Samples generated by permutation test are just random (mean centered around 0). The p-value is defined as probability of achieving result “more extreme than observed difference of means of 1.1” by random permutation test.

We know that in 950 cases we failed to reach the result as extreme as mean difference of 1.1. We only got results as extreme as 1. Therefore we know that only in 50 cases we got a result larger or equal to 1. This is 5% of all tests. Out of these 50 results, some number smaller (or equal) to 50, reached 1.1 or more. Hence p <=0.05. There is less 5% (or less) chance that our observed result came randomly from data with no association between labels (immigrant/native and happiness).

This is why we can conclude that our observed result (real sample we have, not “junk” ones obtained by permutation test) in NOT random and we accept our *alternative hypothesis that immigrants are more happy than natives.*

And in another we just reversed the “<” into “>”

Stating that 950 out of 1000 permutations showed

*Mean happiness of immigrants – Mean happiness of natives > 1*

What can we say about p-value in this case?

Now, only in 50 out of 1000 permutations we have obtained a result as extreme as 1 (1000-950). Thus in less (or equal) than 950 cases we have randomly generated samples as extreme as 1.1. This is why

p<=0.95

Can this really happen? Very unlikely, probably the test was buggy or results were “cherry picked”. Since permutation scores are centered around zero with a bell curve distribution – it is extremely unlikely that 950 of them would show the mean larger than 950 and only 50 get value smaller than 1…. But if this really happened than p was nearly 0.95, since all of this samples were just random (not observed) showing our real observed result could be easily obtained randomly. Thus, we cannot accept our hypothesis that immigrants are more happy than natives. Our permutation test got this result almost 95% of times by \*random chance\*.

**HERE IS PART OF A BLOG on p-values**

<http://blog.minitab.com/blog/adventures-in-statistics-2/how-to-correctly-interpret-p-values>

“Everyone knows that you use P values to determine statistical significance in a [hypothesis test](http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/what-is-a-hypothesis-test/). In fact, P values often determine what studies get published and what projects get funding.

Despite being so important, the P value is a slippery concept that people often interpret incorrectly. How *do* you interpret P values?

In this post, I'll help you to understand P values in a more intuitive way and to avoid a very common misinterpretation that can cost you money and credibility.

**What Is the Null Hypothesis in Hypothesis Testing?**

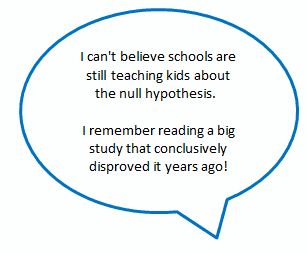
In order to understand P values, you must first understand the null hypothesis.

In every experiment, there is an effect or difference between groups that the researchers are testing. It could be the effectiveness of a new drug, building material, or other intervention that has benefits. Unfortunately for the researchers, there is always the possibility that there is no effect, that is, that there is no difference between the groups. This lack of a difference is called the [null hypothesis](http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/null-and-alternative-hypotheses/), which is essentially the position a devil’s advocate would take when evaluating the results of an experiment.

To see why, let’s imagine an experiment for a drug that we know is totally ineffective. The null hypothesis is true: there is no difference between the experimental groups at the population level.

Despite the null being true, it’s entirely possible that there will be an effect in the sample data due to random sampling error. In fact, it is extremely unlikely that the sample groups will ever exactly equal the null hypothesis value. Consequently, the devil’s advocate position is that the observed difference in the [sample](http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/introductory-concepts/basic-concepts/sample-and-population/) does not reflect a true difference between [populations](http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/introductory-concepts/basic-concepts/sample-and-population/).

**What Are P Values?**

P values evaluate how well the sample data support the devil’s advocate argument that the null hypothesis is true. It measures how compatible your data are with the null hypothesis. How likely is the effect observed in your sample data if the null hypothesis is true?

* High P values: your data are likely with a true null.
* Low P values: your data are unlikely with a true null.

A low P value suggests that your sample provides enough evidence that you can reject the null hypothesis for the entire population.

**How Do You Interpret P Values?**

In technical terms, a P value is the probability of obtaining an effect at least as extreme as the one in your sample data, assuming the truth of the null hypothesis.

For example, suppose that a vaccine study produced a P value of 0.04. This P value indicates that if the vaccine had no effect, you’d obtain the observed difference or more in 4% of studies due to random sampling error.

P values address only one question: how likely are your data, assuming a true null hypothesis? It does not measure support for the alternative hypothesis. This limitation leads us into the next section to cover a very common misinterpretation of P values.

**P Values Are***NOT***the Probability of Making a Mistake**

Incorrect interpretations of P values are very common. The most common mistake is to interpret a P value as the probability of making a mistake by rejecting a true null hypothesis (a [Type I error](http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/type-i-and-type-ii-error/)).

There are several reasons why P values can’t be the error rate.

First, P values are calculated based on the assumptions that the null is true for the population and that the difference in the sample is caused entirely by random chance. Consequently, P values can’t tell you the probability that the null is true or false because it is 100% true from the perspective of the calculations.

Second, while a low P value indicates that your data are unlikely assuming a true null, it can’t evaluate which of two competing cases is more likely:

* The null is true but your sample was unusual.
* The null is false.

Determining which case is more likely requires subject area knowledge and replicate studies.

Let’s go back to the vaccine study and compare the correct and incorrect way to interpret the P value of 0.04:

* **Correct:**Assuming that the vaccine had no effect, you’d obtain the observed difference or more in 4% of studies due to random sampling error.
* **Incorrect:** If you reject the null hypothesis, there’s a 4% chance that you’re making a mistake”

.