

# Likelihood

**Author:** Cole Brookson **Date:** 29 August 2022

It's often the case that we may sample independently from a population, and develop some hypothesis that may indicate that the means for the independent samples are in fact different from one another. Taking the example we noted in the hypothesis testing section, perhaps we care about the size of fiddler crabs, and if the crabs are from multiple sites, we could say that, between two sites of interest, we might believe there to be a size difference. Well, let's perform a statistical test to see if this is true.

It is generally best in statistics, especially when starting our own individual journeys as statisticians, to choose the simplest test that A) will answer our question, and B) we can satisfy the assumptions of.

When comparing two sample means, the standard test to perform here is the unpaired two-sample t-test. The test statistic is calculated as:

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)_0}{SE_{Y_1 - Y_2}}$$

This test assumes three major things:

- 1) Normality
- 2) Homoskedasticity
- 3) Independence

Our question is as it stands is simply that we believe there is a size difference between crabs from different locations. Let's start by attempting to justify these assumptions, and if we can, using the simple test.

## T-test - Two Sided

We'll use data on crab sizes:

There are a number of sites in this dataset, so let's compare crabs from two northern sites and two southern sites. Our northern sites will be "Cape Cod" and "Narragansett Bay NERR" and our southern sites will be "Guana Tolomoto Matanzas NERR" and "Sapelo Island NERR".

It's important to never forget the underlying mathematics of what we're doing, we should never simply rely on our computers exclusively and not understand what we're doing, so let's review the equations underlying the t-test. We'll be calculating the test statistic  $t$  as described above. As we are testing a hypothesis that the sizes are different, we can state this formally as  $H_0 : \mu_1 = \mu_2$  and  $H_1 : \mu_1 \neq \mu_2$ .

Since we stated we believe  $H_0 : \mu_1 = \mu_2$ , this is a two-sided test, we have not suggested that  $\mu_1 > \mu_2$  or  $\mu_1 < \mu_2$ .

## Assumptions

Before we proceed we need to take into account assumptions of our chosen test. We'll assume our samples are independent, but our two-sample t-test requires two other things: 1) normality, and 2) homoskedasticity.

**Normality** To assess the normality of our data, let's first assess it visually and then double-check with a statistical test of normality (the Shapiro-Wilks test). A convenient package for this task is the `ggpubr` package which allows to plot density plots and QQ plots.

Let's use a density plot to check if the data follow the normal/Gaussian curve:

They both look normal. Let's look at the QQ plots:

We're looking for the points to follow the black line and for no points to be outside the grey error bars. Both of these look pretty good. Let's use a Shapiro-Wilk's test to take a look at what we see. Importantly, a Shapiro-Wilk's test is a significance test that has a null hypothesis of normality. If the test is SIGNIFICANT then there is sufficient non-normality to stop our test and re-evaluate.

We can do this in R very easily. The function `shapiro.test()` comes shipped in core R with the stats package. First for the northern sites: Okay, and we see that our p-value is  $> 0.05$  so we'll fail to reject the null. For the southern sites:

Here we see a large p-value,  $\gg 0.05$ , so again we'll reject the null.

Good, we've satisfied the condition of normality .

**Homoskedasticity** Our second assumption was that the variances of the two samples be equal. If they are not (heteroskedasticity), all is not lost and we can use a different version of the t-test. We will test our homoskedasticity with the F-test . Again, similar to the normality test, the null hypothesis of the F-test is that the variances are equal, so if  $p < 0.05$  then for our purposes we could say that the assumption of homoskedasticity is violated.

To do the F-test let's use the stats package function `var.test()`. We will just pass the two sets of data as arguments.

Great,  $p \gg 0.05$ , so we can continue.

### The test

So now we can actually perform the test. It's quite simple, we'll use the `t.test()` function:

We see our result here, and it turns out the p-value is very small, so we can reject the null hypothesis, there is good evidence that  $\mu_1 \neq \mu_2$ . NOTE that if our data were heteroskedastic, then we could specify `var.equal = FALSE` and that would perform a Welch's t-test which allows for this.

If we wanted, we could plot this result as a box plot, or perhaps more usefully, a violin plot (For a full walk through of how to iteratively make this plot, see our section on plotting Values Between Groups):

### T-test - one-sided

In our above example we stated our hypothesis such that we could use a two-sided test. But what if when we were formulating our hypotheses, we'd said we believe the southern crabs would be bigger? Note that we're pretending here that we HAVEN'T seen the results just above, since that would be p-hacking. But let's say we thought "Hey, warmer water, lots of sun, maybe more beaches, that could mean bigger crabs." (I stress this is not an intelligently derived biological hypothesis).

But in this case, we could perform the exact test as above, but instead determine that we wanted a one-sided test. That's easy enough and uses all the same code, with one additional argument. Using the same dataset, we would simply add an argument for `alternative`. The syntax here is saying that in the function, we pass two sets of numbers, `x` and `y`. If we believe `x` to be less than `y`, we pass "less", and if we believe `x` to be greater than `y`, then we pass "greater".

In our case, we are hypothesizing that the southern crabs (`crab_sizes_south`) are greater in size than the northern crabs, so let's set `x` to be the `crab_sizes_south` and the `y` to be `crab_sizes_north`, and then `alternative = "greater"`, so this is testing the alternate hypothesis that  $\mu_{south} > \mu_{north}$ :

So we see here we have a p-value of 1. Given that we do know which group is actually bigger this comes as no surprise. Although this is a good reminder: Just because our one-sided test with an  $H_A : \mu_{south} > \mu_{north}$  resulted in us failing to reject the null, this does NOT mean that we can assume that in fact  $\mu_{south} > \mu_{north}$  because the null hypothesis of these tests is stated that  $H_0 : \mu_{south} = \mu_{north}$ .

### What To Do With Non-Normality?

Going back to our assumptions, let's re-start our process of this test but with a different grouping of data.

What if we just took the MOST northern and southern sites alone, not groups of sites, and did a t-test on those two? say that instead of the two sites acting as our "northern" sites, we chose to use the single

most northern site: “Plum Island Estuary - West Creek” (PIE), and for our southern site, we took the most southern one - “Guana Tolomoto Matanzas NERR” (GTM). If we start again by checking our assumptions, we would first have to check for normality of both our groups.

Let’s start by again dividing our data:

Okay, so these data look somewhat skewed. Let’s do our Shapiro-Wilk’s test:

Okay, so our p-values for both are  $< 0.05$ , so we have to reject the null - these data are non-normal.

Now what? Well there are other tests available to us!! A reminder that parametric tests like the t-test, often have a non-parametric alternative, that usually relax assumptions of normality, and are good for data with small sample sizes. Here we have only 28 observations in each group, and non-normality so let’s turn to the non-parametric tests.

## Non-Parametric Tests

The best alternative for our case is the Mann-Whitney U Test.

There are a few things to note about this test however. This test actually does NOT compare sample means. The Mann-Whitney’s stated use is to determine whether or not the groups or samples are drawn from equal populations, that is, are the shape of the data the same. It can be interpreted that this is analogous to saying the median values are the same for both groups.

## Assumptions

Just because the data do not have to be normal, there are still important assumptions we need to satisfy. The data still need to be independent, but there are two additional assumptions:

- 1) The variable needs to be continuous
- 2) The data are assumed to be similar in shape

The second assumption here is important! While the data do not have to be normal, they have to exhibit a similar skew.

**Continuity** We know that our variable is continuous, so this assumption is met.

**Skew** Measuring a degree of skew is somewhat arbitrary and can be difficult. It is beyond the scope of this tutorial to do so, and so we will satisfy ourselves with simply stating that the density distributions are both left-skewed, and leave it at that.

## The test

The Mann-Whitney is sometimes referred to as the Mann-Whitney-Wilcoxon test, and so is performed using the `wilcox.test()` function, again from stats:

Here we see our p-value is  $\ll 0.05$ , so we can reject our null hypothesis that say that these two medians are equal, and thus conclude (generally) that the northern and southern populations have different sizes. Note again that this is a two-sided test so we’re not testing directionality.