**Analyzing Data**

**Background**

As a scientist, you perform many experiments and generate large amounts of data. How do you know if your results have any practical meaning in the real world? You have to come up with some way to look at your data and determine if your experimental treatment had a different effect on your experimental subjects compared to your control. Looking at your data in this way is called analysis. Luckily, there are a couple of well-established ways to evaluate your results. The first way to examine your data is called quantitative analysis. This type of analysis involves looking at your data and making judgments about it based on what you already know. A much more standardized way of analyzing your data is called qualitative analysis. This type of analysis involves using math to make standardized judgments about your results and is the type of analysis most often used by scientists.

What is data analysis and why is it so important?

* It is a way to mathematically manipulate your raw data
* It allows you to show your results in a way that is easily and universally understood. In science it is important to communicate your data in a way that is universally understood because when you make a discovery or report the results of your experiment, you want everyone who looks at your work to be able to understand your results in the same way.
* It allows you to better understand your data set and interpret your results

**Confidence Intervals**

Often in science, we want to compare two means (or averages). For instance, we may want to answer the question: Are men taller than women? To answer this, we could measure 10 men and 10 women (See Table 1 for raw data). Then we could find the mean height of men by adding up all of the heights we have for men and dividing by the number of men we measured (in this case, 10). We would do the same for the women. As you can see in Table 1, the mean height of the men we measured is 69 inches and the mean height of the women we measured is 64 inches. So now we know the average height of the people *we measured*, but our original questions wasn’t asking if the men we measured would be taller than the women we measured. Our original question was a much broader question about **all** men and women.

The means we just calculated are called sample means because they are calculated from only a portion, or sample, of the entire population of men and women. Because we want to answer our question about the *entire population* of men and women, we need to have some idea about the population means for the height of men and women (i.e. mean heights of **all** men and women on the planet). We can’t realistically measure every person in the world, so we need a way of estimating what the true population means would be.

One method scientists use to determine the true mean of a population is calculating a confidence interval. A confidence interval is a calculated range of values that could contain the true population mean. We can’t determine one single value for the population mean because we didn’t actually measure every single man and woman on the planet, instead we can figure out a range of heights that are reasonable values for the true population mean (a confidence interval).

Table 1. Heights of 10 men and 10 women in inches

|  |  |  |
| --- | --- | --- |
| Men | Women |  |
| 72 | 65 |  |
| 69 | 69 |  |
| 67 | 67 |  |
| 70 | 60 |  |
| 74 | 61 |  |
| 66 | 65 |  |
| 67 | 62 |  |
| 71 | 64 |  |
| 66 | 62 |  |
| 68 | 65 |  |
| 690 | 640 | Sum of heights |
| 69 | 64 | Mean height (Sum of heights/10) |

If we only look at the means (highlighted in the box above), we would say that the means are different, but as scientists we want to know if they are *significantly* different. A significant difference tells us if the means are truly different from a mathematical standpoint. This is important because even though the numbers we calculated are different from one another, because we only measured a small number of individuals, there is some error that needs to be accounted for. Error is something that we calculate (shown below) and then add or subtract from the mean we calculated for each group. If we ended up with a calculated error value of 5 inches for the height of women, we would add and subtract 5 from the mean, 64 inches, to come up with our confidence interval. So with our error included, we end up with a range of heights for women that spans from 59 inches (64 - 5) to 69 inches (64 + 5). Additionally, if we had an error value of 1 inch for the men, our range would be between 68 inches (69-1) and 70 inches (69 +1). Now, with all these overlapping numbers, it is becoming harder to determine if there is a real difference in the heights of men and women, right?

In order to *really* know if our means are mathematically and significantly different, we need to calculate a 95% Confidence Interval for each mean. A 95% Confidence Interval (C.I.) is a calculated range of values that has a 95% chance of containing the true population mean. You might ask why we don’t use a 100% confidence interval. We want to be 100% sure that we are estimating the population mean correctly, right? Unfortunately, using a 100% confidence interval is not a valuable mathematical calculation. The range of values in our confidence interval gets too large to be useful. We would rather have a narrower range of values to estimate our population mean and possibly be wrong 5% of the time than have a huge range of values and be right 100% of the time.

**Calculations for a 95% Confidence Interval**

**Mean**

The first calculation that must be performed in order to obtain a confidence interval is calculating the mean. We have two groups we need to calculate means for – men and women. The mean is calculated by adding all of the observations (in this case the heights) from a group together and then dividing by the number of observations you have for that group.

Example:

**Standard Deviation (SD)**

Next we need to calculate the standard deviation of each group. We won’t use the standard deviation directly to find our confidence intervals, but it is an important step in calculating the error, something we will use to find our confidence intervals. This calculation is just giving us an idea of how far our collected data values stray from the mean.

Example:

**Step 1.** Subtract each data value from the mean of the sample set. This is the deviation value.

**Step 2.** Square each of the deviation values (d).

**Step 3**. Sum the squared deviation values(d2).

 **Men** **Women**

 Data Value Mean Deviation (d) d2 Data Value Mean Deviation (d) d2

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 72 | - | 69 | = |  3 | 9 |  | 65 | - | 64 | = | 1 | 1 |
| 69 | - | 69 | = |  0 | 0 |  | 69 | - | 64 | = | 5 | 25 |
| 67 | - | 69 | = | -2 | 4 |  | 67 | - | 64 | = | 3 | 9 |
| 70 | - | 69 | = |  1 | 1 |  | 60 | - | 64 | = | -4 | 16 |
| 74 | - | 69 | = |  5 | 25 |  | 61 | - | 64 | = | -3 | 9 |
| 66 | - | 69 | = | -3 | 9 |  | 65 | - | 64 | = | 1 | 1 |
| 67 | - | 69 | = | -2 | 4 |  | 62 | - | 64 | = | -2 | 4 |
| 71 | - | 69 | = |  2 | 4 |  | 64 | - | 64 | = | 0 | 0 |
| 66 | - | 69 | = | -3 | 9 |  | 62 | - | 64 | = | -2 | 4 |
| 68 | - | 69 | = | -1 | + 1 |  | 65 | - | 64 | = | 1 | + 1 |

 Sum of d2 = 66 Sum of d2 = 70

**Step 4.** Calculate the Standard Deviation (SD) = $\sqrt{(sum of d^{2})/ (\# of observations – 1)}$

**Men:** SD = $\sqrt{66/ (10 – 1)}$ **Women:** SD = $\sqrt{70/ (10 – 1)}$

 = $\sqrt{66/ 9}$ = $\sqrt{70/ 9}$

 = 2.7080 = 2.7889

**Standard Error (SE)**

As we previously mentioned, error is something that is introduced into every experiment we perform, so it is important that we include it in our calculations. We calculate error by dividing the standard deviation of each group by the square root of the number of observations in the corresponding group.

Example:

**Men:** SE = $2.7080 ÷ \sqrt{10}$ **Women:** SE = $2.7889 ÷ \sqrt{10}$

 = 2.7080 ÷ 3.1623 = 2.7889 ÷ 3.1623

 = 0.8563 = 0.8819

**2 x Standard Error (2xSE)**

This is the shortest calculation you will perform on your way to calculating a confidence interval. You simply multiply the standard error you previously calculated by 2!

Example:

**Men:** 2xSE = 0.8563 x 2 **Women:** 2xSE = 0.8819 x 2

 = 1.7126 = 1.7638

**Lower Confidence Limit (LCL)**

This is the first component of our confidence interval. The lower confidence limit is the lowest number in our range of calculated values where the population mean could truly lie. To calculate the LCL you subtract the 2xSE for each group from the mean of the same group.

Example:

**Men:** LCLMen = MeanMen - 2xSEMen **Women:** LCLWomen = MeanWomen - 2xSEWomen

 = 69 – 1.7126 = 64 – 1.7638

 = 67.2874 ≈ 67 inches = 62.2362 ≈ 62 inches

**Upper Confidence Limit (UCL)**

This is the second component of our confidence interval. The upper confidence limit is the highest number in our range of calculated values where the population mean could truly lie. To calculate the UCL you add the 2xSE for each group to the mean of the same group.

Example:

**Men:** UCLMen = MeanMen + 2xSEMen **Women:** UCLWomen = MeanWomen + 2xSEWomen

 = 69 + 1.7126 = 64 + 1.7638

 = 70.7126 ≈ 71 inches = 65.7638 ≈ 66 inches

**95% Confidence Interval (95% CI)**

We now have all the information we need to find our 95% CI’s for each group! In general, a confidence interval is the range of values between the LCL and UCL for a specific group.

So, the 95% CI for the height of Men is the range of values between 67 inches and 71 inches.

The 95% CI for the height of Women is the range of values between 62 inches and 66 inches.

**Interpreting a 95% Confidence Interval (95% CI)**

Now that we have our confidence intervals, we need to know how to interpret them. Generally when discussing a confidence interval use specific language to ensure that other scientists can understand our results. We typically say: “We are 95% confident that the population mean lies between [insert confidence interval].”

Let’s use this language to discuss the confidence intervals we calculated.

Men: We are 95% confident that the population mean height of men lies between 67

and 71 inches.

Women: We are 95% confident that the population mean height of women lies between 62 and 66 inches.

So what we are really saying with our 95% CI’s is that based on the data we collected, we are confident that if we measured all the men and women in the world their mean heights would be within our confidence intervals.

**Comparing Means Using 95% Confidence Intervals**

Just calculating our confidence intervals doesn’t answer our original question. We now need to use these intervals to compare the means of the two groups to determine if the means are *significantly different* from one another.

We can say two means are significantly different if the 95% CI’s of the two groups do not overlap. This means that no number in the confidence interval of one group can appear in the confidence interval of the other group.

If the numbers within the two 95% CI’s do overlap, we must say that the means are not significantly different.

Let’s look back at our example. Remember that our original question was: “Are men taller than women?” When we calculated the mean heights of men and women they looked different, but we needed a way to scientifically determine if the means were different.

So let’s use our 95% CI’s to answer our question!

95% CI for height of Men: 67 in – 71 in.

95% CI for height of Women: 62 in – 66 in.

We need to know if the numbers in our two confidence intervals overlap. One great way to visualize this is with a number line.

 

By looking at the number line, we can tell that the confidence intervals do not overlap. We can now say that **the mean heights of men and women are significantly different from one another.**

Does this fully answer our question? Not quite. We wanted to know if men are taller than women. Since we know that the mean heights of men and women are significantly different from each other, we need to figure out which group is taller than the other. Look at the number line again and see which group, men or women, has the confidence interval with the highest number values in it. In this case, the confidence interval for the men has higher numbers than women, so **we can say with confidence that men *are* taller than women**.

Another way to visualize our confidence intervals is by graphing them.



In this graph, the boxes are the means of each group and the bars coming off of each box are the confidence intervals.

Do the bars overlap? No. So that means that our confidence intervals do not overlap and the means are significantly different from each other.

Let’s look at a different example for comparing means. Here, our question is do batteries from Group 1 last longer than batteries from Group 2.



By looking at the number line and graph, we can see that the 95% CI’s of the two groups overlap. So, we would have to say that the means of the two groups are *not significantly different*. Once we know that the means of the two groups are not significantly different, we can’t say which group of batteries lasts longer, even if the mean battery life isn’t exactly the same for both groups. This is because when we say something is not significantly different, we are saying that, mathematically, the two means are the same. But how could this be?

We calculated two different means, but our confidence intervals are telling us that the means are effectively the same! This is because our confidence intervals are telling us something about the corresponding *population* means. It makes sense that our individual sample means can be different from one another, while the population means can be the same.

So which is better, *significantly different* or *not significantly different*? This is a trick question - both are equally good results! Just because we don’t find a significant difference between our means doesn’t indicate that our experiment or results aren’t good, it just means that we don’t have the mathematical power to detect a difference between our two means. Sometimes we don’t have the power to detect a difference between our means because we used a small sample size when we were collecting data, other times we can’t see a difference between our means because we introduced too much error into our experiment. But even if we have a gigantic sample size and relatively little error in our experiment, we can still find that there is not a significant difference between our means. This isn’t necessarily a bad thing; sometimes it is good to find that two means are the same. For instance, maybe we originally thought that frogs is one pond had much shorter lifespans than frogs in another pond, but we found that the mean life spans of the frogs in both ponds were not significantly different from one another. We wouldn’t call this a bad result would we? Neither finding (significantly different or not) is inherently better than the other, it all comes down to the context of your question!

**Activity**

A farmer has begun growing corn with a new fertilizer that promises to increase his crop yield. The farmer hires you to test whether or not the new fertilizer really has increased his crop yield over the old fertilizer.

Here is the data he collected (each number represents the crop yield in pounds after harvest):

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| New fertilizer | 67 | 73 | 68 | 66 | 71 | 70 | 67 | 69 | 63 | 66 |
| Old Fertilizer | 61 | 67 | 65 | 62 | 62 | 67 | 62 | 61 | 57 | 56 |

1. Calculate the mean crop yield for the new fertilizer. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ pounds
2. Calculate the mean crop yield for the old fertilizer. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ pounds
3. Calculate the standard deviation (SD) for each fertilizer.

 **New** **Old**

 Data Value Mean Deviation (d) d2 Data Value Mean Deviation (d) d2-

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 67 | - | \_\_\_\_\_\_ | = |  \_\_\_\_\_\_ | \_\_\_\_\_\_ |  | 61 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |
| 73 | - | \_\_\_\_\_\_ | = |  \_\_\_\_\_\_ | \_\_\_\_\_\_ |  | 67 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |
| 68 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |  | 65 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |
| 66 | - | \_\_\_\_\_\_ | = |  \_\_\_\_\_\_ | \_\_\_\_\_\_ |  | 62 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |
| 71 | - | \_\_\_\_\_\_ | = |  \_\_\_\_\_\_ | \_\_\_\_\_\_ |  | 62 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |
| 70 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |  | 67 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |
| 67 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |  | 62 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |
| 69 | - | \_\_\_\_\_\_ | = |  \_\_\_\_\_\_ | \_\_\_\_\_\_ |  | 61 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |
| 63 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |  | 57 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |
| 66 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ |  \_\_\_\_\_\_ |  | 56 | - | \_\_\_\_\_\_ | = | \_\_\_\_\_\_ | \_\_\_\_\_\_ |

 Sum of d2 = Sum of d2 =

Remember:Standard Deviation (SD) = $\sqrt{(sum of d^{2})/ (\# of observations – 1)}$

1. Calculate the standard error (SE) for each fertilizer.
2. Calculate 2 x SE for each fertilizer.
3. Calculate the lower confidence limit (LCL) for each fertilizer.
4. Calculate the upper confidence limit (UCL) for each fertilizer.
5. Write the 95% confidence intervals (95% CI’s) for the mean of each fertilizer.
6. Interpret the 95% CI’s you just found (Hint: use the same wording as we did in the example).
7. Draw a number line and place your 95% CI’s on it like we did in the example.
8. Graph your means and confidence intervals. Remember to label and scale your axes!



1. Do the confidence intervals overlap?
2. Is there a significant difference between the crop yields of the two different fertilizers? If so, does the new fertilizer increase crop yield over the old fertilizer?