19. *Practical*. z- and t- intervals

This lab focuses on applying the concepts from Chapter 17 and Chapter 18 in Jamovi. Specifically, we will practice calculating confidence intervals (CIs). There will be 4 exercises focusing on calculating confidence intervals in Jamovi. To complete the first 2 exercises, you will need the distrACTION module in Jamovi. We downloaded the distrACTION module in the Week 4 practical. If you need to download it again, the instructions to do this are in Chapter 16 Exercise 16.2 (briefly, go to the Modules option and select 'jamovi library', then scroll down until you find the 'distraACTION' module).

The data for this lab are inspired by ongoing work in the Woodland Creation and Ecological Networks (WrEN) project (Fuentes-Montemayor et al., 2022a,b). The Wren project is led by University of Stirling researchers Dr Elisa Fuentes-Montemayor, Dr Robbie Whytock, Dr Kevin Watts, and Prof Kirsty Park (https://www.wren-project.com/). It focuses on questions about what kinds of conservation actions should be prioritised to restore degraded ecological networks.

Figure 19.1.: Images from the WrEN project led by the University of Stirling

The WrEN project encompasses a huge amount of work and data collection from hundreds of surveyed secondary or ancient woodland sites. Here we will focus on observations of tree diameter at breast height (DBH) and grazing to calculate confidence intervals.

19.1. Confidence intervals with distrACTION

First, it is important to download the disrACTION module if it has not been downloaded already. If the distrACTION module has already been downloaded, it should appear in the toolbar of Jamovi (Figure 19.2) If it has not been downloaded, then see the instructions for downloading it with the 'Modules' option (see Figure 19.2) in Exercise 16.2.

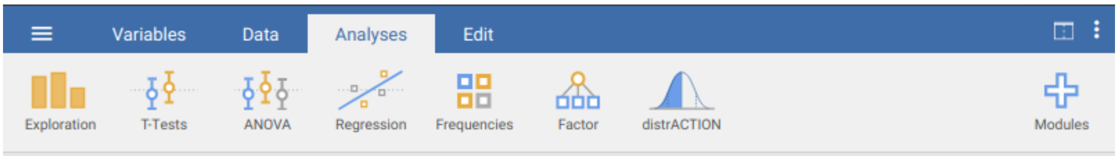

Figure 19.2.: Jamovi tool bar, which includes an added module called distrACTION.

Once the distrACTION module has been made available, download the WrEN trees dataset wren_trees.xlsx dataset and open it in a spreadsheet. Notice that the dataset is not in a tidy format. There are 4 different sites represented by different columns in the dataset. The numbers under each column are measurements of tree diameter at breast height (DBH) in centimeteres. Before doing anything else, it is therefore necessary to put the WrEN dataset into a tidy format. The tidy dataset should include two columns, one for site and the other for DBH.

Once the WrEN trees dataset has been reorganised into a tidy format, save it as a CSV file and open it in Jamovi. In Jamovi, go to Exploration and Descriptives in the toolbar and build a histogram that shows the distribution of DBH. Do these data appear to be roughly normal? Why or why not?

Next, calculate the grand mean and standard deviation of tree DBH (i.e., the mean and standard deviation of trees across all sites).

Grand Mean: ____

Grand Standard Deviation: __________________________

We will use this mean and standard deviation to compute quantiles and obtain 95% z-scores. First, click on the distrACTION icon in the toolbar (see Figure 19.2). From the distrACTION pulldown menu, select 'Normal Distribution'. To the left, you should see boxes to input parameter values for the mean and standard deviation (SD). Below the Parameters options, you should also see different functions for computing probability or quantiles. To the right, you should see a standard normal distribution (i.e., a normal distribution with a mean of 0 and a standard deviation of 1).

For this exercise, we will assume that the population of DBH from which our sample came is normally distributed. In other words, if we somehow had access to *all possible* DBH measurements in the woodland sites (not just the 120 trees sampled), we assume that DBH would be normally distributed. To find the probability of sampling a tree within a given interval of DBH (e.g., greater than 30), we therefore need to build this distribution with the correct mean and standard deviation. We do not know the *true* mean (μ) and standard deviation (σ) of the population, but our best estimate of these values are the mean ($\bar{x}$) and standard deviation (s) of the sample, as reported above (i.e., the grand mean and standard deviation). Using the Mean and SD parameter input boxes in distrACTION, we can build a normal distribution with the same mean and standard deviation as our sample. Do this now by inputting the calculated Grand Mean and Grand Standard Deviation from above in the appropriate boxes. Note that the normal distribution on the right has the same shape, but the table of parameters has been updated to reflect the mean mean and standard deviation.

In the previous practical from Chapter 16, we calculated the probability of sampling a value within a given interval of the normal distribution. If we wanted to do the same exercise here, we might find the probability of sampling a DBH < 30 using the Compute probability function (the answer is P = 0.265). Instead, we are now going to do the opposite using the Compute quantile(s) function. We might want to know, for example, what 75% of DBH values will be less than (i.e., what is the cutoff DBH, below which DBH values will be lower than this cutoff with a probability of 0.75). To find this, uncheck the 'Compute probability' box and check the 'Compute quantile(s)' box. Make sure that the 'cumulative quantile' radio button is selected, then set p = 0.75 (Figure 19.3).

From Figure 19.3, we can see that the cumulative 0.75 quantile is 44.3, so if DBH is normally distributed with the mean and standard deviation calculated above, 75% of DBH values in a population will be below 36.1 cm. Using the same principles, what is the cumulative 0.4 quantile for the DBH data?

Quantile: ________________________________ cm

We can also use the Compute quantile(s) option in Jamovi to compute interval quantiles. For example, if we want to know the DBH values within which 95% of the probability density is contained, we can set p = 0.95, then select the radio button 'central interval quantiles'. Do this for the DBH data. From the Results table on the right, what interval of DBH values will contain 95% of the probability density around the mean?

Interval: ________________________________ cm

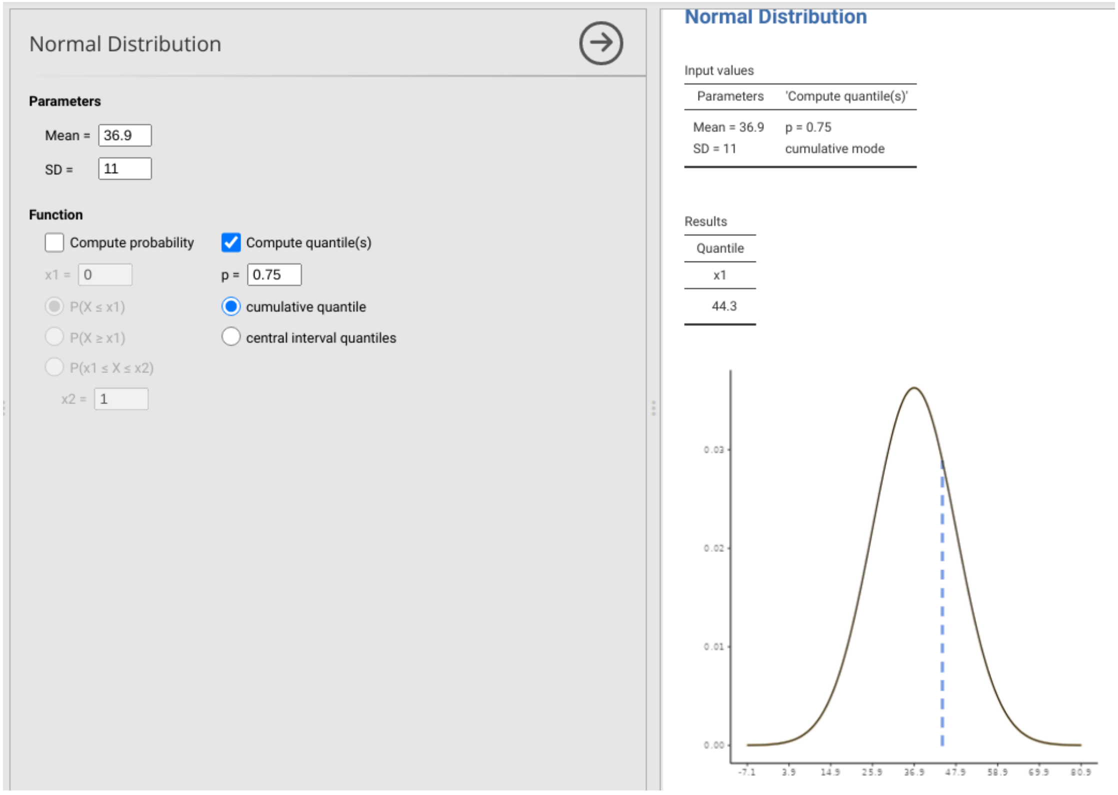

Figure 19.3.: Jamovi interface for the 'distrACTION' module, in which quantiles have been computed to find the diameter at breast height (DBH) below which 75 per cent of DBHs will be given a normal distribution with a mean of 36.9 and standard deviation of 11. Data for these parameter values were collected from in Scotland as part of the Woodland Creation and Ecological Networks (WrEN) project.

Remember that we are looking at the full sample distribution of DBH. That is, getting intervals for the probability of sampling DBH values around the mean, *not* confidence intervals around the mean as introduced in Chapter 17. How would we get confidence intervals around the mean? That is, what if we want to say that we have 95% confidence that the *mean* lies between 2 values? We would need to use the standard deviation *of the sample mean* $\bar{x}$ around the true mean μ, rather than the sample standard deviation. Recall from Chapter 12.6 that the standard error is the standard deviation of $\bar{x}$ values around μ. We can therefore use the standard error to calculate confidence intervals around the mean value of DBH. From the Descriptives panel in Jamovi (recall that this is under the 'Exploration' button), find the standard error of DBH,

Std. error of Mean: ____________________________

Now, go back to the distrACTION Normal Distribution and put the DBH mean into the parameters box as before. But this time, put the standard error calculated above into the box for SD. Next, choose the Compute quantile(s) option and set p = 0.95 to calculate a 95% confidence interval. Based on the Results table, what can you infer are the lower and upper 95% confidence intervals (CIs) around the mean?

Lower 95% CI: _______________________

Upper 95% CI: _______________________

Remember that this assumed that the sample means ($\bar{x}$) are normally distributed around the true mean (μ). But as we saw in Chapter 18, when we assume that our sample standard deviation (s) is the same as the population standard deviation (σ), then the shape of the normal distribution will be at least a bit off. Instead, we can get a more accurate estimate of CIs using a t-distribution. Jamovi usually does this automatically when calculating CIs outside of the distrACTION module. To get 95% CIs, go back to the Descriptives panel in Jamovi, then choose DBH (cm) as variable. Scroll down to the Statistics options and check 'Confidence interval for Mean' under the **Mean Dispersion** options, and make sure that the number in the box is 95 for 95% confidence. Confidence intervals will appear in the Descriptives table on the right. From this Descriptives table now, write the lower and upper 95% CIs below.

Lower 95% CI: _______________________

Upper 95% CI: _______________________

You might have been expecting a bit more of a difference, but remember, for sufficiently large sample sizes (around N = 30), the normal and t-distributions are very similar (see Chapter 18). We really do not expect much of a difference until sample sizes become small, which we will see in Exercise 19.3.

19.2. Confidence intervals from z- and t-scores

While Jamovi can be very useful for calculating confidence intervals from a dataset, you might also need to calculate CIs from just a set of summary statistics (e.g., the mean, standard error, and sample size). This activity will demonstrate how to calculate CIs from z- and t-scores. Recall the formula for lower and upper confidence intervals from Chapter 17.1,

$$LCI = \bar{x} - (z \times SE),$$

$$UCI = \bar{x} + (z \times SE).$$

We could therefore calculate 95% confidence intervals for DBH with just the sample mean ($\bar{x}$), z-score (z), and standard error (SE). We have already calculated $\bar{x}$ and SE for the DBH in Exercise 19.1 above, so we just need to figure out z. Recall that z-scores are *standard normal deviates*; that is, deviations from the mean given a standard normal distribution, in which the mean equals 0 and standard deviation equals 1. For example, $z = -1$ is 1 standard deviation below the mean of a standard normal distribution, and $z = 2$ is 2 standard deviations above the mean of a standard normal distribution. What values of z contain 95% of the probability density of a standard normal distribution? We can use the distrACTION module again to find this out. Select 'Normal Distribution' from the pulldown of the distrACTION module. Notice that by default, a standard normal distribution is already set (Mean = 0 and SD = 1). All that we need to do now is compute quantiles for p = 0.95. From these quantiles, what is the proper z-score to use in the equations for LCI and UCI above?

z-score: ________________________

Now, use the values of $\bar{x}$, z, and SE for DBH in the equations above to calculate lower and upper 95% confidence intervals again.

Lower 95% CI: ________________________

Upper 95% CI: ________________________

Are these confidence intervals the same as what you calculated in Exercise 19.1?

Lastly, instead of using the z-score, we can do the same with a t-score. We can find the appropriate t-score from the t-distribution in the distrACTION module. To get the t-score, click on the distrACTION module button and choose 'T-Distribution' from the pulldown. To get quantiles with the t-distribution, we need to know the degrees of freedom (df) of the sample. Chapter 18 explains how to calculate df from the sample size N. What are the appropriate df for DBH?

df: ____________________________

Put the df in the Parameters box. Ignore the box for lambda (λ); this is not needed. Under the **Function** options, choose 'Compute quantile(s)' as before to calculate Quantiles. From the Results table, what is the proper t-score to use in the equations for LCI and UCI?

t-score: _________________________

Again, use the values of $\bar{x}$, t, and SE for DBH in the equations above to calculate lower and upper 95% confidence intervals.

Lower 95% CI: __________________________

Upper 95% CI: __________________________

How similar are the estimates for lower and upper CIs when using z- versus t-scores. Reflect on any similarities or differences that you see in all of these different ways of calculating confidence intervals

19.3. Confidence intervals for different sample sizes (t- and z-)

In Exercises 19.1 and 19.2, the sample size of DBH was fairly large (N = 120). Now, we will calculate confidence intervals for the mean DBH of each of the 4 different sites using both z- and t-scores. These sites have much different sample sizes. From the Descriptives tool in Jamovi, write the sample sizes for DBH split by site below.

Site 1182: N = _______________

Site 1223: N = _______________

Site 3008: N = _______________

Site 10922: N = _______________

For which of these sites would you predict CIs calculated from z-scores versus t-scores to differ the most?

Site: _______________________

The next part of this exercise is self-guided. In Exercises 19.1 and 19.2, you used different approaches for calculating 95% CIs from the normal and t-distributions. Now, fill in the table below reporting 95% CIs calculated using each distribution from the 4 sites using any method you prefer.

Table 19.1.: 95 per cent Confidence intervals calculated for tree diameter at breast height (DBH) in cm. Data for these parameter values were collected in Scotland as part of the Woodland Creation and Ecological Networks (WrEN) project.

Site	N	95% CIs (Normal)	95% CIs (t-distribution)
1182			
1223			
3008			
10922			

Next, do the same, but now calculate 99% CIs instead of 95% CIs.

Table 19.2.: 99 per cent Confidence intervals calculated for tree diameter at breast height (DBH) in cm. Data for these parameter values were collected in Scotland as part of the Woodland Creation and Ecological Networks (WrEN) project.

Site	N	99% CIs (Normal)	99% CIs (t-distribution)
1182			
1223			
3008			
10922			

What do you notice about the difference between CIs calculated from the normal distribution versus the t-distribution across the different sites?

In your own words, based on this practical and what you have read from the lab workbook and any other material, what do these confidence intervals *actually mean*?

We will now move on to calculating confidence intervals for proportions.

19.4. Proportion confidence intervals

We will now try calculating confidence intervals for proportional data using the WrEN Sites dataset, which you can download here (right click and "Save Link As..."). Notice that there are more sites included than there were in the dataset used in Exercises 19.1-19.3, and that some of these sites are grazed while others are not (column 'Grazing'). From the Descriptives options, find the number of sites grazed versus not grazed (hint, remember from the lab practical in Chapter 16 to put 'Grazing' in the variable box and click the 'Frequency tables' checkbox).

Grazed: ____________________

Not Grazed: ______________________

From these counts above, what is the estimate ($\hat{p}$) of the proportion of sites that are grazed?

$\hat{p}$: _________________

Chapter 17.2 explained how to calculate lower and upper CIs for binomial distributions (i.e., proportion data). To do this, we can use equations similar to the ones used for LCI and UCI from Exercise 19.2 above,

$$LCI = \hat{p} - z \times SE(p),$$

$$UCI = \hat{p} + z \times SE(p),$$

We have already calculated $\hat{p}$, and we can find z-scores for confidence intervals in the same way that we did in Exercise 19.2 (i.e., the z-scores associated with 95% confidence intervals do not change just because we are working with proportions). All that leaves for calculating LCI and UCI are the standard errors of the proportions. Remember from Chapter 17.2 that these are calculated differently from a standard deviation of continuous values such as diameter breast height. The formula for standard error of a proportion is,

$$SE(p) = \sqrt{\frac{p\left(1-p\right)}{N}}.$$

We can estimate p using $\hat{p}$, and N is the total sample size. Using the above equation, what is the standard error of p?

SE(p): ____________________

Using this standard error, what are the lower and upper 95% confidence intervals around $\hat{p}$?

$LCI_{95\%} =$ ____________________

$UCI_{95\%} =$ ____________________

Next, find the lower and upper 99% CIs around $\hat{p}$ and report them below (hint: the only difference here from the calculation of the 95% CIs are the z-scores).

$LCI_{99\%} =$ ____________________

$UCI_{99\%} =$ ____________________

19.5. Another proportion confidence interval

If you have sufficient time during the lab practical, try one more proportional confidence interval. This time, find the 80%, 95%, and 99% CIs for the proportion of sites that are classified as Ancient woodland. First consider an 80% CI (hint, use the distrACTION module again to find the z-scores).

$LCI_{80\%} =$ ____________________

$UCI_{80\%} =$ ____________________

Next, calculate 95% CIs for the proportion of sites classified as Ancient woodland.

$LCI_{95\%} =$ ____________________

$UCI_{95\%} =$ ____________________

Finally, calculate 99% CIs for the proportion of sites classified as Ancient woodland.

$LCI_{99\%} =$ ____________________

$UCI_{99\%} =$ ____________________

Reflect again on what these values actually mean. For example, what does it mean to have 95% confidence that the proportion of sites classified as Ancient woodland are between two values? Are there any situations in which this might be useful, from a scientific or conservation standpoint? There is no right or wrong answer here, but confidence intervals are very challenging to understand conceptually, so having now done the calculations to get them, it is a good idea to think again about what they mean.