Power and Sample Size Calculation – Software workflow for Statulator and PS

To access these online calculators:

- Statulator
- Power and Sample Size (PS)
- Note: Results generated in Statulator and PS may vary from your output in other sample size calculation software.

1. Difference between 2 means

Example: Chicken welfare – Bone density

The bone density of chickens is an important indication of their welfare. You want to test to see if (mineral) bone density can be improved from 120 to at least 130 mg/cm^3.

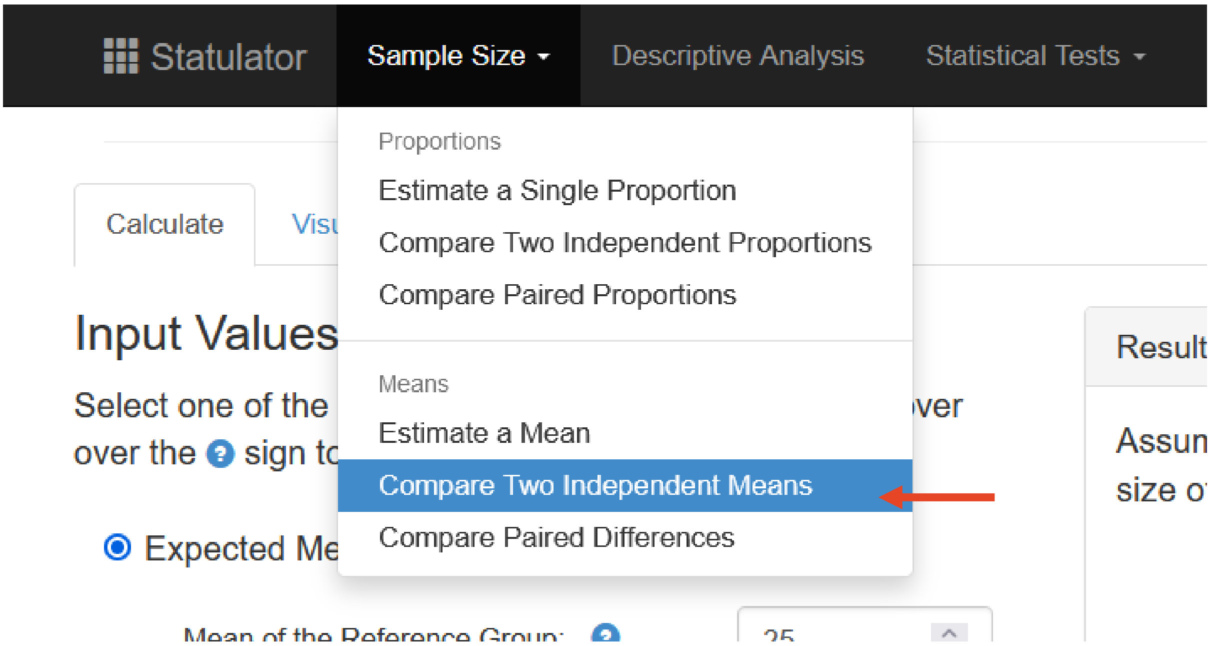
- Treatment group (high mineral diet)
- Control group (normal diet)
- Response variable: Measure the tibia bone density after 6 weeks growth

Question 1: How many chickens do you need to detect a difference in bone density of 10mg/cm^3?

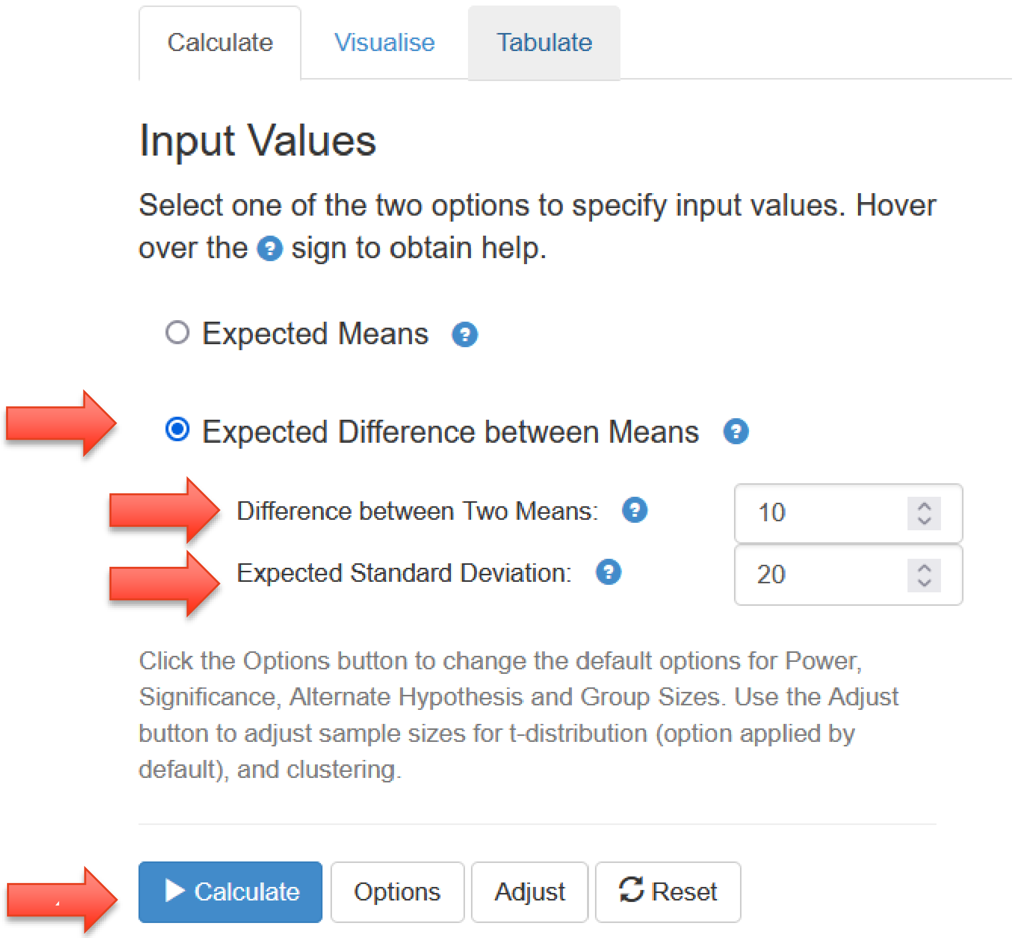
Step 1	Determine experiment type and statistical test	T-test (assume normality)
Step 2	Set α and $1 - \beta$	$\alpha = 0.05$ and $1 - \beta = 0.8$
Step 3	Set the smallest effect size of interest (SESOI)	SESOI = 10mg/cm^3
Step 4	Estimate the variance	You know from previous studies what the typical variation in bone density is for the control diet. You don't know about the treatment diet. You will use an estimate from the control diet of SD = 20mg/cm^3. Assume equal group sizes, n1 = n2.
Step 5	Calculate the minimum sample size	Put all the information into Statulator, and use Statulator defaults.
Step 6	Explore scenarios	Consider how much your within-study standard deviation could vary from your point estimate. You will also run the sample size calculation using SD = 15 (possible min value; optimistic) and SD = 30 (possible max value; pessimistic). You can also plot SD=10, SESOI = 5mg/cm^3 (possible min value; pessimistic), 15mg/cm^3 (possible max value, optimistic)

Methods:

1. Select Sample Size > Compare Two Independent Means



2. Enter the values for the chick experiment



3. Click OK. Your output should be as follows:

Results and Live Interpretation Download

Assuming a pooled standard deviation of 20 units, the study would require a sample size of:

63

for each group (i.e. a total sample size of 126, assuming equal group sizes), to achieve a power of 80% and a level of significance of 5% (two sided), for detecting a true difference in means between the test and the reference group of 10 units.

In other words, if you select a random sample of 63 from each population, and determine that the difference in the two means is 10 units, and the pooled standard deviation is 20 units, you would have 80% power to declare that the two groups have significantly different means, i.e. a two sided p-value of less than 0.05.

Reference: Dhand, N. K., & Khatkar, M. S. (2014). Statulator: An online statistical calculator. Sample Size Calculator for Comparing Two Independent Means. Accessed 13 January 2025 at http://statulator.com/SampleSize/ss2M.html

Note: Statulator used the input values of a power of 80%, a two sided level of significance of 5% and equal group sizes for sample size calculation and adjusted the sample size for t-distribution. You may change the options by clicking here or the 'Options' button and the adjustments by clicking here or the 'Adjust' button.

4. Repeat the process using SD = 15 mg/cm^3.(optimistic scenario) Your output should be as follows:

Results and Live Interpretation Download

Assuming a pooled standard deviation of 15 units, the study would require a sample size of:

36

for each group (i.e. a total sample size of 72, assuming equal group sizes), to achieve a power of 80% and a level of significance of 5% (two sided), for detecting a true difference in means between the test and the reference group of 10 units.

In other words, if you select a random sample of 36 from each population, and determine that the difference in the two means is 10 units, and the pooled standard deviation is 15 units, you would have 80% power to declare that the two groups have significantly different means, i.e. a two sided p-value of less than 0.05.

Reference: Dhand, N. K., & Khatkar, M. S. (2014). Statulator: An online statistical calculator. Sample Size Calculator for Comparing Two Independent Means. Accessed 10 February 2025 at http://statulator.com/SampleSize/ss2M.html

Note: Statulator used the input values of a power of 80%, a two sided level of significance of 5% and equal group sizes for sample size calculation and adjusted the sample size for t-distribution. You may change the options by clicking here or the 'Options' button and the adjustments by clicking here or the 'Adjust' button.

5. Repeat the process using SD = 30 mg/cm^3.(pessimistic scenario). Your output should be as follows:

Results and Live Interpretation — Download

Assuming a pooled standard deviation of 30 units, the study would require a sample size of:

142

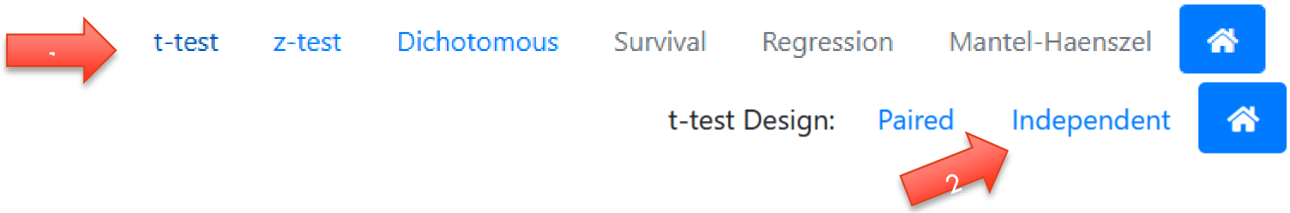
for each group (i.e. a total sample size of 284, assuming equal group sizes), to achieve a power of 80% and a level of significance of 5% (two sided), for detecting a true difference in means between the test and the reference group of 10 units.

In other words, if you select a random sample of 142 from each population, and determine that the difference in the two means is 10 units, and the pooled standard deviation is 30 units, you would have 80% power to declare that the two groups have significantly different means, i.e. a two sided p-value of less than 0.05.

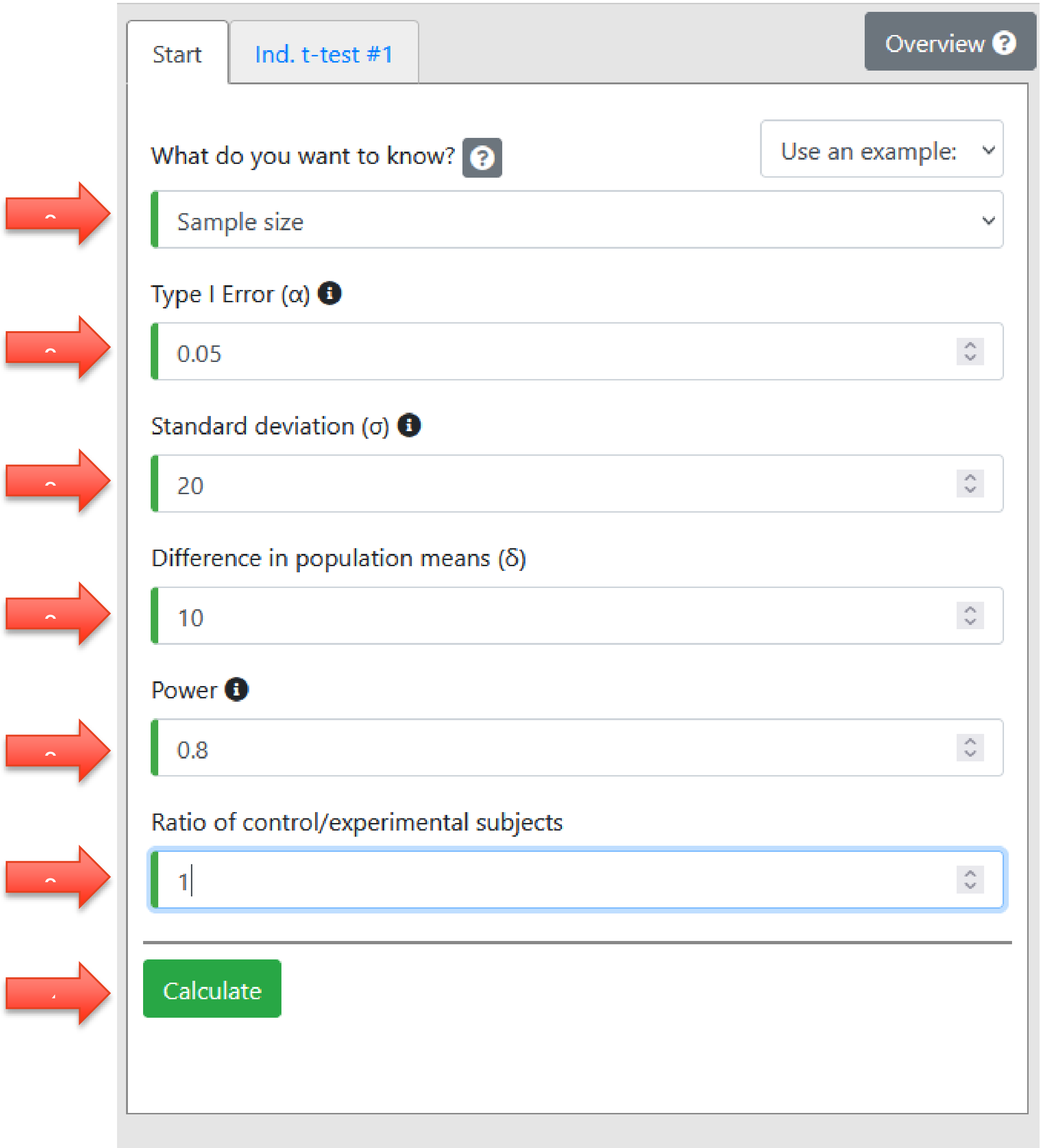
Reference: Dhand, N. K., & Khatkar, M. S. (2014). Statulator: An online statistical calculator. Sample Size Calculator for Comparing Two Independent Means. Accessed 10 February 2025 at http://statulator.com/SampleSize/ss2M.html

Note: Statulator used the input values of a power of 80%, a two sided level of significance of 5% and equal group sizes for sample size calculation and adjusted the sample size for t-distribution. You may change the options by clicking here or the 'Options' button and the adjustments by clicking here or the 'Adjust' button.

6. For a specified SESOI and SD, you can compare sample sizes for different levels of power. To include these on the same plot we need to switch to PS (Power and Sample Size). Go to PS, and select the appropriate sample size calculation type



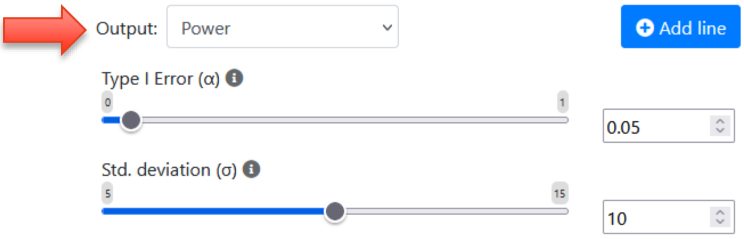
Then enter the same values as you did in Statulator (including Statulator defaults as shown below)



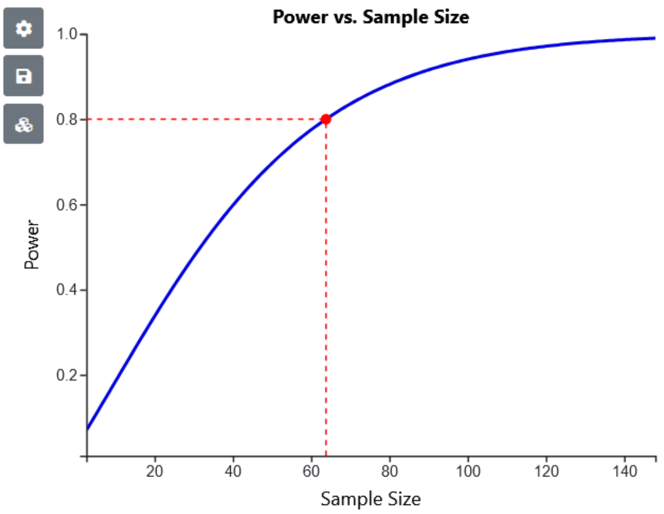
7. Click Calculate. Your output should be as follows:

"We are planning a study of a continuous response variable from independent control and experimental subjects with 1.00 control(s) per experimental subject. In a previous study the response within each subject group was normally distributed with standard deviation 20.00. If the true difference in the experimental and control means is 10.00, we will need to study 64 experimental subjects and 64 control subjects to be able to reject the null hypothesis that the population means of the experimental and control groups are equal with probability (power) 0.80. The Type I error probability associated with this test of this null hypothesis is 0.05."

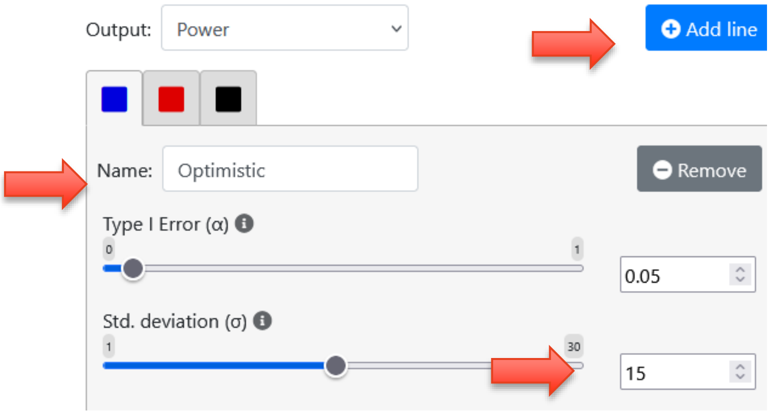
8. You will also see a series of plots. We will change the axes of the top left plot. Find the 'Output' drop-down and select 'Power'



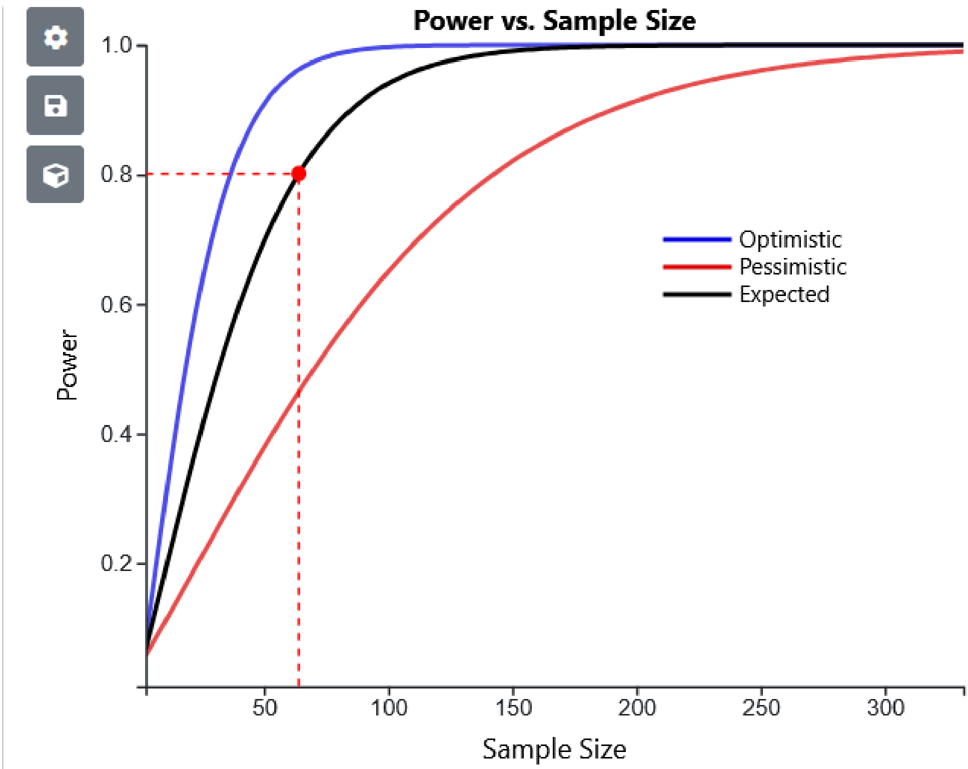
You will see the plot change:



We can add two more lines to see the same relationship for our 'Optimistic' and 'Pessimistic' scenario. Change the Std. deviation to =15 (Optimistic) and repeat process for =30 (Pessimistic). Add appropriate Name (label) for each scenario



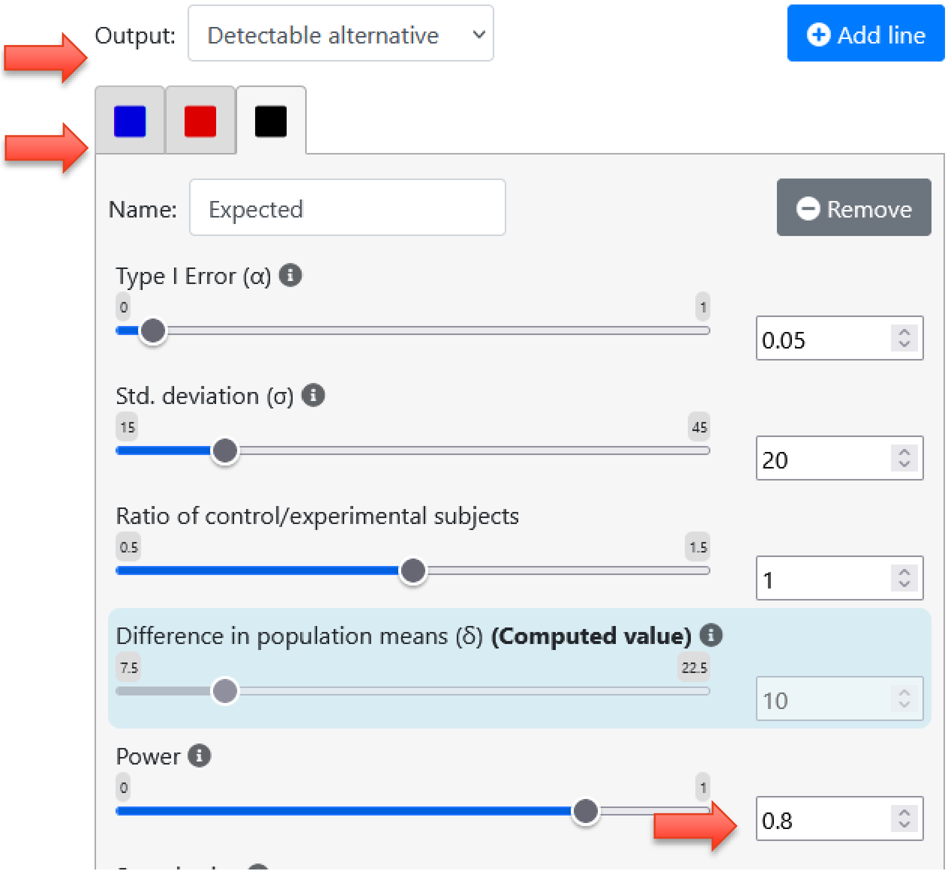
The resulting plot should look like this:



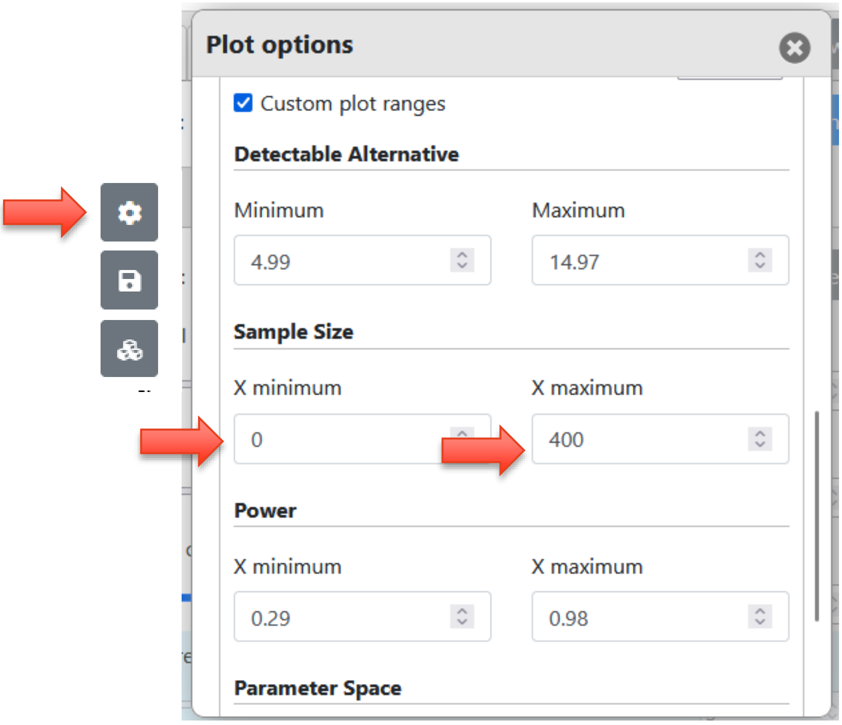
Note that you can move your cursor over the curve to see a pop-up display of the exact value of Sample Size and Power at each point on the curve.

We can also plot the sample size vs. detectable effect size for a fixed value of power

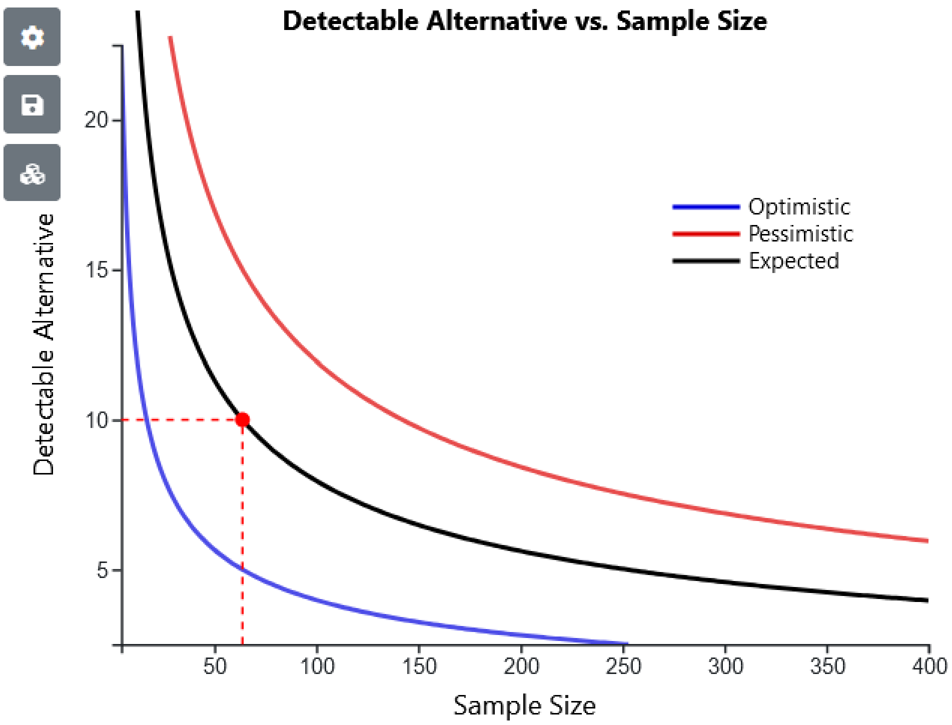
9. Change the output to ‘Detectable alternative’ The software will try to keep the curves consistent, and to do so change the Power. Click through each tab and for each of the lines, ensure the power is set to 0.8 (80%)



10. The x-axis gets a little bit out of hand at this point. So choose the gear wheel at the top left of the plots, scroll down and change the Sample Size X [axis] minimum to 0 and maximum to 400

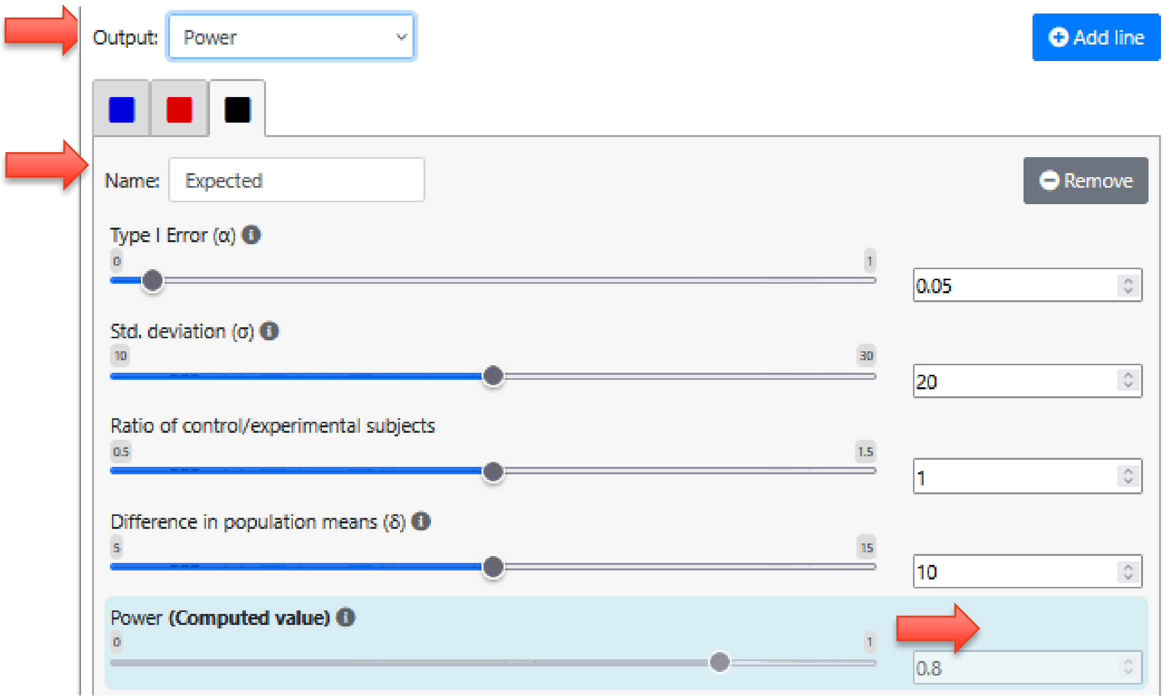


The resulting plots should look like this

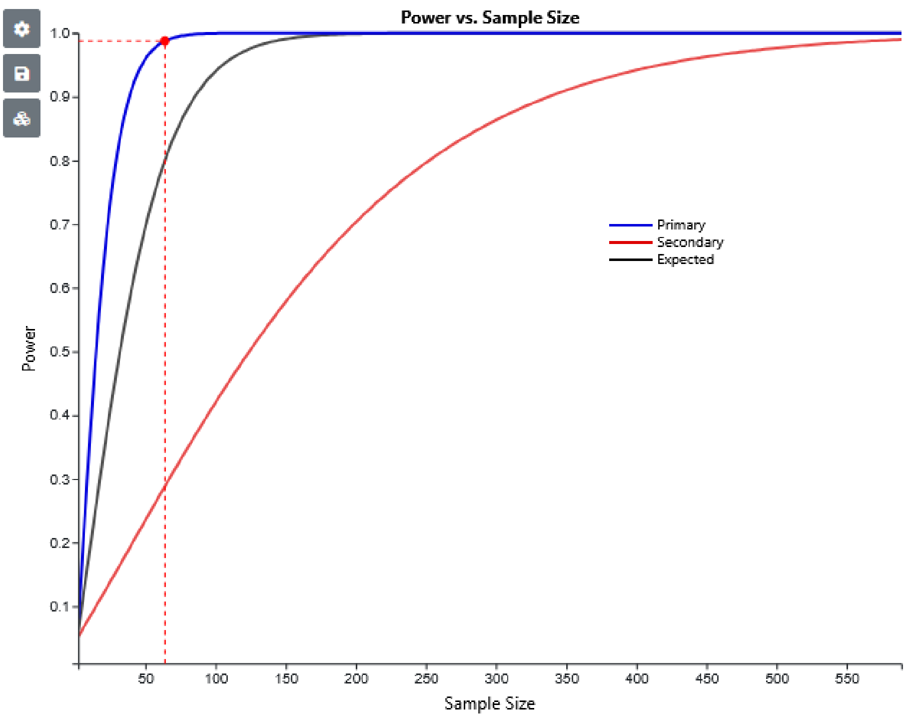


1. We can also investigate our optimistic and pessimistic scenarios for effect size. 3 scenarios for variance x 3 scenarios for effect size = 9 different scenarios, which is too many lines for this graph (only five different colours are supported). Instead, we will fix the variance at our expected 20 mg/cm^3, and evaluate the three effect-size scenarios. Change the output back to power, click through the three lines and change the 'Difference in population means to'

   Expected: 10 Optimistic: 15 Pessimistic: 5



Your plot will look like this:



Difference between 2 means (Mann-Whitney)

Example: Happiness Survey

You want to measure happiness using the Lyubomirsky & Lepper scale. Each item response ranges from 1 (unhappy) to 7 (happy). The score is the sum of 4 items, so the range is 4~28.

A pilot study on two groups produced the following results that can be used for the power calculation:

	Values		Ranks	
	Single	Married	Single	Married
	12	20	3	1
	11	15	4	2
	10	9	5	6
	6	8	8	7
Avg	9.8	13.0	5	4
SD	2.6	5.6		

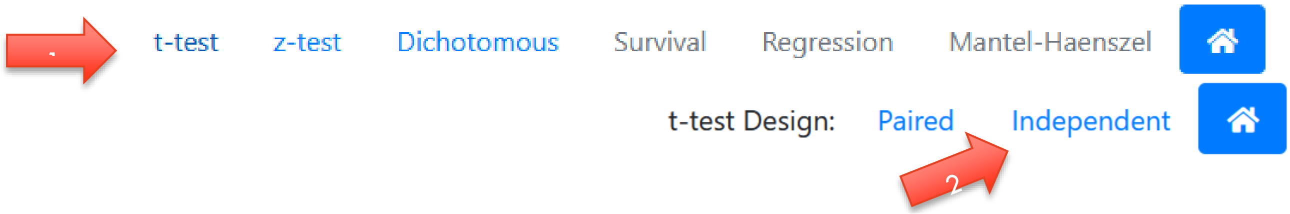
You want to apply the happiness survey to different groups of people (e.g. single vs. married) to see if there is a difference in scores.

Question 2: What is a meaningful difference?

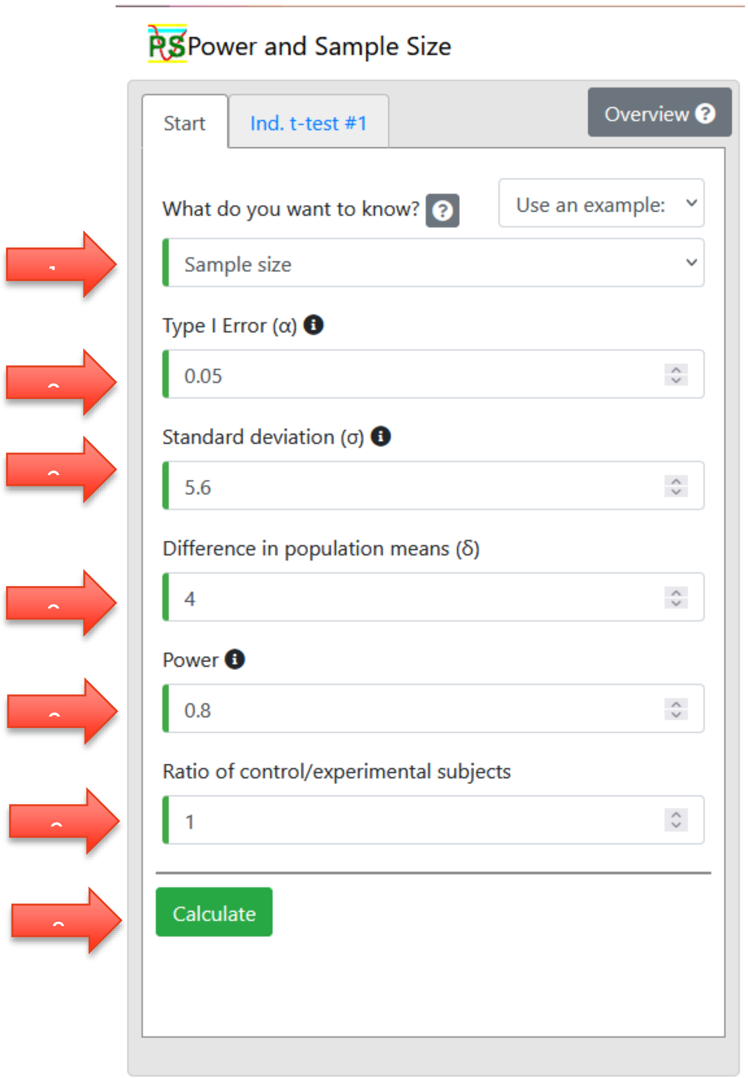
Step 1	Determine experiment type and statistical test	Mann-Whitney (also called Wilcoxon rank sum)
Step 2	Set α and $1 - \beta$	$\alpha = 0.05$ and $1 - \beta = 0.8$
Step 3	Set the smallest effect size of interest (SESOI)	SESOI = 4 points
Step 4	Estimate the variance	SD1 = 2.6 and SD2 = 5.6
Step 5	Calculate the minimum sample size	Heuristic method – Do the calculations as if performing the corresponding parametric test (i.e., the t-test), then add 15% to the sample size. Note: There are currently no options in Statulator or PS to run sample size calculations for non-parametric tests, so the heuristic method will be used.

Methods:

1. Go to PS, and select the appropriate sample size calculation type



2. Enter the values for the happiness experiment. For now we will use the group with the larger SD (SD2) as the '[Within group] Standard deviation'



3. Click OK. Your output should be as follows:

"We are planning a study of a continuous response variable from independent control and experimental subjects with 1.00 control(s) per experimental subject. In a previous study the response within each subject group was normally distributed with standard deviation 5.60. If the true difference in the experimental and control means is 4.00, we will need to study **32 experimental subjects** and 32 control subjects to be able to reject the null hypothesis that the population means of the experimental and control groups are equal with probability (power) 0.80. The Type I error probability associated with this test of this null hypothesis is 0.05."

4. Add 15% for non-parametric. N = 32 x 1.15 = 36.8. Since you cannot have 0.8 of a person, you should round up the sample size to 37 people per group.

5. Using the group with the larger SD is a conservative choice, but it may be overly conservative. The pooled standard deviation from both groups can be calculated as

$$\sigma' = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}$$

Which in this example is:

$$\sqrt{\frac{2.6^2+5.6^2}{2}} = 4.37$$

Entering 4.37 as the SD yields this output:

"We are planning a study of a continuous response variable from independent control and experimental subjects with 1.00 control(s) per experimental subject. In a previous study the response within each subject group was normally distributed with standard deviation 4.37. If the true difference in the experimental and control means is 4.00, we will need to study **20 experimental subjects** and 20 control subjects to be able to reject the null hypothesis that the population means of the experimental and control groups are equal with probability (power) 0.80. The Type I error probability associated with this test of this null hypothesis is 0.05. "

Add 15% for non-parametric. N = 20 x 1.15 = 23.

3. Difference between 2 proportions

Example: Happiness Survey

The survey scores could also be analysed as proportions by considering how many report a value above a threshold (say >14 means "happy").

- Singles group: P1 = proportion of subjects who respond "happy"
- Married group: P2 = proportion of subjects who respond "happy"

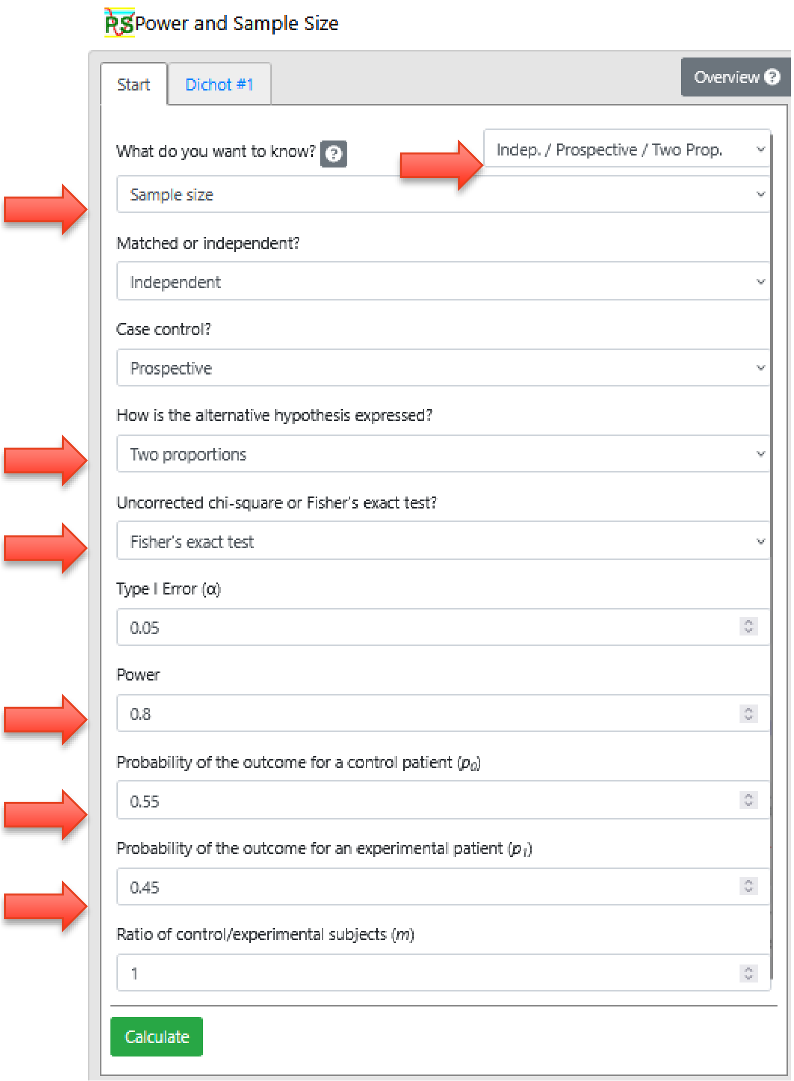
Question 3: Say you want to find a minimum difference in proportions of P1 – P2 = 0.1. What sample size is required?

Step 1	Determine experiment type and statistical test	Fisher's Exact test (should be used when sample sizes are going to be small but can also be used for larger sample sizes providing you with a more conservative estimate).
Step 2	Set α and $1 - \beta$	$\alpha = 0.05$ and $1 - \beta = 0.8$
Step 3	Set the smallest effect size of interest (SESOI)	SESOI = P1 – P2 = 0.1
Step 4	Estimate the variance	You also need to estimate the two proportions. Let's first assume that there will be maximum variance (p = 0.50), so let's try using P1 = 0.55 and P2 = 0.45.
Step 5	Calculate the minimum sample size	Put all the information into PS.
Step 6	Explore scenarios	You will also run the sample with P1 = 0.85 and P2 = 0.95.

Methods:

1. Analyze > Power Analysis > Proportions > Independent-Samples Binomial Test

Enter as below



2. Click 'Calculate'. Your output should be as follows:

"We are planning a study of independent cases and controls with 1 control(s) per case. Prior data indicate that the failure rate among controls is 0.55. If the true failure rate for experimental subjects is 0.45, we will need to study 4**11 experimental subjects** and 411 control subjects to be able to reject the null hypothesis that the failure rates for experimental and control subjects are equal with probability (power) 0.80. The Type I error probability associated with this test of this null hypothesis is 0.05. We will use a continuity-corrected chi-squared statistic or Fisher's exact test to evaluate this null hypothesis."

3. Repeat the process using P1 = 0.85 and P2 = 0.95. Your output should be as follows:

"We are planning a study of independent cases and controls with 1 control(s) per case. Prior data indicate that the failure rate among controls is 0.85. If the true failure rate for experimental subjects is 0.95, we will need to study 159 experimental subjects and 159 control subjects to be able to reject the null hypothesis that the failure rates for experimental and control subjects are equal with probability (power) 0.80. The Type I error probability associated with this test of this null hypothesis is 0.05. We will use a continuity-corrected chi-squared statistic or Fisher's exact test to evaluate this null hypothesis."