



GraphPad

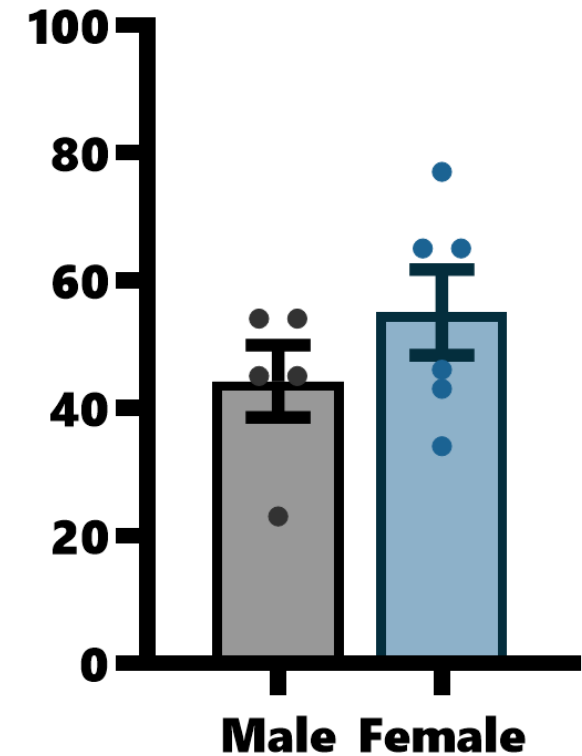
The Null Hypothesis and basic GraphPad 9 use




Robert Midence, PhD

Introduction

- Measured a variable in two groups
- Means are different.... is that by chance?



The background is a light beige color. It features several large, semi-circular decorative elements: a light blue one with a pattern of small dashes in the top left, a white one with three red stars in the top center, and a light green one with a wavy line pattern in the top right. There are also smaller solid circles in shades of blue, purple, and teal scattered around. In the bottom left, there is a light teal semi-circle containing two strings of blue and dark blue triangular bunting flags.

**I don't care just tell me
how to get my stars!!**

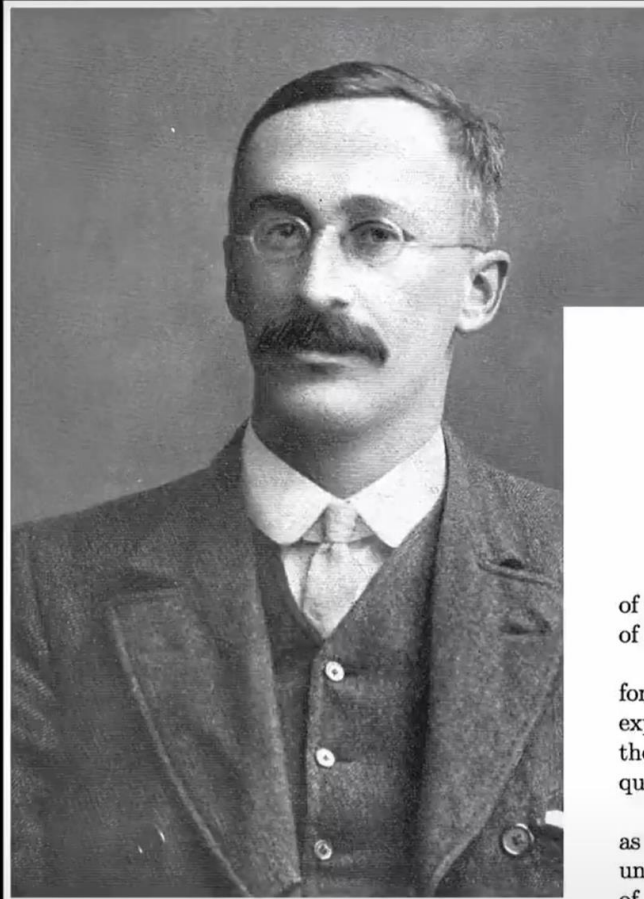


Sources

- Bozeman Science
- Prism Academy
- StatQuest



The Student t-test



William Sealy Gosset



THE PROBABLE ERROR OF A MEAN

By STUDENT

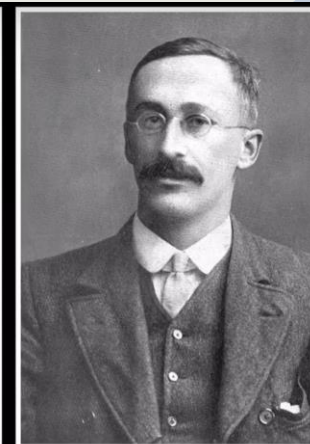
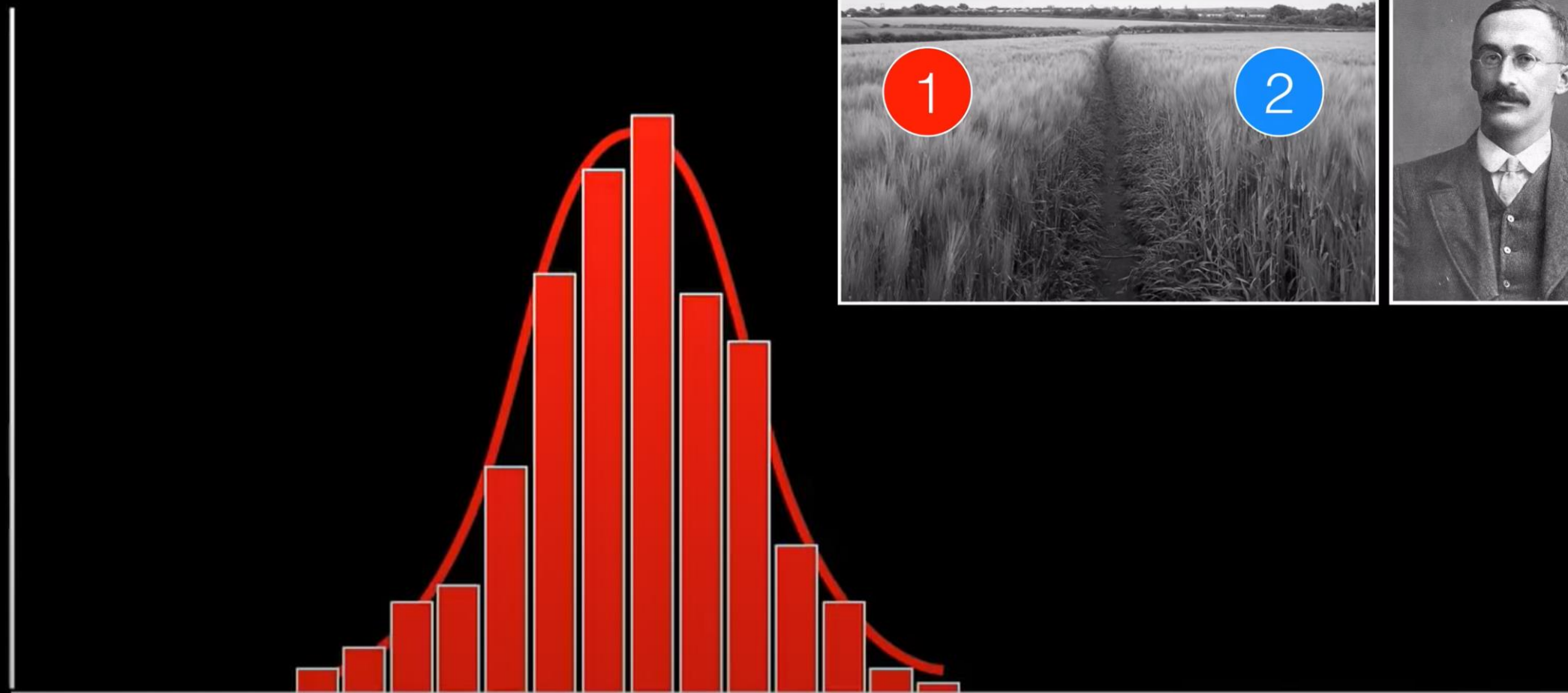
Introduction

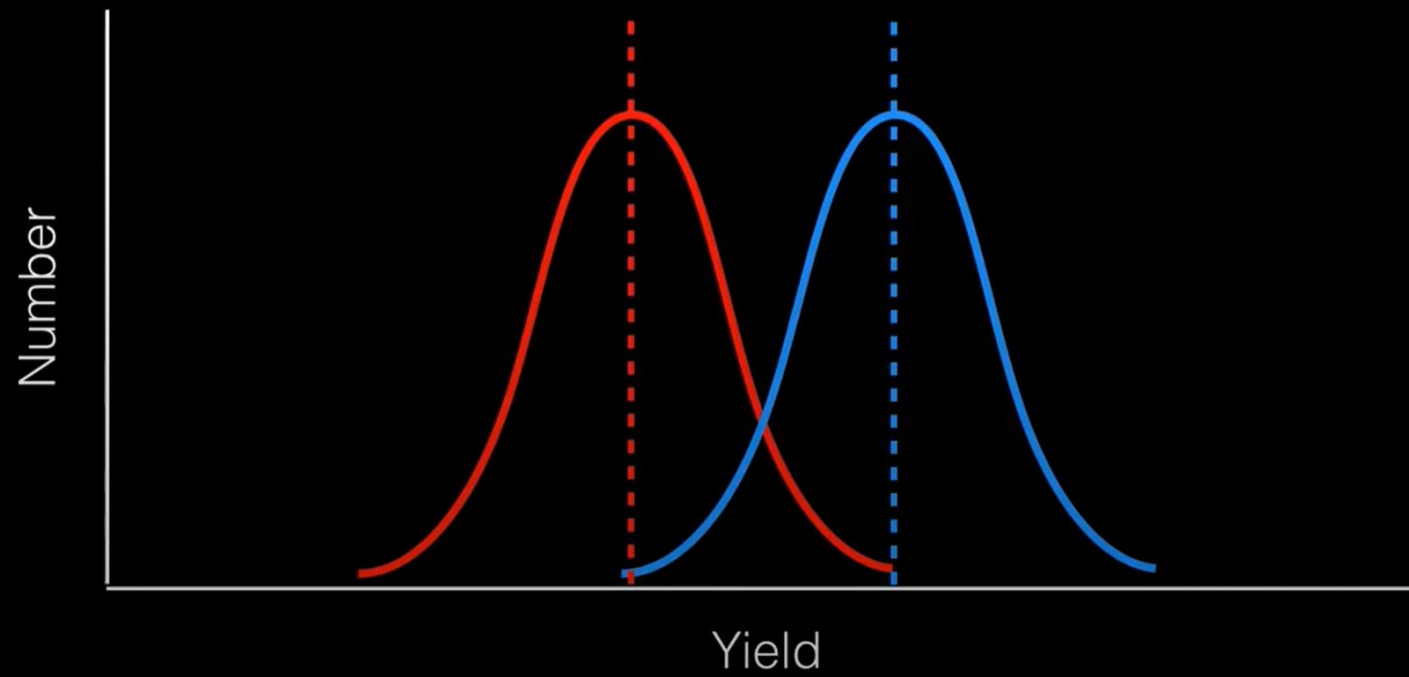
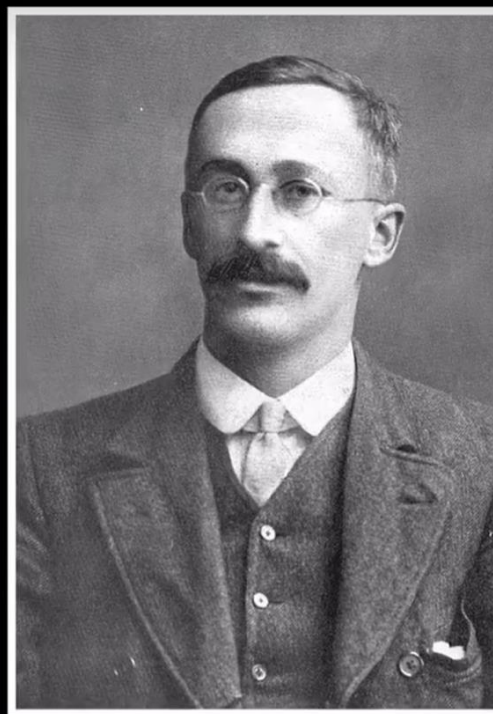
Any experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a greater number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

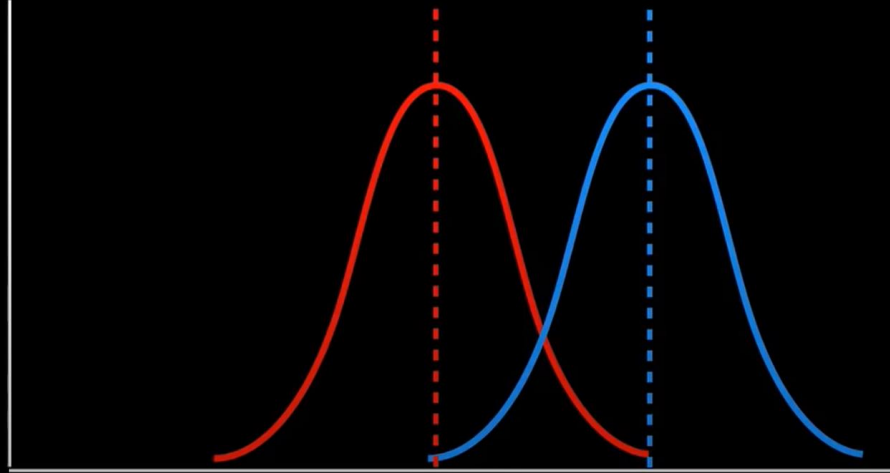
If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty: (1) owing to the "error of random sampling" the mean of our series of experiments deviates more or less widely from the mean of the population, and (2) the sample is not sufficiently large to determine what is the law of distribution of individuals. It is usual, however, to assume a normal distribution, because, in a very large number of cases, this gives an approximation so close that a small sample will give no real information as to the manner in which the population deviates from normality: since some law of distribution must be assumed it is better to work with a curve whose area and ordinates are tabled, and whose properties are well known. This assumption is accordingly made in the present paper, so that its conclusions are not strictly applicable to populations known not to be normally distributed; yet it appears probable that the deviation from normality must be very extreme to lead to serious error. We are concerned here solely with the first of these two sources of uncertainty.

Number



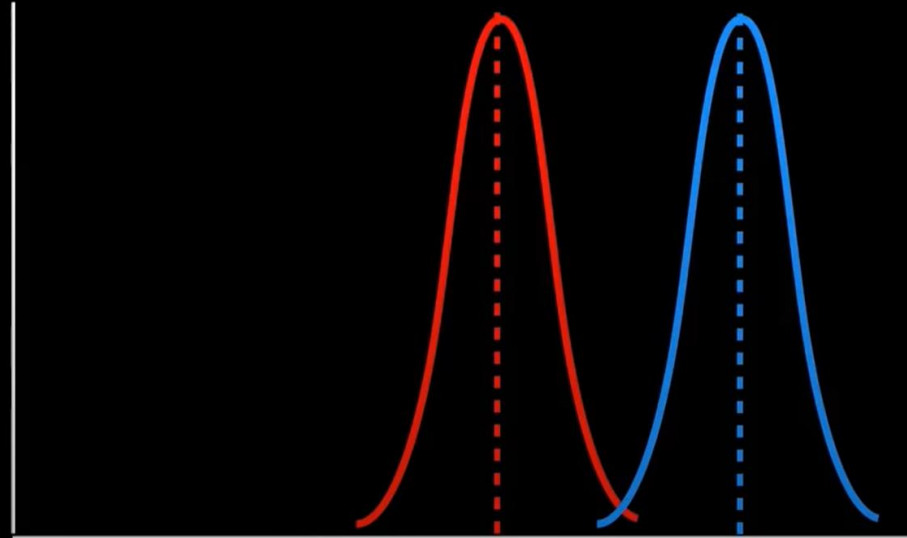


Number



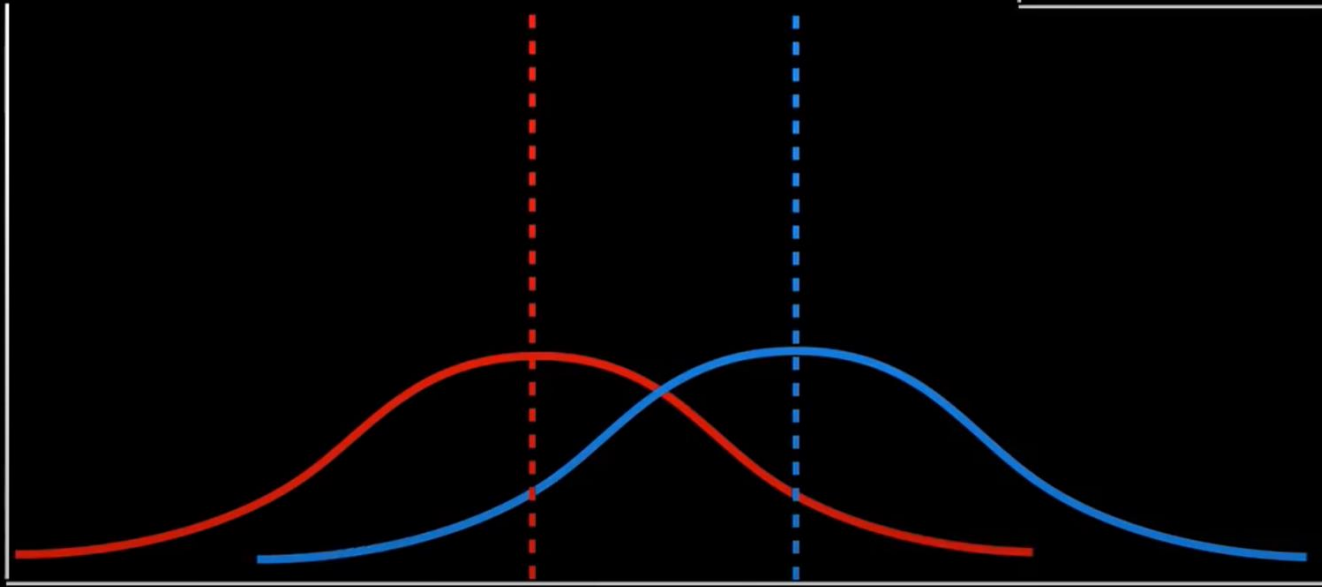
Yield

Number



Yield

Number



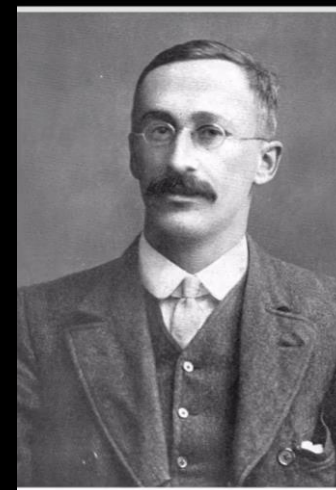
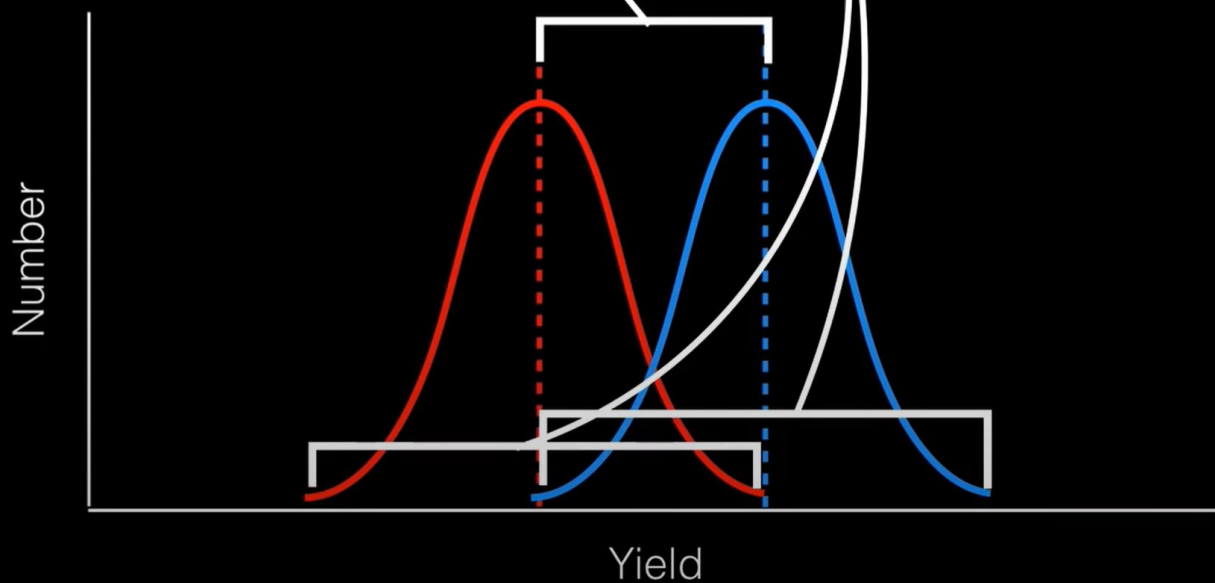
Yield

The t-value

$$\frac{\text{signal}}{\text{noise}}$$

$$\frac{\text{signal}}{\text{noise}} = \frac{\text{difference between group means}}{\text{variability of groups}} = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

t-value



	Field 1	Field 2
	15.2	15.9
	15.3	15.9
	16.0	15.2
	15.8	16.6
	15.6	15.2
	14.9	15.8
	15.0	15.8
	15.4	16.2
	15.6	15.6
	15.7	15.6
	15.5	15.8
	15.2	15.5
	15.5	15.5
	15.1	15.5
	15.3	14.9
	15.0	15.9
Mean	15.381	15.681
St Dev	0.312	0.407
Variance	0.098	0.166

$$\begin{aligned}
 \text{t-value} &= \frac{15.38 - 15.68}{\sqrt{\frac{.097}{16} + \frac{.165}{16}}} \\
 &= \frac{0.30}{.13} = 2.3
 \end{aligned}$$

NULL HYPOTHESIS EXAMPLES

THE NULL HYPOTHESIS ASSUMES THERE IS NO RELATIONSHIP BETWEEN TWO VARIABLES AND THAT CONTROLLING ONE VARIABLE HAS NO EFFECT ON THE OTHER.

CATS SHOW
NO PREFERENCE
FOR FOOD
BASED ON SHAPE.



PLANT GROWTH IS
NOT AFFECTED
BY LIGHT COLOR.



AGE HAS
NO EFFECT
ON
MUSICAL ABILITY.



The P Value and its Interpretation



Many people misunderstand what a P value means. Let's assume that you compared two means and obtained a P value equal to 0.03.

- *Correct definitions of this P value:*

There is a 3% chance of observing a difference as large as you observed even if the two population means are identical (the null hypothesis is true).

or

Random sampling from identical populations would lead to a difference smaller than you observed in 97% of experiments, and larger than you observed in 3% of experiments.

The P Value and its Interpretation

- **Wrong:**

~~There is a 97% chance that the difference you observed reflects a real difference between populations, and a 3% chance that the difference is due to chance.~~



H_0

There is no
statistically significant
difference between the
samples

t-test

Don't
Reject

Critical
Value

Reject

H₀

Degrees of Freedom	p=0.05	p=0.025	p=0.01
1	12.71	25.45	63.66
2	4.30	6.20	9.92
3	3.18	4.17	5.84
4	2.78	3.50	4.60
5	2.57	3.16	4.03
6	2.45	2.97	3.71
7	2.36	2.84	3.50
8	2.31	2.75	3.36
9	2.26	2.68	3.25
10	2.23	2.63	3.17
11	2.20	2.59	3.11
12	2.18	2.56	3.05
13	2.16	2.53	3.01
14	2.14	2.51	2.98
15	2.13	2.49	2.95
16	2.12	2.47	2.92
17	2.11	2.46	2.90
18	2.10	2.44	2.88
19	2.09	2.43	2.86
20	2.09	2.42	2.84
21	2.08	2.41	2.83
22	2.07	2.41	2.82
23	2.07	2.40	2.81
24	2.06	2.39	2.80
25	2.06	2.38	2.79
26	2.06	2.38	2.78
27	2.05	2.37	2.77
28	2.05	2.37	2.76
29	2.04	2.36	2.76
30	2.04	2.36	2.75
40	2.02	2.33	2.70
60	2.00	2.30	2.66
120	1.98	2.27	2.62
infinity	1.96	2.24	2.58

12	2.18	2.56	3.05
13	2.16	2.53	3.01
14	2.14	2.51	2.98
15	2.13	2.49	2.95
16	2.12	2.47	2.92
17	2.11	2.46	2.90
18	2.10	2.44	2.88
19	2.09	2.43	2.86
20	2.09	2.42	2.84
21	2.08	2.41	2.83
22	2.07	2.41	2.82
23	2.07	2.40	2.81
24	2.06	2.39	2.80
25	2.06	2.38	2.79
26	2.06	2.38	2.78
27	2.05	2.37	2.77
28	2.05	2.37	2.76
29	2.04	2.36	2.76
30	2.04	2.36	2.75
40	2.02	2.33	2.70
60	2.00	2.30	2.66
120	1.98	2.27	2.62

2.3

$$d_f = n_1 + n_2 - 2$$

H₀

Degrees of Freedom	p=0.05	p=0.025	p=0.01
1	12.71	25.45	63.66
2	4.30	6.20	9.92
3	3.18	4.17	5.84
4	2.78	3.50	4.60
5	2.57	3.16	4.03
6	2.45	2.97	3.71
7	2.36	2.84	3.50
8	2.31	2.75	3.36
9	2.26	2.68	3.25
10	2.23	2.63	3.17
11	2.20	2.59	3.11
12	2.18	2.56	3.05
13	2.16	2.53	3.01
14	2.14	2.51	2.98
15	2.13	2.49	2.95
16	2.12	2.47	2.92
17	2.11	2.46	2.90
18	2.10	2.44	2.88
19	2.09	2.43	2.86
20	2.09	2.42	2.84
21	2.08	2.41	2.83
22	2.07	2.41	2.82
23	2.07	2.40	2.81
24	2.06	2.39	2.80
25	2.06	2.38	2.79
26	2.06	2.38	2.78
27	2.05	2.37	2.77
28	2.05	2.37	2.76
29	2.04	2.36	2.76
30	2.04	2.36	2.75
40	2.02	2.33	2.70
60	2.00	2.30	2.66
120	1.98	2.27	2.62
infinity	1.96	2.24	2.58

12	2.18	2.56	3.05
13	2.16	2.53	3.01
14	2.14	2.51	2.98
15	2.13	2.49	2.95
16	2.12	2.47	2.92
17	2.11	2.46	2.90
18	2.10	2.44	2.88
19	2.09	2.43	2.86
20	2.09	2.42	2.84
21	2.08	2.41	2.83
22	2.07	2.41	2.82
23	2.07	2.40	2.81
24	2.06	2.39	2.80
25	2.06	2.38	2.79
26	2.06	2.38	2.78
27	2.05	2.37	2.77
28	2.05	2.37	2.76
29	2.04	2.36	2.76
30	2.04	2.36	2.75
40	2.02	2.33	2.70
60	2.00	2.30	2.66
120	1.98	2.27	2.62

2.3

$$d_f = n_1 + n_2 - 2$$

Assumptions!



- Normal Distribution
- Similar Variance (or correct it!)
- Datapoints are same/similar!

Untitled — Edited

File

Sheet

Undo

Clipboard

Analysis

Change

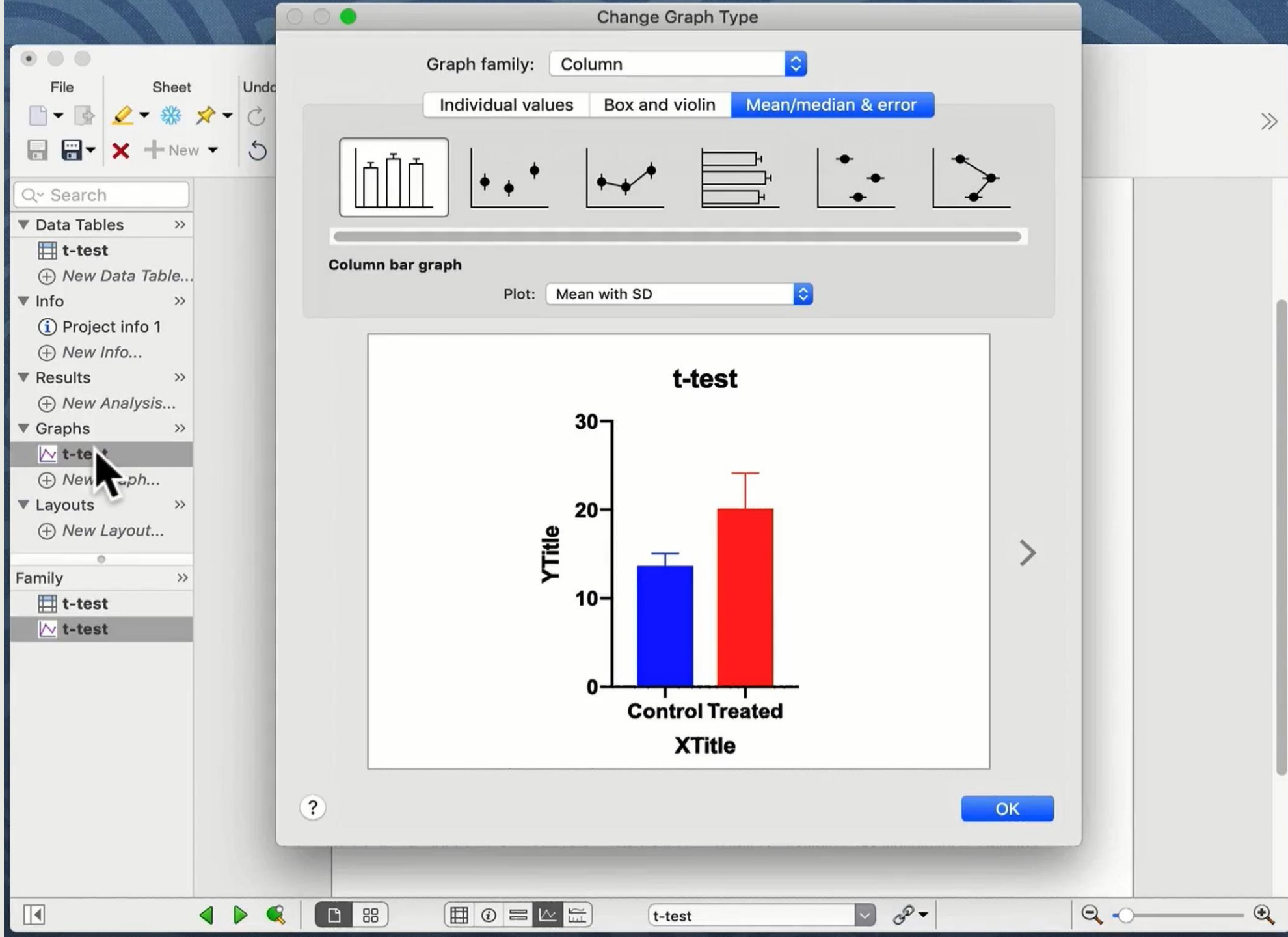
Import

Draw

Write

>>

Q Search		Group A	Group B	Group C	Group D	Group E	Group F	Group
▼ Data Tables >>		Control	Treated	Title	Title	Title	Title	Title
■ t-test		Y	Y	Y	Y	Y	Y	Y
⊕ New Data Table...	1	12	23					
▼ Info >>	2	14	21					
ⓘ Project info 1	3	13	16					
⊕ New Info...	4	15	23					
▼ Results >>	5	16	26					
⊕ New Analysis...	6	14	21					
▼ Graphs >>	7	13	15					
📈 t-test	8	12	16					
⊕ New Graph...	9							
▼ Layouts >>	10							
⊕ New Layout...	11							
Family >>	12							
■ t-test	13							
📈 t-test	14							
	15							
	16							
	17							
	18							
	19							

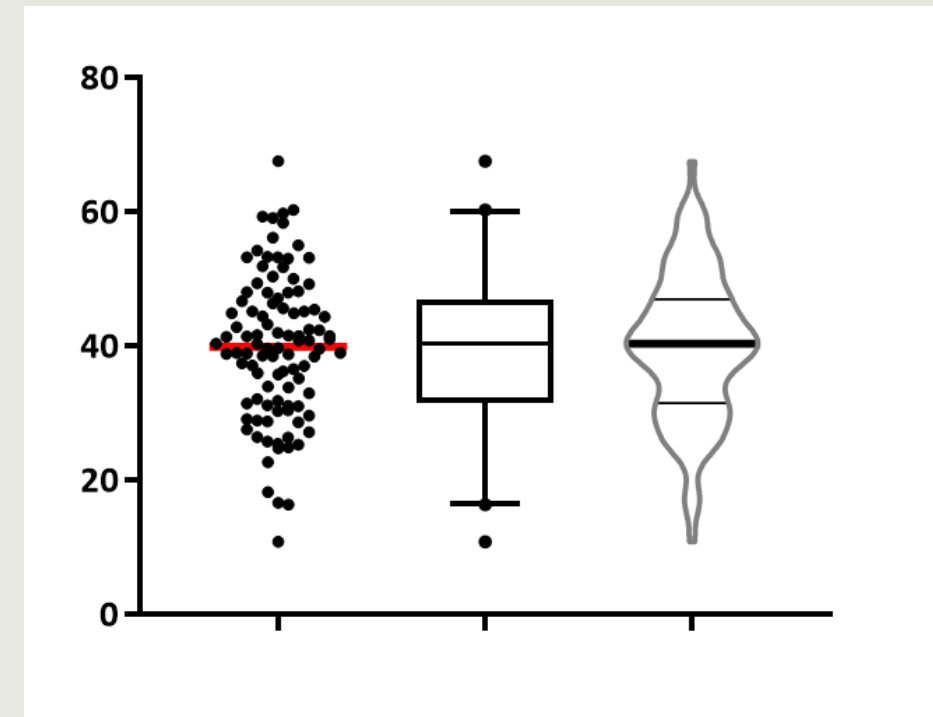


Which plot to choose?

Want to show the variation in your data?

If each value represents a different individual, you probably want to show the variation among values. Even if each value represents a different lab experiment, it often makes sense to show the variation.

- If you are plotting a column graph fewer than 100 or so values per data set, create a scatter plot that shows every value.
- If your data set has more than 100 or so values, a scatter plot becomes messy. Alternatives are to show a box-and-whiskers plot, a frequency distribution (histogram), or a cumulative frequency distribution.

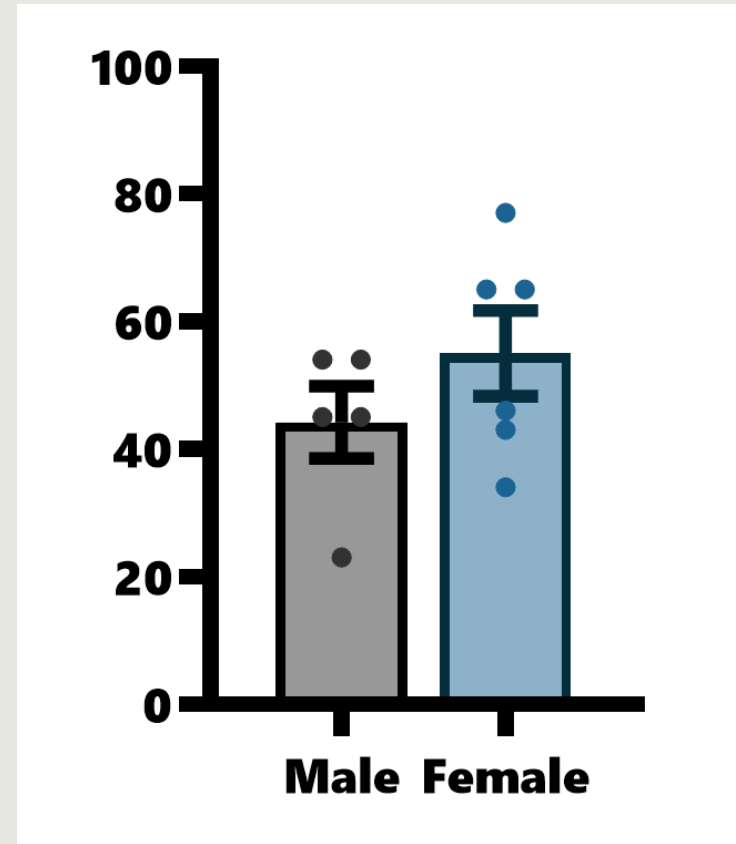


Which plot to choose?

Want to show the variation in your data?

- If you want to plot mean and error bar, the SD quantifies variability among replicates. So does a graph of median with interquartile range or full range.

When plotting a graph with error bars, be sure to explain how the error bars were computed in the figure itself or in its legend.



SD or SEM?

It is easy to be confused about the difference between the standard deviation (SD) and the standard error of the mean (SEM). Here are the key differences:

The SD quantifies scatter — how much the values vary from one another.

The SEM quantifies how precisely you know the true mean of the population. It takes into account both the value of the SD and the sample size.

Both SD and SEM are in the same units -- the units of the data.

The SEM, by definition, is always smaller than the SD.

SD or SEM?

The SEM gets smaller as your samples get larger. This makes sense, because the mean of a large sample is likely to be closer to the true population mean than is the mean of a small sample. With a huge sample, you'll know the value of the mean with a lot of precision even if the data are very scattered.

The SD does not change predictably as you acquire more data. The SD you compute from a sample is the best possible estimate of the SD of the overall population. As you collect more data, you'll assess the SD of the population with more precision. But you can't predict whether the SD from a larger sample will be bigger or smaller than the SD from a small sample. (This is not strictly true. It is the variance -- the SD squared -- that doesn't change predictably, but the change in SD is trivial and much much smaller than the change in the SEM.)

Analyze which data sets?

- Transform
- Transform concentrations (X)
- Normalize
- Prune rows
- Remove baseline and column math
- Transpose X and Y
- Fraction of Total

▼ Column analyses

t tests (and nonparametric tests)

One-way ANOVA (and nonparametric or m

One sample t and Wilcoxon test

Descriptive statistics

Normality and Lognormality Tests

Frequency distribution

ROC Curve

Bland-Altman method comparison

Identify outliers

Analyze a stack of P values

► Grouped analyses

☒ A:Control

☒ B:Treated

Select All

Deselect All

Cancel

OK

[illegible]

Parameters: t Tests (and Nonparametric Tests)

Experimental Design

Residuals

Options

Experimental design

☒ Unpaired☐ Paired

	Group A	Group B
	Control	Treated
	Y	Y
1		
2		
3		
4		
5		

Assume Gaussian distribution?

☒ Yes. Use parametric test.☐ No. Use nonparametric test.

Choose test

☒ Unpaired t test. Assume both populations have the same SD☐ Unpaired t test with Welch's correction. Do not assume equal SDs

Cancel

OK

Paired test examples!

You measure a variable in each subject before and after an intervention.

You recruit subjects as pairs, matched for variables such as age, ethnic group, and disease severity. One of the pair gets one treatment; the other gets an alternative treatment.

You run a laboratory experiment several times, each time with a control and treated preparation handled in parallel.

You measure a variable in twins or child/parent pairs.

Matching should be determined by the experimental design, and definitely should not be based on the variable you are comparing. If you are comparing blood pressures in two groups, it is OK to match based on age or postal code, but it is not OK to match based on blood pressure

Assume Gaussian Distribution?

Choosing when to use a nonparametric test is not straightforward. Here are some considerations:

- **Off-scale values.** With some kinds of experiments, one, or a few, values may be "off scale" -- too high or too low to measure. Even if the population is Gaussian, it is impossible to analyze these data with a t test or ANOVA. If you exclude these off scale values entirely, you will bias the results. If you estimate the value, the results of the t test depend heavily on your estimate. The solution is to use a nonparametric test.
- **Transforming can turn a nongaussian distribution into a Gaussian distribution.** If you are sure the data do not follow a Gaussian distribution, pause before choosing a nonparametric test. Instead, consider transforming the data, perhaps using logarithms or reciprocals. Often a simple transformation will convert non-Gaussian data to a Gaussian distribution. Then analyze the transformed values with a conventional test.

Assume Gaussian Distribution?

- **Noncontinuous data.** Clearly the population is far from Gaussian in these cases. The problem with using nonparametric tests is that so many values will tie for the same rank. Nonparametric tests have special corrections built-in to deal with tied ranks, but I am not sure how well those work when there are lots of tied ranks. An alternative would be to do a chi-square test.
- **Small samples.** If you have tiny samples (a few subjects in each group), the nonparametric tests have little or no power to find a significant difference.
- **Normality tests** should not be used to automatically decide whether or not to use a nonparametric test. But they can help you make the decision. (recommend the **D'Agostino-Pearson** normality test.)

**Mann-
Whitney test**

**Kolmogorov-
Smirnov test**

**Power to detect a shift
in the median?**

**More
power**

**Less
Power**

**Power to detect
differences in the shape
of the distributions?**

**Less
power**

**More
Power**

Parameters: t Tests (and Nonparametric Tests)

Experimental Design

Residuals

Options

Experimental design

☒ Unpaired☐ Paired

	Group A	Group B
	Control	Treated
	Y	Y
1		
2		
3		
4		
5		

Assume Gaussian distribution?

☒ Yes. Use parametric test.☐ No. Use nonparametric test.

Choose test

☒ Unpaired t test. Assume both populations have the same SD☐ Unpaired t test with Welch's correction. Do not assume equal SDs

?

Cancel

OK

Parameters: t Tests (and Nonparametric Tests)

Experimental Design Residuals **Options**

Calculations

P value: ☐ One-tailed ☒ Two-tailed (recommended)

Report differences as: Treated - Control

Confidence level: 95%

Definition of statistical significance: $P < 0.05$

Graphing options

- ☐ Graph differences (paired)
- ☐ Graph ranks (nonparametric)
- ☐ Graph correlation (paired)
- ☐ Graph CI of difference between means

Additional results

- ☐ Descriptive statistics for each dataset
- ☐ t Test: Also compare models using AICc
- ☐ Mann-Whitney: Also compute the CI of difference between medians
Assumes both distributions have the same shape.
- ☐ Wilcoxon: When both values on a row are identical, use method of Pratt
If this option is unchecked, those rows are ignored and the results will match prior version of Prism

Output

Show this many significant digits (for everything except P values): 4

P value style: GP: 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.000... N= 6

☐ Make options on this tab be the default for future tests.

Cancel OK

File Sheet

Search

Data Tables >>

- t-test
- + New Data Table...

Info >>

- Project info 1
- + New Info...

Results >>

- + New Analysis...

Graphs >>

- t-test
- + New Graph...

Layouts >>

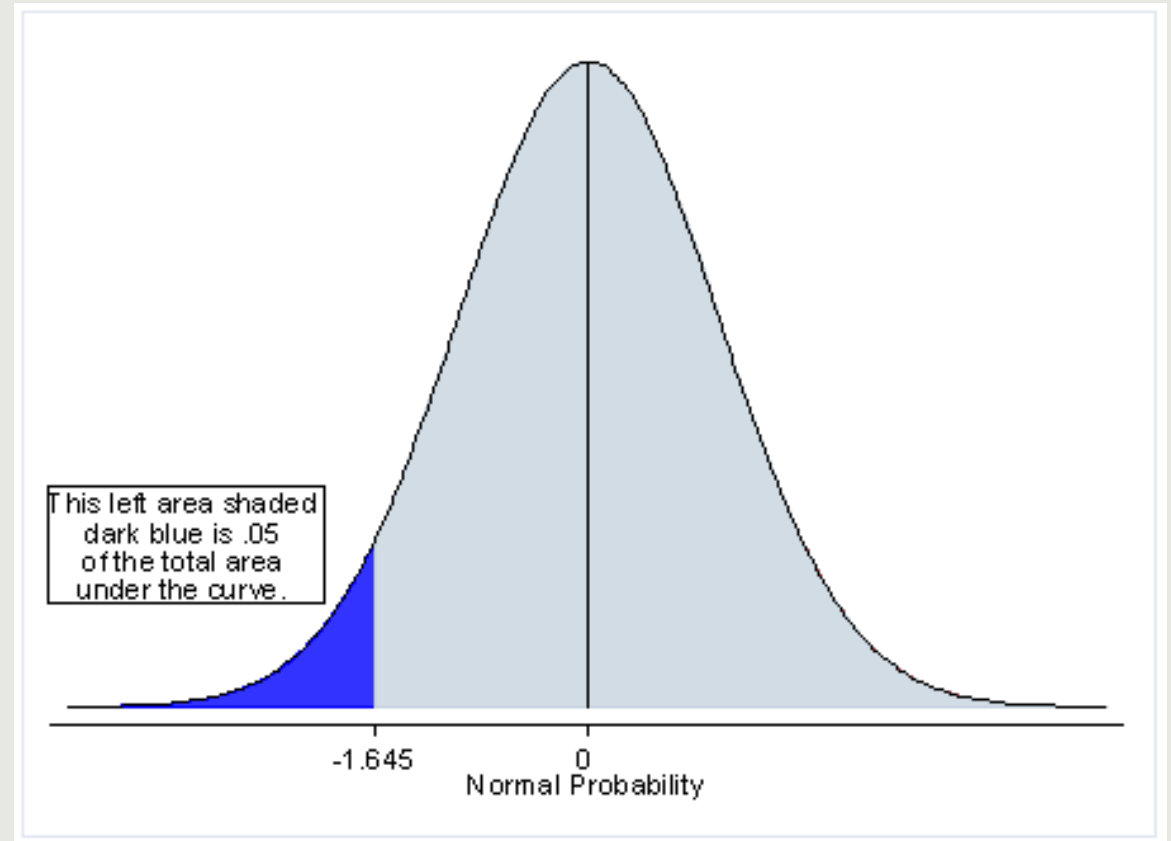
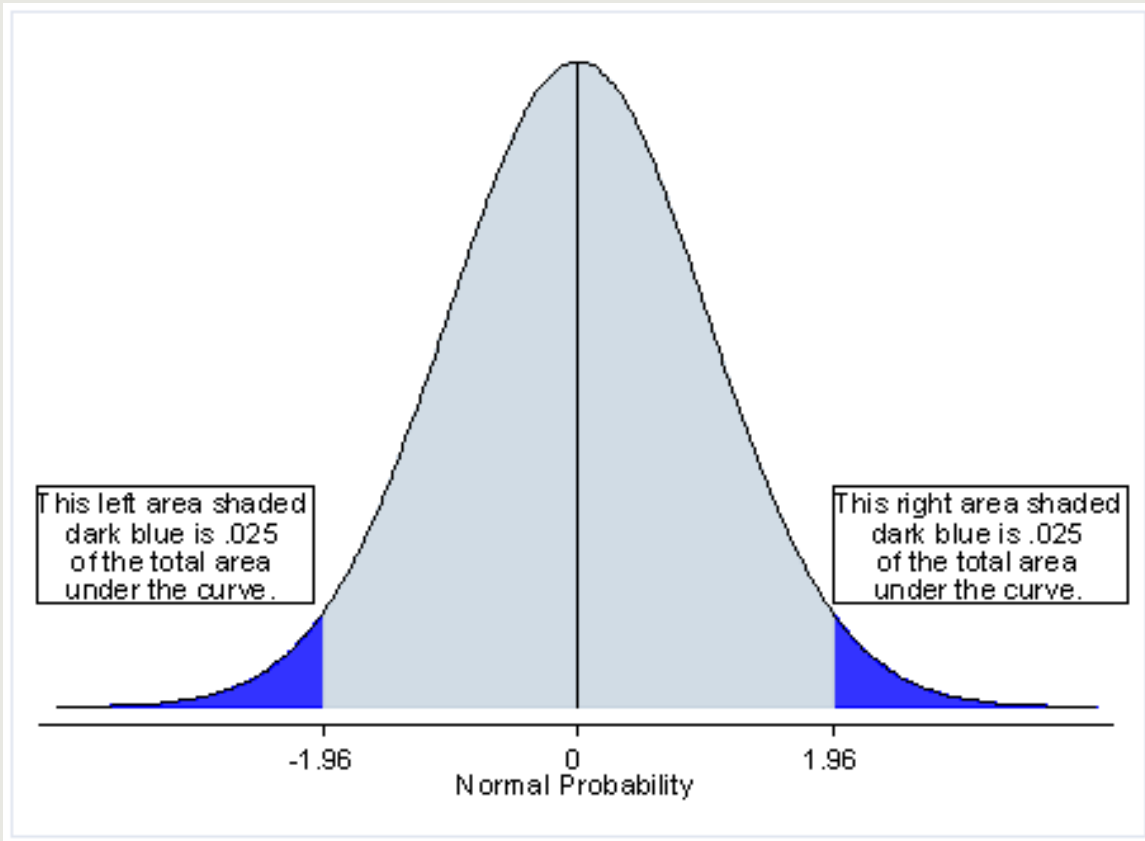
- + New Layout...

Family >>

- t-test
- t-test

Row 9, A: Cor

Two-tailed vs one-tailed t-tests



Student-t Checklist

Are the populations distributed according to a Gaussian distribution?

The unpaired t test assumes that you have sampled your data from populations that follow a Gaussian distribution. This assumption matters less with large samples due to the Central Limit Theorem.

Prism can perform normality tests as part of the Column Statistics analysis

If your data do not come from Gaussian distributions, you have three options. Your best option is to transform the values (perhaps to logs or reciprocals) to make the distributions more Gaussian. Another choice is to use the Mann-Whitney nonparametric test instead of a t test. A final option is to use t test anyway, knowing that it is fairly robust to violations of a Gaussian distribution with large samples.

Student-t Checklist

Do the two populations have the same variances?

- The unpaired t test assumes that the two populations have the same variances (and thus the same standard deviation)

Are the data unpaired?

- The unpaired t test works by comparing the difference between means with the standard error of the difference, computed by combining the standard errors of the two groups. If the data are paired or matched, then you should choose a paired t test instead. If the pairing is effective in controlling for experimental variability, the paired t test will be more powerful than the unpaired test.

Student-t Checklist

Are the “errors” independent?

- The term “error” refers to the difference between each value and the group mean. The results of a t test only make sense when the scatter is random – that whatever factor caused a value to be too high or too low affects only that one value. Prism cannot test this assumption. You must think about the experimental design.

For example, the errors are not independent if you have six values in each group, but these were obtained from two animals in each group (in triplicate). In this case, some factor may cause all triplicates from one animal to be high or low.

Student-t Checklist

Are you comparing exactly two groups?

- Use the t test only to compare two groups. To compare three or more groups, use one-way ANOVA followed by multiple comparison tests. It is not appropriate to perform several t tests, comparing two groups at a time.

Making multiple comparisons increases the chance of finding a statistically significant difference by chance and makes it difficult to interpret P values and statements of statistical significance. Even if you want to use planned comparisons to avoid correcting for multiple comparisons, you should still do it as part of one-way ANOVA to take advantage of the extra degrees of freedom that brings you.



GraphPad

Thank you!

Mann-Whitney Checklist

Are the “errors” independent?

- The term “error” refers to the difference between each value and the group median. The results of a Mann-Whitney test only make sense when the scatter is random – that whatever factor caused a value to be too high or too low affects only that one value. Prism cannot test this assumption. You must think about the experimental design. For example, the errors are not independent if you have six values in each group, but these were obtained from two animals in each group (in triplicate). In this case, some factor may cause all triplicates from one animal to be high or low.

Are the data unpaired?

- The Mann-Whitney test works by ranking all the values from low to high, and comparing the mean rank in the two groups. If the data are paired or matched, then you should choose a Wilcoxon matched pairs test instead.

Are you comparing exactly two groups?

- Use the Mann-Whitney test only to compare two groups. To compare three or more groups, use the Kruskal-Wallis test followed by post tests. It is not appropriate to perform several Mann-Whitney (or t) tests, comparing two groups at a time.

Mann-Whitney Checklist

Do the two groups follow data distributions with the same shape?

- If the two groups have distributions with similar shapes, then you can interpret the Mann-Whitney test as comparing medians. If the distributions have different shapes, you really cannot interpret the results of the Mann-Whitney test.

Do you really want to compare medians?

- The Mann-Whitney test compares the medians of two groups (well, not exactly). It is possible to have a tiny P value – clear evidence that the population medians are different – even if the two distributions overlap considerably.

Mann-Whitney Checklist

If you chose a one-tail P value, did you predict correctly?

- If you chose a one-tail P value, you should have predicted which group would have the larger median before collecting any data. Prism does not ask you to record this prediction, but assumes that it is correct. If your prediction was wrong, then ignore the P value reported by Prism and state that $P > 0.50$. One- vs. two-tail P values.

Are the data sampled from non-Gaussian populations?

- By selecting a nonparametric test, you have avoided assuming that the data were sampled from Gaussian distributions, but there are drawbacks to using a nonparametric test. If the populations really are Gaussian, the nonparametric tests have less power (are less likely to give you a small P value), and that difference is quite noticeable with small sample sizes.