BCBB Workshop

Statistical Testing

Qinlu (Claire) Wang, M.S.
Statistician

Bioinformatics and Computational Biosciences Branch (BCBB)
Office of Cyber Infrastructure and Computational Biology (OCICB)
National Institute of Allergy and Infectious Diseases (NIAID)

NIH National Institute of Allergy and Infectious Diseases

NIAID

OCICB Bioinformatics and Computational Biosciences Branch (BCBB)

- Part of NIAID
- Group of ~50
- Software developers
- Computational Biologists
- Project Management & Analysis Professionals
- Biostatistics, Phylogenetics, Genomics, Structural Biology, Programming

NIH National Institute of Allergy and Infectious Diseases

NIAID

How to contact us?

1. Submit Request

   Send emails to bioinformatics@niaid.nih.gov

2. Attend our seminars

   http://www.eventzilla.net/user/NIAID_OCICB_BCBB

3. Tell us what statistical topics you want to learn about

   BCBB Statistical Training – Suggest a class!

NIH National Institute of Allergy and Infectious Diseases

NIAID

Learning materials

- Yale Statistical Courses:

  http://www.stat.yale.edu/Courses/

- Statistics Glossary v1.1

  This site is a glossary of statistical terms searchable by topic or in alphabetical order. Created by Valerie J. Easton & John H. McColl

  http://www.stats.gla.ac.uk/steps/glossary/

NIH National Institute of Allergy and Infectious Diseases

NIAID

www.nature.com/nature/authors/gta

nature

Statistical checklist

Authors providing statistical analysis in their papers are asked to check the following points to ensure statistical adequacy.

It is helpful to the editors and the referees if authors confirm in their submission and resubmission cover letters that they have applied all relevant checks from the list below to the work described in their paper.

IN THE METHODS SECTION:

Type and applicability of test used

- Comparisons of interest are clearly defined
- Name of tests applied are clearly stated
- All statistical methods identified unambiguously
- Justification for use of test is given
- Data meet all assumptions of tests applied (with particular attention paid to non-normal data sets or small sample sizes, which should be identified in the text as such)
- Adjustments made for multiple testing is explained

Details about the test

- *n* is reported at the start of the study and for each analysis thereafter
- Sample size calculation (or justification) is given
- Unit of analysis is given for all comparisons
- Alpha level is given for all statistical tests
- Tests are clearly identified as one or two-tailed
- Randomization procedures or other ways to eliminate bias in sampling (in particular for experiments involving animals) described
- Actual *P* values are given for primary analyses

Descriptive statistics summary

- *n* for each data set is clearly stated
- A clearly labelled measure of centre (e.g. mean or median) is given
- A clearly labelled measure of variability (e.g. standard deviation or range[]) is given
- All numbers following a ± sign are identified as standard errors (s.e.m.) or standard deviations (s.d.)

Anomalies

- Any unusual or complex statistical methods are clearly defined and explained for Nature's wide readership. (Authors are encouraged to use Supplementary Information for long explanations.)
- Any data exclusions are stated and explained
- Any discrepancies in the value of *n* between analyses are clearly explained and justified
- Any method of treatment assignment (randomization, etc) is explained and justified
- Any data transformations are clearly described and justified

WITHIN INDIVIDUAL GRAPHS:

Distortions

- Any distorted effect sizes (e.g. by truncation of *y* axis) are clearly labelled and justified

Clear labelling

- Error bars are present on all graphs, where applicable.
- All error bars are clearly labelled

PLEASE HELP US

Many statistical analyses published in *Nature* are highly sophisticated and outside the scope of this checklist, particularly in the case of some studies in physical sciences disciplines. Authors and referees who have specific suggestions for additional entries to this list are encouraged to send them by e-mail to authors@nature.com or referees@nature.com. Nature will update this checklist at intervals in an effort to ensure that papers published are statistically robust.

NIH National Institute of Allergy and Infectious Diseases

NIAID

Outline

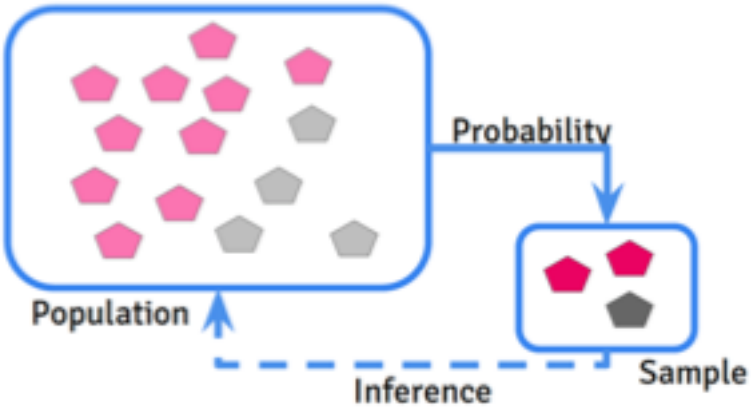
1. Statistical Testing Process
   - 1.1. Formulate Null and Alternative hypotheses
   - 1.2. Calculate the appropriate test statistics
   - 1.3. Compute the probability of observing the test statistic under the null hypothesis
   - 1.4. Make a statistical decision
   - 1.5. Make a biological conclusion
2. Common Statistical Tests
   - 2.1.Choose appropriate tests
   - 2.2.Two-sample T-test
   - 2.3.One-way ANOVA
   - 2.4.Two-way ANOVA
   - 2.5.Multiple Comparisons
3. Application in Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

1. Statistical Testing Process

The best way to determine whether a statistical hypothesis is true would be to examine the entire population. Since that is often impractical, researchers typically examine a random sample from the population. If sample data are not consistent with the statistical hypothesis, the hypothesis is rejected.



NIH National Institute of Allergy and Infectious Diseases

NIAID

1.1. Formulate Null and Alternative hypotheses
- *E.g. (null)* H_0: $\mu_1 = \mu_2$ *vs. (alternative)* H_A: $\mu_1 \neq \mu_2$

1.2. Calculate the appropriate test statistics
- *E.g. Student's t-test, Wilcoxon test, …*

1.3. Compute the probability of observing the test statistic (i.e. your sample data) under the null hypothesis
- *i.e. Compute a p-value*

1.4. Make a statistical decision
- *"Reject the null hypothesis" or "fail to reject the null hypothesis"*

1.5. Make a biological conclusion
- *E.g. New drug reduces viral load, vitamin C helps prevent cancer*

NIH National Institute of Allergy and Infectious Diseases

NIAID

Outline

1. Statistical Testing Process
   - 1.1. Formulate Null and Alternative hypotheses
   - 1.2. Calculate the appropriate test statistics
   - 1.3. Compute the probability of observing the test statistic under the null hypothesis
   - 1.4. Make a statistical decision
   - 1.5. Make a biological conclusion
2. Common Statistical Tests
   - 2.1.Choose appropriate tests
   - 2.2.Two-sample T-test
   - 2.3.One-way ANOVA
   - 2.4.Two-way ANOVA
   - 2.5.Multiple Comparisons
3. Application in Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

1.1. Formulate Null And Alternative Hypotheses

- Two Types of Statistical Hypotheses
- One-Sided vs. Two-sided Tests

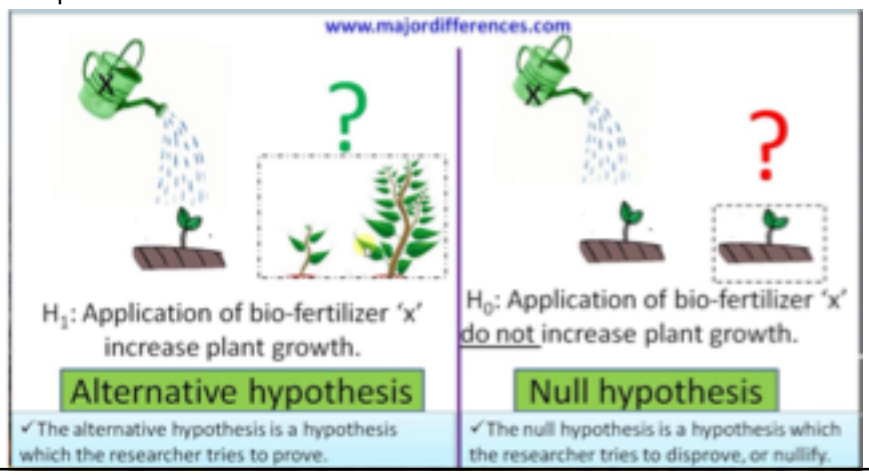
NIH National Institute of Allergy and Infectious Diseases

NIAID

➢ Two Types of Statistical Hypotheses

Null hypothesis: states that a population parameter (such as the mean, the standard deviation, and so on) is equal to a hypothesized value. The null hypothesis is often an initial claim that is based on previous analyses or specialized knowledge.

Alternative hypothesis: states that a population parameter is smaller, greater, or different than the hypothesized value in the null hypothesis. The alternative hypothesis is what you might believe to be true or hope to prove true.



NIH National Institute of Allergy and Infectious Diseases

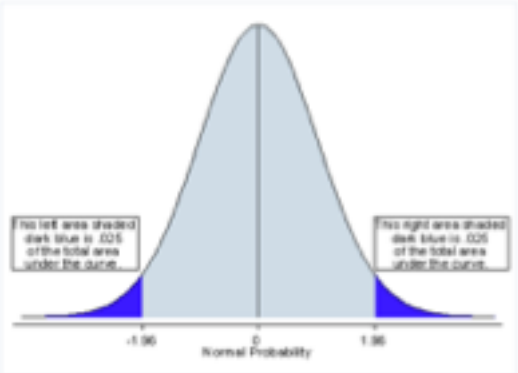
NIAID

➢ One-Sided vs. Two-sided Tests

Two-sided Test

If you are using a significance level of α, a two-sided test allots half of your α to test the statistical significance in one direction and half of your α to test statistical significance in the other direction.

For example, we may wish to compare the mean of a sample to a given value μ using a t-test. Our null hypothesis is that the mean is equal to μ. A two-sided test will test both if the mean is significantly greater than μ **AND** if the mean significantly less than μ. The mean is considered significantly different from μ if the test statistic is in the top 2.5% or bottom 2.5% of its probability distribution, resulting in a p-value less than 0.05.



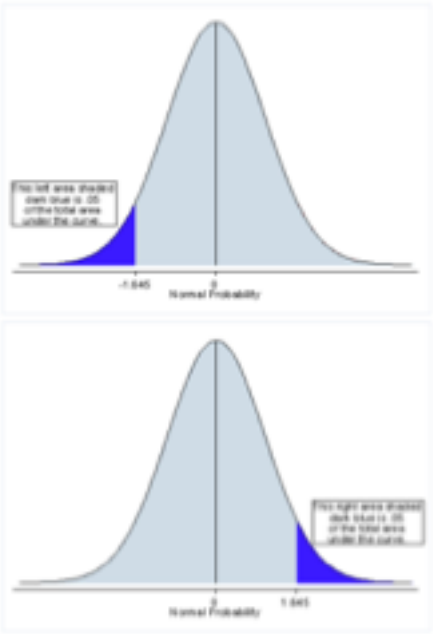
NIH National Institute of Allergy and Infectious Diseases

NIAID

One-sided Test

If you are using a significance level of α, an one-sided test allots all of your α to test the statistical significance in one direction of interest.

Let's return to our example comparing the mean of a sample to a given value μ using a t-test. Our null hypothesis is that the mean is equal to μ. An one-sided test will test either if the mean is significantly greater than μ **OR** less than μ, but not both. Then, depending on the chosen tail, the mean is significantly greater than or less than *x* if the test statistic is in the top 5% of its probability distribution or bottom 5% of its probability distribution, resulting in a p-value less than 0.05. The one-sided test provides more power to detect an effect in one direction by not testing the effect in the other direction.



NIH National Institute of Allergy and Infectious Diseases

NIAID

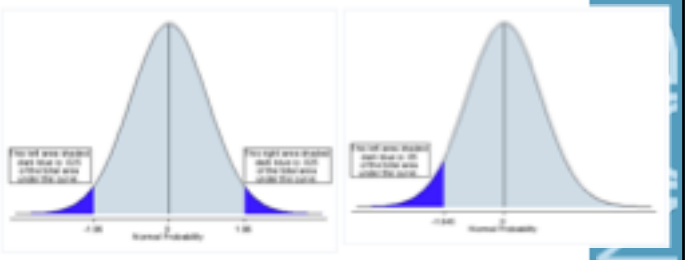
When is a one-sided test appropriate?

If you consider the consequences of missing an effect in the untested direction and conclude that they are negligible and in no way irresponsible or unethical, then you can proceed with an one-sided test.
For example, developing a new and cheaper drug.

When is an one-sided test NOT appropriate?

- When you have no idea of the direction or it is unacceptable to ignore the effect in the untested direction
- Choosing a one-tailed test for the sole purpose of attaining significance is not appropriate. Choosing a one-tailed test after running a two-tailed test that failed to reject the null hypothesis is not appropriate, no matter how "close" to significant the two-tailed test was.

	PROS	CONS
One-Tailed Tests	• Requires less traffic • Gains significance faster (read: why significance does not equal validity)	• Only accounts for one scenario • Can lead to inaccurate and biased results
Two-Tailed Tests	• Accounts for all three scenarios • Leads to accurate and reliable results	• Requires more traffic • Takes longer to gain significance



NIH National Institute of Allergy and Infectious Diseases

Outline

1. Statistical Testing Process
   - 1.1. Formulate Null and Alternative hypotheses
   - **1.2. Calculate the appropriate test statistics**
   - 1.3. Compute the probability of observing the test statistic under the null hypothesis
   - 1.4. Make a statistical decision
   - 1.5. Make a biological conclusion
2. Common Statistical Tests
   - 2.1.Choose appropriate tests
   - 2.2.Two-sample T-test
   - 2.3.One-way ANOVA
   - 2.4.Two-way ANOVA
   - 2.5.Multiple Comparisons
3. Application in Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

1.2. Statistical Tests

- What is Statistical Test
- Variance, Standard Deviation and Spread
- Common Statistical Tests

NIH National Institute of Allergy and Infectious Diseases

NIAID

➢ **What is statistical test**

$$\text{Test} = \frac{\text{Difference}}{\text{Error}} = \frac{\text{Statistic} - \text{Null Value}}{\text{Error}}$$

- Almost all tests used in inferential statistics can be generalized as the ratio of a "difference" over an "error"
  - Difference between a statistic and null value (usually 0)
  - A *statistic* is nothing more than a numeric summary of the experimental data with respect to the null hypothesis
  - A *null value* is an assumption about the population under the null hypothesis
  - An *error* is an estimate of the sampling distribution error

NIH National Institute of Allergy and Infectious Diseases

NIAID

Example: Two-sample Student's T-test

$$T^* = \frac{\overline{X}_1 - \overline{X}_2 - 0}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}}$$

statistic → $\overline{X}_1 - \overline{X}_2$; null value → 0 ; standard error → $\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$

- The "statistic" in a two-sample *t*-test is a difference between the two sample means and the null value is zero
  - The hypothesis $\mu_1 = \mu_2$ implies $\mu_1 - \mu_2 = 0$
- The standard error is an estimate of the common variance

NIH National Institute of Allergy and Infectious Diseases

NIAID

Variance, Standard Deviation and Spread

The standard deviation of the mean (SD) is the most commonly used measure of the spread of values in a distribution. SD is calculated as the square root of the variance (the average squared deviation from the mean).

Variance in a population is:

$$\sigma^2 = \frac{\sum_{\text{pop}}(x-\mu)^2}{n}$$

[x is a value from the population, μ is the mean of all x, n is the number of x in the population, Σ is the summation]

Variance is usually estimated from a sample drawn from a population. The unbiased estimate of population variance calculated from a sample is:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

[x_i is the *i*th observation from a sample of the population, x-bar is the sample mean, n (sample size) -1 is degrees of freedom, Σ is the summation]

The spread of a distribution is also referred to as dispersion and variability. All three terms mean the extent to which values in a distribution differ from one another.

SD is the best measure of spread of an approximately normal distribution. This is not the case when there are extreme values in a distribution or when the distribution is skewed, in these situations interquartile range or semi-interquartile are preferred measures of spread. Interquartile range is the difference between the 25th and 75th centiles. Semi-interquartile range is half of the difference between the 25th and 75th centiles. For any symmetrical (not skewed) distribution, half of its values will lie one semi-interquartile range either side of the median, i.e. in the interquartile range. When distributions are approximately normal, SD is a better measure of spread because it is less susceptible to sampling fluctuation than (semi-)interquartile range.

If a variable y is a linear (y = a + bx) transformation of x then the variance of y is b^2 times the variance of x and the standard deviation of y is b times the variance of x.

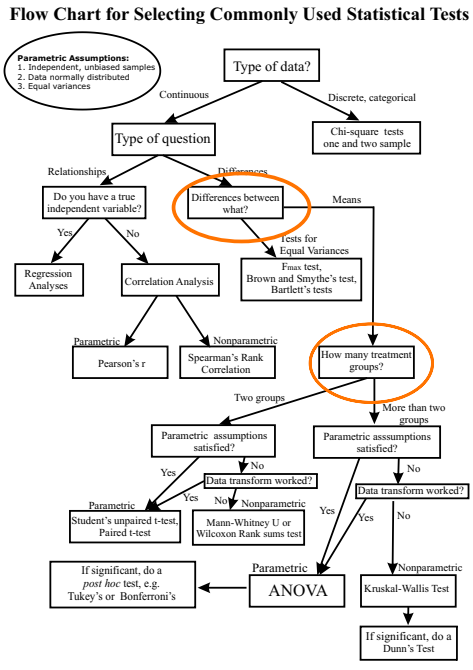
The standard error of the mean is the expected value of the standard deviation of means of several samples, this is estimated from a single sample as:

$$\text{SEM} = \frac{s}{\sqrt{n}}$$

[s is standard deviation of the sample mean, n is the sample size]

NIAID

How to choose appropriate statistical test?



NIH National Institute of Allergy and Infectious Diseases

NIAID

Parametric vs. Non-parametric Tests?

In practice:

- Parametric statistical procedures rely on assumptions about the shape of the distribution (i.e., assume a normal distribution) in the underlying population and about the form or parameters (i.e., means and standard deviations) of the assumed distribution.
- Nonparametric statistical procedures rely on **NO** or **FEW** assumptions about the shape or parameters of the population distribution from which the sample was drawn.

National Institute of Allergy and Infectious Diseases

NIAID

Why don't we always use nonparametric tests?

Two Main Drawbacks.

- Nonparametric tests generally are less statistically powerful than parametric procedures when the data truly are approximately normal. If you are planning a study and trying to determine how many patients to include, a nonparametric test will require a slightly larger sample size to have the same power as the corresponding parametric test.
- The results of nonparametric tests are often harder to interpret than the results of parametric tests. Many nonparametric tests use rankings of the values in the data rather than using the actual data. Knowing that the difference in mean ranks between two groups is five does not really help our intuitive understanding of the data.

--> In short, nonparametric procedures are useful in many cases and necessary in some, but they are not a perfect solution.

National Institute of Allergy and Infectious Diseases

NIAID

Outline

1. Statistical Testing Process
   - 1.1. Formulate Null and Alternative hypotheses
   - 1.2. Calculate the appropriate test statistics
   - **1.3. Compute the probability of observing the test statistic under the null hypothesis**
   - 1.4. Make a statistical decision
   - 1.5. Make a biological conclusion
2. Common Statistical Tests
   - 2.1.Choose appropriate tests
   - 2.2.Two-sample T-test
   - 2.3.One-way ANOVA
   - 2.4.Two-way ANOVA
   - 2.5.Multiple Comparisons
3. Application in Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

1.3. Probability of Observing the Test Statistic under the Null Hypothesis

- P-value
  - *What is p-value*
  - *How to calculate p–value*
  - *How to interpret p-value*
- Confidence Interval
  - *What is confidence interval*
  - *How to calculate confidence interval*
  - *Z-intervals vs. T-intervals*
  - *How to interpret confidence interval*

NIH National Institute of Allergy and Infectious Diseases

NIAID

P-value

P-value is measure of strength of the evidence against null hypothesis. It is the probability of getting the observed value of test statistic, or value with even greater evidence against null hypothesis (H_0), if the null hypothesis of a study question is true.

In short, a p-value is the probability of observing your data given that the null hypothesis is actually true

NIH National Institute of Allergy and Infectious Diseases

NIAID

How to calculate p-value?

1. Obtained according to the asymptotic distribution of the test statistic
   - *When the sample size is sufficiently large.*
   - *E.g: Student's t-test*

2. Calculate p-value empirically.
   - *E.g: Exact test, like Fisher's exact test*

3. Resampling-based procedures
   - *When the asymptotic distribution is either unreliable due to insufficient sample size or unavailable for complicated test statistics*
   - *E.g: permutation or bootstrap*

NIH National Institute of Allergy and Infectious Diseases

NIAID

Example:

Suppose a pharmaceutical company manufactures ibuprofen pills. They need to perform some quality assurance to ensure they have the correct dosage, which is supposed to be 500 milligrams. This is a two-sided test because if the company's pills are deviating significantly in either direction, meaning there are more than 500 milligrams or less than 500 milligrams, this will indicate a problem.

H_0: $\mu = 500\ mg$

H_A: $\mu \neq 500\ mg$

In a random sample of 125 pills, there is an average dose of 499.3 milligrams with a standard deviation of 6 milligrams. Because this is quantitative data, 500 mg is the population mean. We can use the following formula to calculate the z-score:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\bar{x}$ = *sample mean* = 499.3 *mg*

μ = *population mean* = 500 *mg*

σ = *sample standard deviation* = 6 *mg*

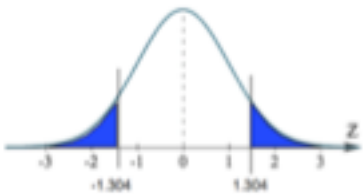
n = *sample size* = 123

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{499.3 - 500}{\frac{6}{\sqrt{125}}} = \frac{-0.7}{\frac{6}{\sqrt{125}}} = \frac{-0.7}{\frac{6}{11.18}} = \frac{-0.7}{0.53667} = -1.304$$

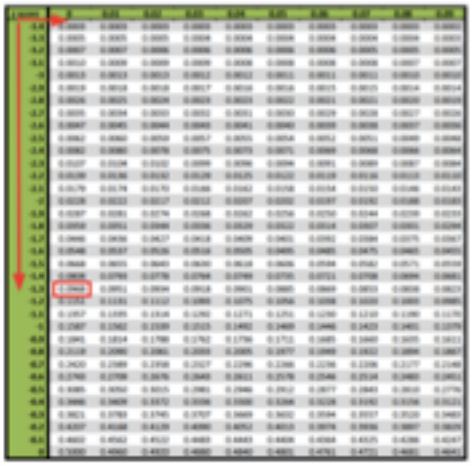
NIH National Institute of Allergy and Infectious Diseases

NIAID

We get a z-score of negative 1.304. Because this is a two-sided test, it is not enough to just look at the left tail. We also have to look at the equivalent of the right tail, or a positive 1.304.



The first way to find the p-value is to use the z-table. In the z-table, the left column will show values to the tenths place, while the top row will show values to the hundredths place. If we have a z-score of -1.304, we need to round this to the hundredths place, or -1.30. In the left column, we will first find the tenths place, or -1.3. In the top row, we will find the hundredths place, or 0.



NIH National Institute of Allergy and Infectious Diseases

This results in a p-value of 0.0968, or 9.68%, for a z-score of negative 1.304. We also need to take the positive 1.304 into account, which is the upper right tail. To calculate the true p-value, we just need to multiply 0.0968 by two, or 0.1936. This would be a p-value of 19.36%.

NIAID

How to interpret p-value?

In the majority of analyses, an alpha of 0.05 is used as the cutoff for significance.

- If the p-value is less than 0.05, we reject the null hypothesis that there's no difference between the means and conclude that a significant difference does exist.
- If the p-value is larger than 0.05, we *cannot* conclude that a significant difference exists.

NIH National Institute of Allergy and Infectious Diseases

NIAID

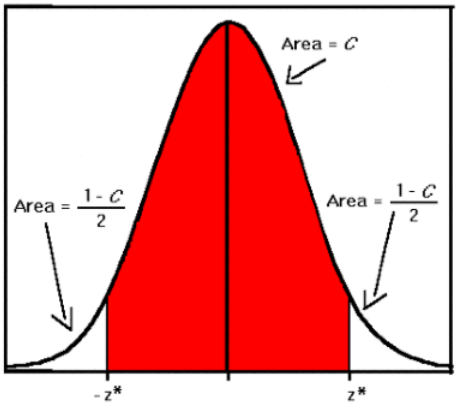
Confidence Interval

In statistical inference, one wishes to estimate population parameters using observed sample data.

"A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data."

---------- *Valerie J. Easton and John H. McColl's* Statistics Glossary v1.1

Strictly speaking a 95% confidence interval means that if we were to take 100 different samples and compute a 95% confidence interval for each sample, then approximately 95 of the 100 confidence intervals will contain the true mean value (**μ**).



NIH National Institute of Allergy and Infectious Diseases

NIAID

How to calculate confidence interval?

1. Obtained according to the asymptotic distribution of the test statistic
   - *When the sample size is sufficiently large.*
2. Resampling-based procedures
   - *When the asymptotic distribution is either unreliable due to insufficient sample size or unavailable for complicated test statistics*
   - *E.g: permutation or bootstrap*
   - *Permutation will be covered in the RNAseq training next week*

NIH National Institute of Allergy and Infectious Diseases

NIAID

The **confidence interval estimate (CI)** is a range of likely values for the population parameter based on:

- Point estimate, e.g., the sample mean
- Investigator's desired level of confidence (most commonly 95%)
- Sampling variability or the standard error of the point estimate.

The common notation for the parameter in question is θ. Often, this parameter is the population mean μ, which is estimated by the sample mean $\bar{x}$

The level C of a confidence interval gives the probability that the interval produced by the method employed includes the true value of the parameter θ.

For estimating the means, there are two types of confidence intervals that can be used: z-intervals and t-intervals.

NIH National Institute of Allergy and Infectious Diseases

NIAID

Z-Intervals

Use a z-interval when:
the sample size is greater than or equal to 30 and population standard deviation known OR Original population normal with the population standard deviation known.

Formula for the z-interval

If these conditions hold, we will use this formula for calculating the confidence interval:

$$\bar{x} \pm z_c \left(\frac{\sigma}{\sqrt{n}} \right)$$

where z_c is a critical value from the normal distribution (see below) and n is the sample size.

Common values of z_c are:

CONFIDENCE LEVEL	CRITICAL VALUE
90%	1.645
95%	1.96
99%	2.575

NIH National Institute of Allergy and Infectious Diseases

Reference: Understanding the *t*-distribution and its normal approximation

NIAID

Example using a z-interval

> Suppose that in a sample of 50 college students in Illinois, the mean credit card debt was $346. Suppose that we also have reason to believe (from previous studies) that the population standard deviation of credit card debts for this group is $108. Use this information to calculate a 95% confidence interval for the mean credit card debt of all college students in Illinois.

Solution

Since we wish to estimate the mean, we immediately know we will be using either a t-interval or a z-interval. Looking a bit closer, we see that we have a large sample size ($n = 50$) and we know the population standard deviation. Therefore, we will use a z-interval with $z_c = 1.96$. From reading the problem, we also have:

- Mean is $346: $\bar{x} = 346$
- Population standard deviation is 108: $\sigma = 108$

Applying the formula:

$$\bar{x} \pm z_c \left(\frac{\sigma}{\sqrt{n}} \right)$$
$$346 \pm 1.96 \left(\frac{108}{50} \right)$$

The $\pm$ indicates that we need to perform two different operations: a subtraction and an addition.

Left hand endpoint:

$$346 - 1.96 \left(\frac{108}{50} \right) = 316.1$$

Right hand endpoint:

$$346 + 1.96 \left(\frac{108}{50} \right) = 375.9$$

This gives our 95% confidence interval for μ, the population mean, as $(316.1, 375.9)$.

Interpretation
We are 95% confident that the mean amount of credit card debt for all college students in Illinois is between $316.10 and $375.90.
Of course this is a very particular statement, so please make sure you study how to interpret confidence intervals in general and so you can understand exactly what this means.

NIAID

T-Intervals: The much more realistic scenario is using a t-interval to estimate an unknown population mean. This interval relies on our sample standard deviation in calculating the margin of error. All this means for us is that the formula will be very similar, but the critical value will no longer come from the normal distribution. Instead, it will come from the student's t distribution.

Use a t-interval when:
Population standard deviation UNKNOWN and original population normal OR sample size greater than or equal to 30 and Population standard deviation UNKNOWN.

Formula for the t-interval

The formula for a t-interval is:

$$\bar{x} \pm t_c \left(\frac{s}{\sqrt{n}} \right)$$

where t_c is a critical value from the t-distribution, s is the sample standard deviation and n is the sample size.

Finding t_c

The value of t_c depends on the sample size through the use of "degrees of freedom" where $df = n - 1$. We will use this to look up the value of t_c in a table (a nice free version of that table can be found here, or typically in the back of your textbook if you are currently taking a class).

NIH National Institute of Allergy and Infectious Diseases

NIAID

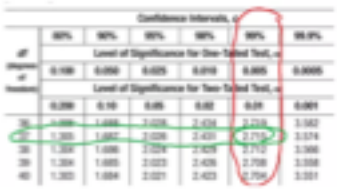
Example using a t-interval

> Suppose that a sample of 38 employees at a large company were surveyed and asked how many hours a week they thought the company wasted on unnecessary meetings. The mean number of hours these employees stated was 12.4 with a standard deviation of 5.1. Calculate a 99% confidence interval to estimate the mean amount of time all employees at this company believe is wasted on unnecessary meetings each week.

Solution

As before, since we are estimating a mean with a confidence interval, we know it will either be a t-interval or a z-interval. In this case, we have a large sample ($n = 38$), but we only have the sample standard deviation. If you aren't sure of that – read closely. The standard deviation of 5.1 was in the context of the sample, so $s = 5.1$. Thus, we will go ahead and use a t-interval since σ is unknown.

Before we can do that however, we need to look up the critical value. To know which row in the t-table to look at, we find the degrees of freedom which is $n - 1 = 38 - 1 = 37$. Using the table linked here:



Now that we have that, we plug the values into the formula and do the calculations to get our two endpoints. Remember that we have:

- Sample mean: $\bar{x} = 12.4$
- Sample size: $n = 38$
- Sample standard deviation: $s = 5.1$
- Critical value: $t_c = 2.715$

NIH National Institute of Allergy and Infectious Diseases

NIAID

Therefore the interval is:

$$\bar{x} \pm t_c \left(\frac{s}{\sqrt{n}} \right)$$

$$12.4 \pm 2.715 \left(\frac{5.1}{\sqrt{38}} \right)$$

This gives us the following two endpoints for our interval.

Left hand endpoint:

$$12.4 - 2.715 \left(\frac{5.1}{\sqrt{38}} \right) = 10.2$$

Right hand endpoint:

$$12.4 + 2.715 \left(\frac{5.1}{\sqrt{38}} \right) = 14.6$$

99% Confidence Interval for μ: $(10.2, 14.6)$

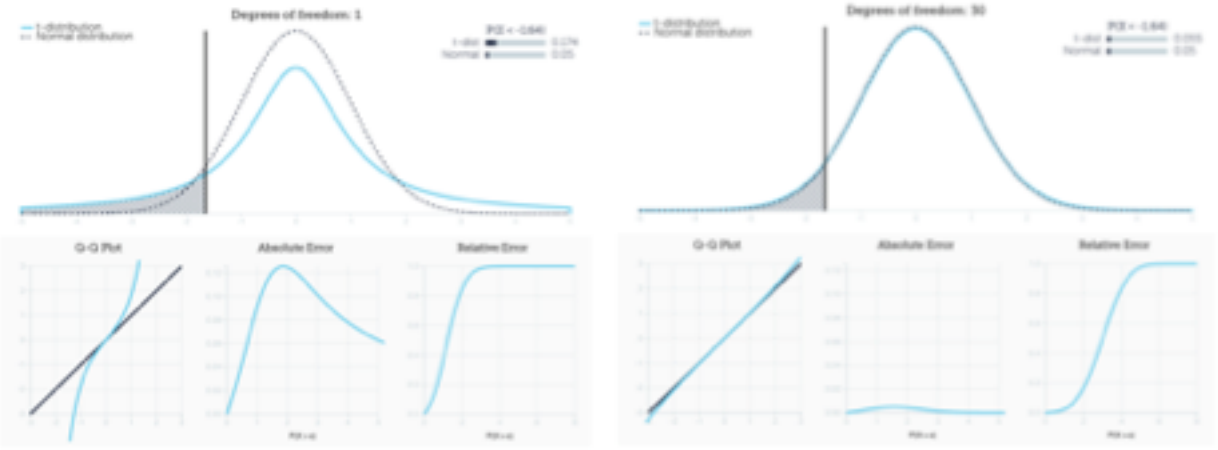
Interpretation

"We are 99% confident that the mean amount of time that all employees at this company think is wasted on meetings each week is between 10.2 and 14.6 hours."

NIH National Institute of Allergy and Infectious Diseases

NIAID

Understanding the t-distribution and its normal approximation

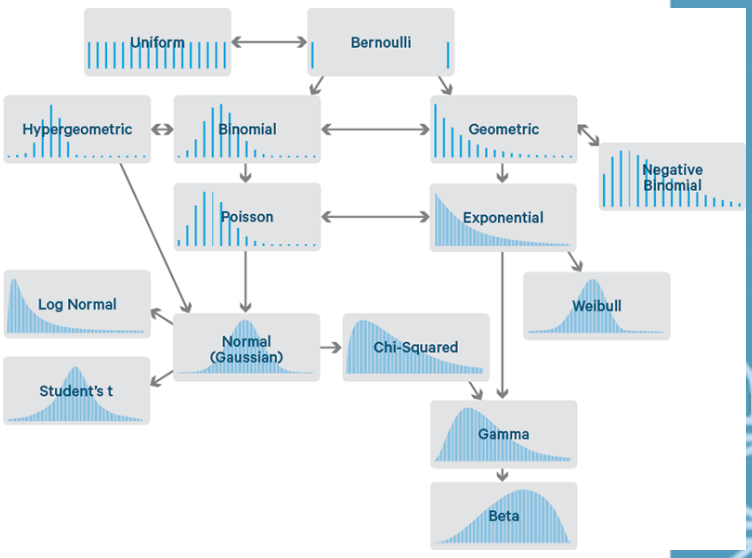


NIH National Institute of Allergy and Infectious Diseases

Reference: Interactive reference with different degrees of freedom

NI

Ins @limhengswee



National Institute o
Allergy and
Infectious Diseases

How to interpret confidence interval?

Once you have taken a sample and constructed the confidence interval (CI), there is no longer any "randomness" left in a CI statement (except for the Bayesian point of view which thinks of μ as being a random variable).

That is, when reporting a CI:

- "I am 95% confident that the mean is between 25.1 and 32.6" is correct.
- "There is a 95% probability that the mean is between 25.1 and 32.6" is WRONG. Either μ is in that interval or not; there is no probability associated with it.

National Institute of
Allergy and
Infectious Diseases

95% Confidence Interval simply means that if I take out samples a large number of times (tending to infinity) from a population with 0.05 significance level then 95% of those intervals will encompass the true parameter value.

There are two cases throwing a ball into a mug or throwing mug over a fixed lying ball. Most people think Confidence Interval as the first case but the things are other way round.

Randomness is not associated with parameter but the randomness is associated with interval. True parameters are fixed. But intervals estimated are random.

Roughly you can understand it like this. It is like, a ball (true parameter) is lying and you are throwing a mug(CI with 0.05 significance level) over it a large number of times. And asymptotically 95% of those throws will cover the ball.

So if I say in terms of frequentist then in future if you take out a sample with 95% CI then there is 0.95 probability that it will encompass true parameter value.

Ring Toss Game

NIH National Institute of Allergy and Infectious Diseases

Resource:
https://www.quora.com/How-do-you-interpret-a-confidence-interval

NIAID

Outline

1. Statistical Testing Process
   - 1.1. Formulate Null and Alternative hypotheses
   - 1.2. Calculate the appropriate test statistics
   - 1.3. Compute the probability of observing the test statistic under the null hypothesis
   - **1.4. Make a statistical decision**
   - **1.5. Make a biological conclusion**
2. Common Statistical Tests
   - 2.1.Choose appropriate tests
   - 2.2.Two-sample T-test
   - 2.3.One-way ANOVA
   - 2.4.Two-way ANOVA
   - 2.5.Multiple Comparisons
3. Application in Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

1.4&1.5. Statistical Decision and Biological Conclusions

- Statistical Decision
  - Type I and Type II errors
- Biological Conclusions

NIH National Institute of Allergy and Infectious Diseases

NIAID

- A **statistical decision** is a choice to "reject the null hypothesis" or "fail to reject the null hypothesis"
  - The decision is based on a critical value or decision rule
  - E.g. Reject the null hypothesis if p-value < 0.05
- A **biological conclusion** is the final interpretation of the statistical testing process in plain language
  - E.g. Vitamin C prevents cancer, drug reduced viral loads, …
  - Make sure conclusion can be justified by the hypotheses

NIH National Institute of Allergy and Infectious Diseases

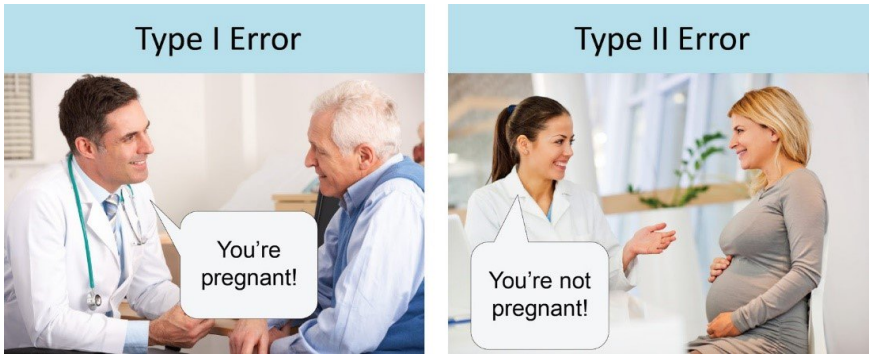
NIAID

Statistics is a game of probability, and we can never know for certain whether our statistical conclusions are correct. Whenever there is uncertainty, there is the possibility of making an error. In statistics, there are two types of statistical conclusion errors possible when you are testing hypotheses: Type I and Type II.

Type I and Type II Errors

- **Type I error** ("**false positive**"): The error of rejecting a TRUE null hypothesis.
- **Type II error** ("**false negative**"): The error of not rejecting a FALSE null hypothesis

For example, when the null hypothesis is "You're not pregnant", and the alternative hypothesis is "You are pregnant":



National Institute of Allergy and Infectious Diseases

NIAID

The chances of committing these two types of errors are inversely proportional—that is, decreasing Type I error rate increases Type II error rate, and vice versa. Your risk of committing a Type I error is represented by your alpha level (the *p* value below which you reject the null hypothesis). The commonly accepted α = .05 means that you will incorrectly reject the null hypothesis approximately 5% of the time.

- *To decrease the chance of committing Type I error: make your alpha (p) value more stringent.*
- *To decrease the chance of committing Type II error: increase your analyses' power by either increasing your sample size or relaxing your alpha level*

Depending on your field and your specific study, one type of error may be costlier than the other. Suppose you conducted a study looking at whether a plant derivative could prevent deaths from certain cancers. If you falsely concluded that it could not prevent cancer-related deaths when it really could (Type II error), you could potentially cost people their lives! If you were looking at whether people's happiness levels were higher when they held versus looked at a puppy, either type of error might not be so important.

	H_0 rejected	Fail to reject H_0
H_0 false	Correct	Type II error
H_0 true	Type I error	correct

Alpha (α) = Prob (Type I error)

Beta (β) = Prob (Type II error)

Power = $1 - \beta$

National Institute of Allergy and Infectious Diseases

NIAID

Outline

1. Statistical Testing Process
   - 1.1. Formulate Null and Alternative hypotheses
   - 1.2. Calculate the appropriate test statistics
   - 1.3. Compute the probability of observing the test statistic under the null hypothesis
   - 1.4. Make a statistical decision
   - 1.5. Make a biological conclusion

2. Common Statistical Tests

- 2.1.Choose appropriate tests
- 2.2.Two-sample T-test
- 2.3.One-way ANOVA
- 2.4.Two-way ANOVA
- 2.5.Multiple Comparisons

3. Application in Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

Outline

1. Statistical Testing Process
   - 1.1. Formulate Null and Alternative hypotheses
   - 1.2. Calculate the appropriate test statistics
   - 1.3. Compute the probability of observing the test statistic under the null hypothesis
   - 1.4. Make a statistical decision
   - 1.5. Make a biological conclusion

2. Common Statistical Tests

- 2.1.Choose appropriate tests
- 2.2.Two-sample T-test
- 2.3.One-way ANOVA
- 2.4.Two-way ANOVA
- 2.5.Multiple Comparisons

3. Application in Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

2.1.Choose Appropriate Tests

We might need to answer some questions:

1. *How many independent variables/factors?*
2. *How many levels do the independent variables have? i.e, how many groups do we have?*
3. *What is the nature of the independent variables/factors? Continuous? Categorical?*
4. *What is the nature of the dependent variable? ordinal or norminal?*
5. *Is the dependent variable normally distributed?*

……..

NIH National Institute of Allergy and Infectious Diseases

NIAID

Useful References for choosing appropriate tests

1. Choosing the Correct Statistical Test in SAS, Stata, SPSS and R (link)
   - The flow chart to choose the best statistical test and the related codes with SAS, Stata, SPSS and R

2. Summary and Analysis of Extension Program Evaluation in R (link)
   - A dictionary of statistical testing methods and the related R codes

3. Analysis Data Model (ADaM) Examples in Commonly Used Statistical Analysis Methods (link)
   - Commonly used statistical analysis methods in clinical trials (proved by FDA)

NIH National Institute of Allergy and Infectious Diseases

NIAID

Common Statistical Tests

Type of Test:	Use:
Correlational	These tests look for an association between variables
Pearson correlation	Tests for the strength of the association between two continuous variables
Spearman correlation	Tests for the strength of the association between two ordinal variables (does not rely on the assumption of normal distributed data)
Chi-square	Tests for the strength of the association between two categorical variables
Comparison of Means:	***look for the difference between the means of variables***
Paired T-test	Tests for difference between two related variables
Independent T-test	Tests for difference between two independent variables
ANOVA	Tests the difference between group means after any other variance in the outcome variable is accounted for
Regression:	***assess if change in one variable predicts change in another variable***
Simple regression	Tests how change in the predictor variable predicts the level of change in the outcome variable
Multiple regression	Tests how change in the combination of two or more predictor variables predict the level of change in the outcome variable
Non-parametric:	***are used when the data does not meet assumptions required for parametric tests***
Wilcoxon rank-sum test	Tests for difference between two independent variables - takes into account magnitude and direction of difference
Wilcoxon sign-rank test	Tests for difference between two related variables - takes into account magnitude and direction of difference
Sign test	Tests if two related variables are different – ignores magnitude of change, only takes into account direction

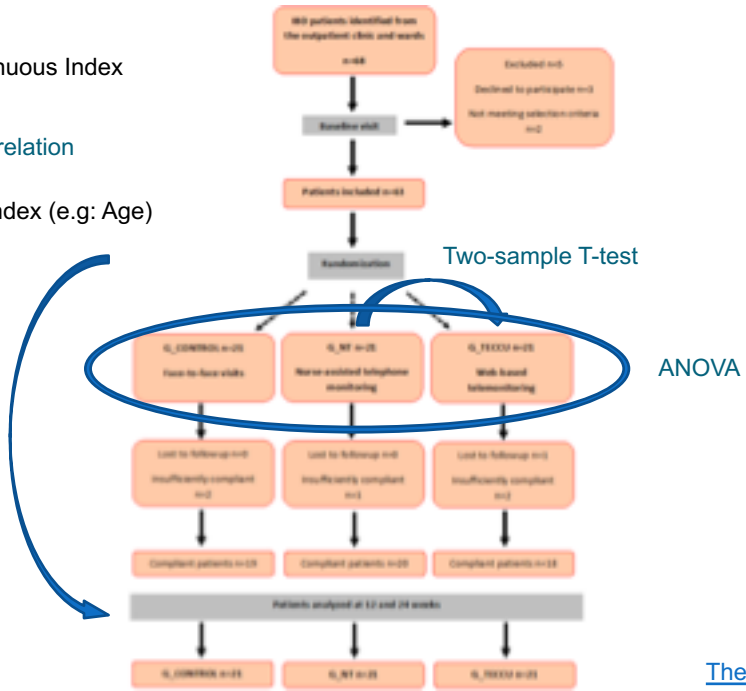
NIH

NIAID

Main Interest: A continuous Index

Correlation

Another continuous Index (e.g: Age)



NIH National Institute of Allergy and Infectious Diseases

The publication

NIAID

Analysis Type	Example	Parametric Procedure	Nonparametric Procedure
Compare means between two distinct/independent groups	Is the mean systolic blood pressure (at baseline) for patients assigned to placebo different from the mean for patients assigned to the treatment group?	Two-sample t-test	Wilcoxon rank-sum test
Compare two quantitative measurements taken from the same individual	Was there a significant change in systolic blood pressure between baseline and the six-month follow-up measurement in the treatment group?	Paired t-test	Wilcoxon signed-rank test
Compare means between three or more distinct/independent groups	If our experiment had three groups (e.g., placebo, new drug #1, new drug #2), we might want to know whether the mean systolic blood pressure at baseline differed among the three groups?	Analysis of variance (ANOVA)	Kruskal-Wallis test
Estimate the degree of association between two quantitative variables	Is systolic blood pressure associated with the patient's age?	Pearson coefficient of correlation	Spearman's rank correlation

NIH National Institute of Allergy and Infectious Diseases

NIAID

Outline

1. Statistical Testing Process
   - 1.1. Formulate Null and Alternative hypotheses
   - 1.2. Calculate the appropriate test statistics
   - 1.3. Compute the probability of observing the test statistic under the null hypothesis
   - 1.4. Make a statistical decision
   - 1.5. Make a biological conclusion
2. Common Statistical Tests
   - 2.1.Choose appropriate tests
   - **2.2.Two-sample T-test**
   - 2.3.One-way ANOVA
   - 2.4.Two-way ANOVA
   - 2.5.Multiple Comparisons
3. Application in Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

2.2.Two-sample T-test

Two-sample t-test measures whether the means of the measurement variable are different in the two groups. There are two kinds of two-sample t-test: paired and unpaired.

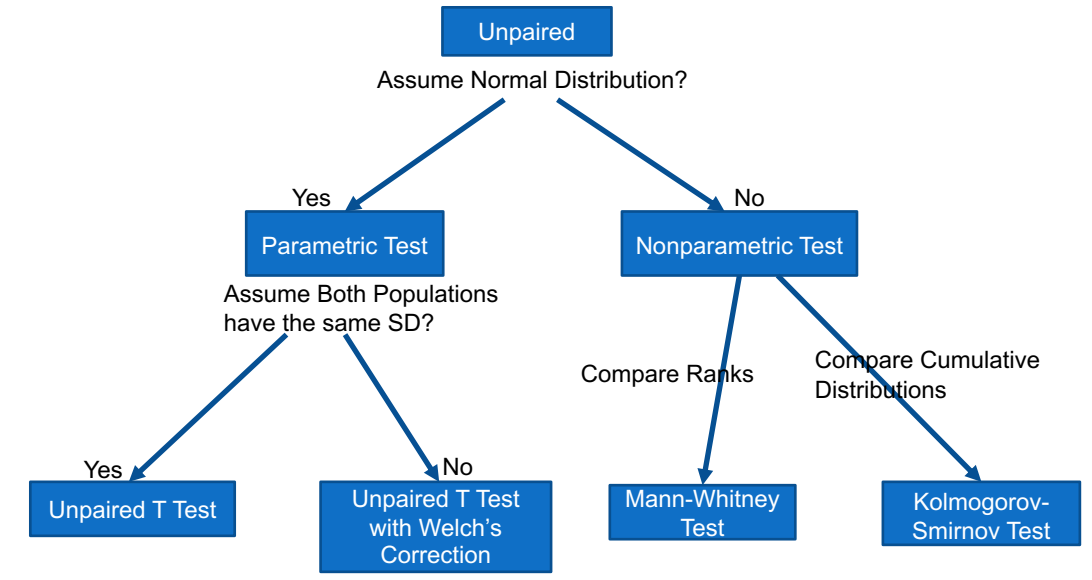
Paired

Choose a paired test when the columns of data are matched. That means that values on the same row are related to each other. Here are some 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.*

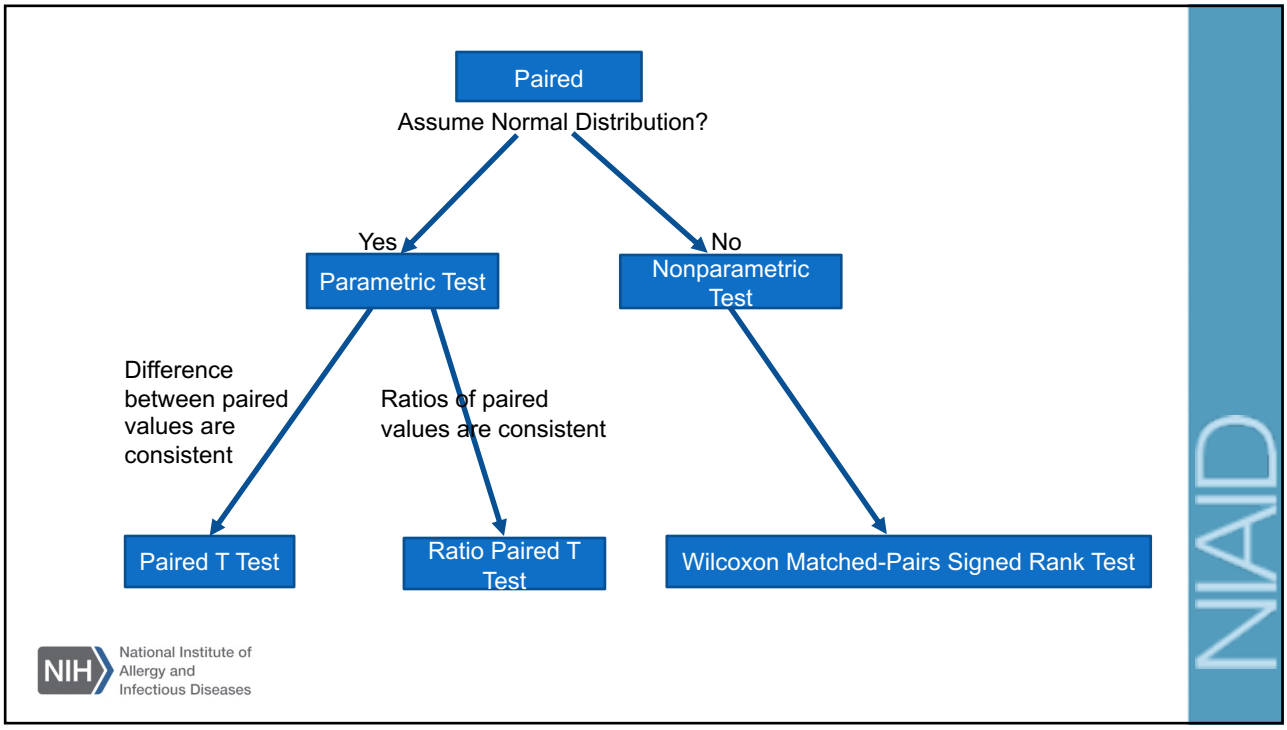
NIH National Institute of Allergy and Infectious Diseases

NIAID



NIH National Institute of Allergy and Infectious Diseases

NIAID



Unpaired t test Example in Prism:

There are two groups, the male group which includes 5 subjects while the female group includes 6 subjects:

	Group A	Group B
	Male	Female
	Y	Y
1	54	43
2	23	34
3	45	65
4	54	77
5	45	46
6		65

NIH National Institute of Allergy and Infectious Diseases
NIAID

Normal distribution?

Prism offers four normality tests. The normality tests all report a P value. To understand any P value, you need to know the null hypothesis. In this case, the null hypothesis is that all the values were sampled from a population that follows a Normal distribution.

- When p-value is high: The data are not inconsistent with a Normal distribution
- When p-value is low: Reject that null hypothesis and so accept the alternative hypothesis that the data are not sampled from a Gaussian population.

[The difference between different normality tests]

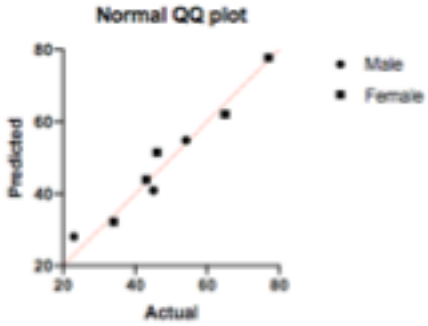
	Normality and Lognormality Tests Tabular results	A Male Y	B Female Y
1	**Test for normal distribution**		
2	**Anderson-Darling test**		
3	A2*	% too small	% too small
4	P value		
5	Passed normality test (alpha=0.05)?		
6	P value summary		
7			
8	**D'Agostino & Pearson test**		
9	K2	% too small	% too small
10	P value		
11	Passed normality test (alpha=0.05)?		
12	P value summary		
13			
14	**Shapiro-Wilk test**		
15	W	0.8070	0.9035
16	P value	0.0923	0.6076
17	Passed normality test (alpha=0.05)?	Yes	Yes
18	P value summary	ns	ns
19			
20	**Kolmogorov-Smirnov test**		
21	KS distance	0.3252	0.2256
22	P value	0.0906	>0.1000
23	Passed normality test (alpha=0.05)?	Yes	Yes
24	P value summary	ns	ns
25			
26	**Number of values**	5	6

National Institute of Allergy and Infectious Diseases

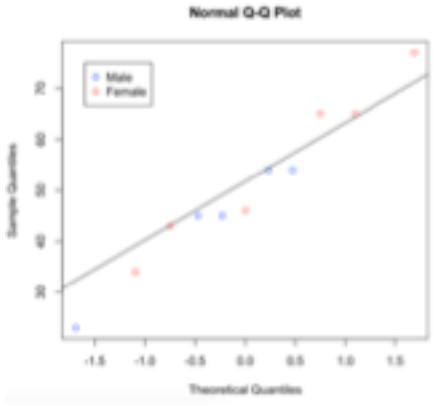
NIAID

QQ Normality Plots

[QQ plot in Prism]



[QQ plot in R]



National Institute of Allergy and Infectious Diseases

NIAID

How to check Same SD?

Formula of F-test

The test statistic can be obtained by computing the ratio of the two variances S_A^2 and S_B^2.

$$F = \frac{S_A^2}{S_B^2}$$

The degrees of freedom are $n_A - 1$ (for the numerator) and $n_B - 1$ (for the denominator).

✔ Note that, the more this ratio deviates from 1, the stronger the evidence for unequal population variances.

✘ Note that, the F-test requires the two samples to be normally distributed.

National Institute of Allergy and Infectious Diseases

NIAID

Check same SDs

In Prism, no matter what t test you choose, unpaired t test or unpaired t test with Welch's correction, it will provide the comparison between variance.
If the answer to the question "Significantly different" is No, unpaired t test is preferred.

```
> res.ftest

        F test to compare two variances

data:  Y by Gender
F = 1.6881, num df = 5, denom df = 4, p-value = 0.6354
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1794174 12.4127534
sample estimates:
ratio of variances
          1.688149
```

	Unpaired t test			Welch's t test	
1	Table Analyzed	Unpaired t test data	1	Table Analyzed	Unpaired t test data
2			2		
3	Column B	Female	3	Column B	Female
4	vs.	vs.	4	vs.	vs.
5	Column A	Male	5	Column A	Male
6			6		
7	**Unpaired t test**		7	**Unpaired t test with Welch's correction**	
8	P value	0.2613	8	P value	0.2501
9	P value summary	ns	9	P value summary	ns
10	Significantly different (P < 0.05)?	No	10	Significantly different (P < 0.05)?	No
11	One- or two-tailed P value?	Two-tailed	11	One- or two-tailed P value?	Two-tailed
12	t, df	t=1.166, df=9	12	Welch-corrected t, df	t=1.230, df=8.972
13			13		
14	**How big is the difference?**		14	**How big is the difference?**	
15	Mean of column A	44.20	15	Mean of column A	44.20
16	Mean of column B	55.00	16	Mean of column B	55.00
17	Difference between means (B - A) ± SEM	10.80 ± 9.010	17	Difference between means (B - A) ± SEM	10.80 ± 8.793
18	95% confidence interval	-9.583 to 31.18	18	95% confidence interval	-9.078 to 30.68
19	R squared (eta squared)	0.1377	19	R squared (eta squared)	0.1442
20			20		
21	**F test to compare variances**		21	**F test to compare variances**	
22	F, DFn, DFd	1.680, 5, 4	22	F, DFn, DFd	1.680, 5, 4
23	P value	0.6354	23	P value	0.6354
24	P value summary	ns	24	P value summary	ns
25	Significantly different (P < 0.05)?	No	25	Significantly different (P < 0.05)?	No
26			26		
27	**Data analyzed**		27	**Data analyzed**	
28	Sample size, column A	5	28	Sample size, column A	5
29	Sample size, column B	6	29	Sample size, column B	6

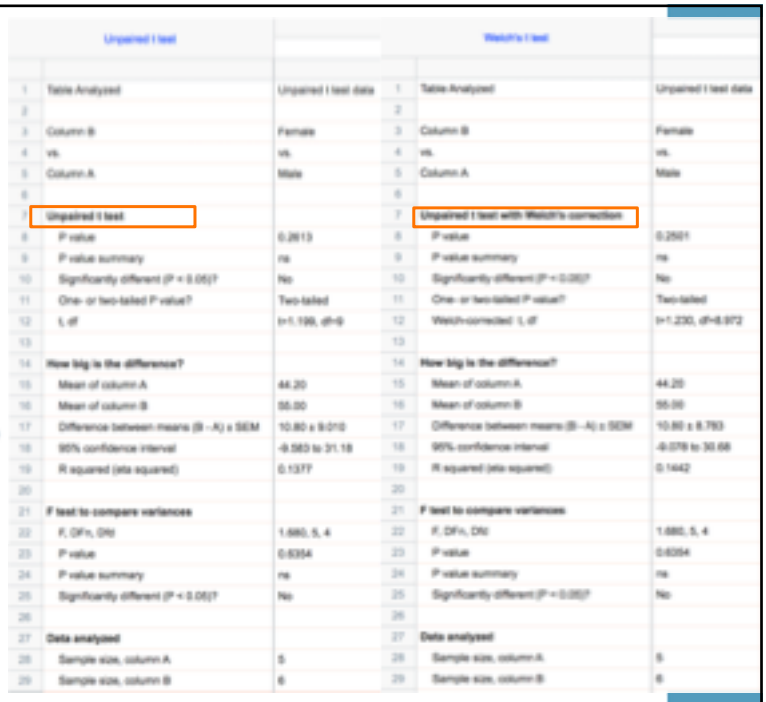
National Institute of Allergy and Infectious Diseases

Paired T-Test

```
> res

        Two Sample t-test

data:  Y by Gender
t = 1.1986, df = 9, p-value = 0.2613
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9.583046 31.183046
sample estimates:
mean in group Female   mean in group Male
                55.0                 44.2
```



	Unpaired t test	
1	Table Analyzed	Unpaired t test data
2		
3	Column B	Female
4	vs.	vs.
5	Column A	Male
6		
7	Unpaired t test	
8	P value	0.2613
9	P value summary	ns
10	Significantly different (P < 0.05)?	No
11	One- or two-tailed P value?	Two-tailed
12	t, df	t=1.199, df=9
13		
14	How big is the difference?	
15	Mean of column A	44.20
16	Mean of column B	55.00
17	Difference between means (B - A) ± SEM	10.80 ± 9.010
18	95% confidence interval	-9.583 to 31.18
19	R squared (eta squared)	0.1377
20		
21	F test to compare variances	
22	F, DFn, Dfd	1.680, 5, 4
23	P value	0.6334
24	P value summary	ns
25	Significantly different (P < 0.05)?	No
26		
27	Data analyzed	
28	Sample size, column A	5
29	Sample size, column B	6

	Welch's t test	
1	Table Analyzed	Unpaired t test data
2		
3	Column B	Female
4	vs.	vs.
5	Column A	Male
6		
7	Unpaired t test with Welch's correction	
8	P value	0.2501
9	P value summary	ns
10	Significantly different (P < 0.05)?	No
11	One- or two-tailed P value?	Two-tailed
12	Welch-corrected t, df	t=1.230, df=6.972
13		
14	How big is the difference?	
15	Mean of column A	44.20
16	Mean of column B	55.00
17	Difference between means (B - A) ± SEM	10.80 ± 8.783
18	95% confidence interval	-9.078 to 30.68
19	R squared (eta squared)	0.1442
20		
21	F test to compare variances	
22	F, DFn, Dfd	1.680, 5, 4
23	P value	0.6334
24	P value summary	ns
25	Significantly different (P < 0.05)?	No
26		
27	Data analyzed	
28	Sample size, column A	5
29	Sample size, column B	6

National Institute of Allergy and Infectious Diseases

How to explain the results?

- When the **p-value** of the test is less than the significance level alpha = 0.05, We can conclude that male's average is significantly different from female's average.
- When the **p-value** of the test is larger than the significance level alpha = 0.05, we *cannot* conclude that a significant difference exists between male and female.

National Institute of Allergy and Infectious Diseases

NIAID

Try It!

~ 5 mins

- Try the same process of unpaired two-sample t-test in Prism
- Try to run the R codes
- Compare the results of Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

Outline

1. Statistical Testing Process
   - 1.1. Formulate Null and Alternative hypotheses
   - 1.2. Calculate the appropriate test statistics
   - 1.3. Compute the probability of observing the test statistic under the null hypothesis
   - 1.4. Make a statistical decision
   - 1.5. Make a biological conclusion
2. Common Statistical Tests
   - 2.1.Choose appropriate tests
   - 2.2.Two-sample T-test
   - **2.3.One-way ANOVA**
   - 2.4.Two-way ANOVA
   - 2.5.Multiple Comparisons
3. Application in Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

2.3.One-Way ANOVA

The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are statistically significantly different from each other.

Specifically, it tests the null hypothesis:

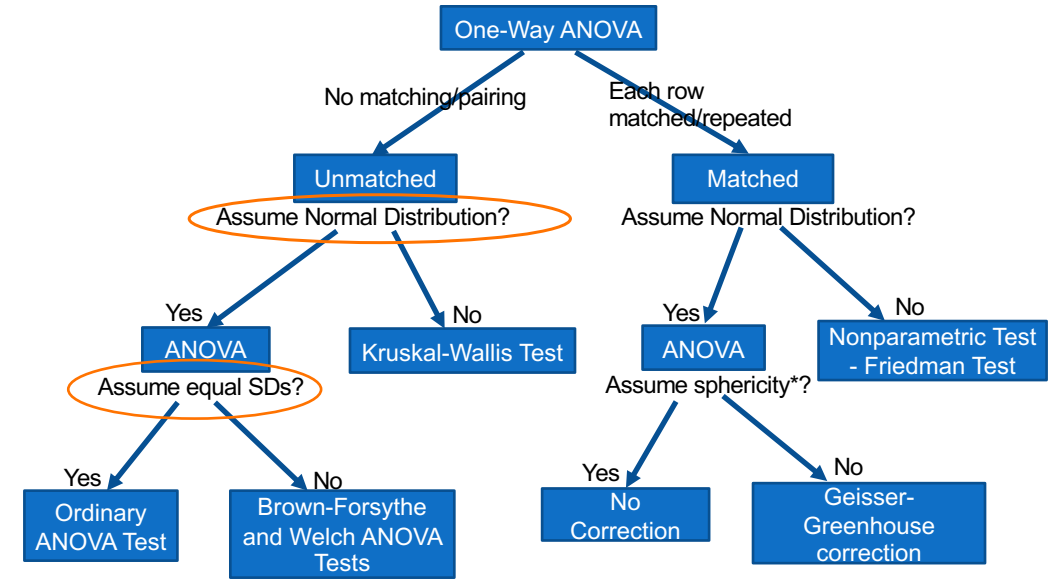
$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

Where μ = group mean and k = number of groups. If the one-way ANOVA returns a statistically significant result, we reject the null hypothesis and conclude that that there are at least two group means that are statistically significantly different from each other.

Note: One-way ANOVA is an **omnibus** test statistic

National Institute of Allergy and Infectious Diseases

NIAID



National Institute of Allergy and Infectious Diseases

*Sphericity: Equal variability of differences

NIAID

Sphericity

Briefly sphericity means that you waited long enough between treatments for any treatment effect to wash away.

		A	B	C	D
		Control	Treatment 1	Treatment 2	Treatment 3
		Y	Y	Y	Y
1	GS	54	43	78	111
2	JM	23	34	65	99
3	HM	45	65	99	78
4	DR	54	77	79	90
5	PS	45	46	87	95

Should we assume sphericity?

- If your experimental design relies on matching rather than repeated measurements, then you can assume sphericity, as violations are essentially impossible.
- If your experiment design is repeated measures, we recommend that you do not assume sphericity.

In this case, it is ANOVA with repeated measures.

NIH National Institute of Allergy and Infectious Diseases

NIAID

Unmatched One-Way ANOVA example

The following is an example provided by Prism. In this case, there are three groups: *Control* (n = 6), *Treated* (n = 5) and *Treated + Antagonist* (n = 6).

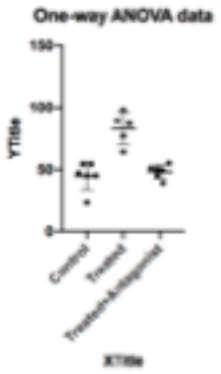
	Group A	Group B	Group C
	Control	Treated	Treated+Antagonist
	Y	Y	Y
1	54	87	45
2	23	98	39
3	45	64	51
4	54	77	49
5	45	89	50
6	47		55

NIH National Institute of Allergy and Infectious Diseases

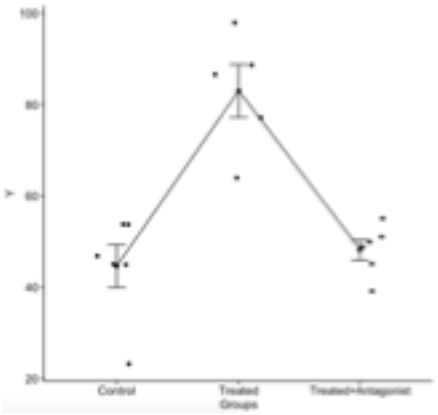
NIAID

Plots

box-plot in Prism



box-plot in R



NIH National Institute of Allergy and Infectious Diseases

NIAID

Considering that the first group, the control group doesn't pass the normality test, it is better to use nonparametric test – Kruskal-Wallis Test

	Normality and Lognormality Tests Tabular results	A Control	B Treated	C Treated+Antagonist
		Y	Y	Y
1	**Test for normal distribution**			
2	**Anderson-Darling test**			
3	A2*	N too small	N too small	N too small
4	P value			
5	Passed normality test (alpha=0.05)?			
6	P value summary			
7				
8	**D'Agostino & Pearson test**			
9	K2	N too small	N too small	N too small
10	P value			
11	Passed normality test (alpha=0.05)?			
12	P value summary			
13				
14	**Shapiro-Wilk test**			
15	W	0.7690	0.9643	0.9517
16	P value	0.0467	0.8376	0.7537
17	Passed normality test (alpha=0.05)?	No	Yes	Yes
18	P value summary	*	ns	ns
19				
20	**Kolmogorov-Smirnov test**			
21	KS distance	0.3450	0.2210	0.2268
22	P value	0.0245	>0.1000	>0.1000
23	Passed normality test (alpha=0.05)?	No	Yes	Yes
24	P value summary	*	ns	ns
25				
26	**Number of values**	6	5	6

NIH National Institute of Allergy and Infectious Diseases

NIAID

Kruskal-Wallis Test?

Kruskal-Wallis test by ranks is also called as one-way ANOVA on ranks, which extends the Mann–Whitney U test, which is used for comparing only two groups

To perform this test, Prism first ranks all the values from low to high, paying no attention to which group each value belongs. The smallest number gets a rank of 1. The largest number gets a rank of N, where N is the total number of values in all the groups. The discrepancies among the rank sums are combined to create a single value called the Kruskal-Wallis statistic (some books refer to this value as H). A large Kruskal-Wallis statistic corresponds to a large discrepancy among rank sums.

- As the p-value is less than the significance level 0.05, we can conclude that there are significant differences between these treatment groups.

```
> kruskal.test(Y ~ Group, data = unmatched_one_way_anova)

        Kruskal-Wallis rank sum test

data:  Y by Group
Kruskal-Wallis chi-squared = 10.18, df = 2, p-value = 0.006158
```

	Kruskal-Wallis test	
1	Table Analyzed	One-way ANOVA data
2		
3	**Kruskal-Wallis test**	
4	P value	0.0015
5	Exact or approximate P value?	Exact
6	P value summary	**
7	Do the medians vary signif. (P < 0.05)?	Yes
8	Number of groups	3
9	Kruskal-Wallis statistic	10.18
10		
11	**Data summary**	
12	Number of treatments (columns)	3
13	Number of values (total)	17

NIH National Institute of Allergy and Infectious Diseases

NIAID

Assumption: Normal distribution?

The assumption - Normal distribution, is not too important with large samples, but it is important with small sample sizes (especially with unequal sample sizes).

If your data do not come from Normal distributions, you have three options:

- *Your best option is to transform the values (perhaps to logs or reciprocals) to make the distributions more normal.*
- *Use the Kruskal-Wallis nonparametric test instead of ANOVA.*
- *Use ANOVA anyway, knowing that it is fairly robust to violations of a Normal distribution with large samples.*

National Institute of Allergy and Infectious Diseases

NIAID

Assumption: same standard deviation?

- The assumption, same standard deviation, is not very important when all the groups have the same (or almost the same) number of subjects, but is very important when sample sizes differ.
- Prism tests for equality of variance with two tests: The Browne-Forsythe test and Bartlett's test.
- Don't base your conclusion solely on these tests. Also think about data from other similar experiments. If you have plenty of previous data that convinces you that the variances are really equal, ignore these tests (unless the P value is really tiny) and interpret the ANOVA results as usual.

National Institute of Allergy and Infectious Diseases

NIAID

Let's run ANOVA anyway to see the output provided by Prism and R.

** BUT in this case, ANOVA might not be acceptable because of the violation of Normal distribution.

```
> summary(res.aov)
            Df Sum Sq Mean Sq F value   Pr(>F)
Group        2   4760  2380.2   22.57 4.16e-05 ***
Residuals   14   1476   105.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ordinary one-way ANOVA

1	Table Analyzed	One-way ANOVA data
2	Data sets analyzed	A-C
3		
4	**ANOVA summary**	
5	F	22.57
6	P value	<0.0001
7	P value summary	****
8	Significant diff. among means (P < 0.05)?	Yes
9	R square	0.7633
10		
11	**Brown-Forsythe test**	
12	F (DFn, DFd)	0.7307 (2, 14)
13	P value	0.4991
14	P value summary	ns
15	Are SDs significantly different (P < 0.05)?	No
16		
17	**Bartlett's test**	
18	Bartlett's statistic (corrected)	2.966
19	P value	0.2247
20	P value summary	ns
21	Are SDs significantly different (P < 0.05)?	No
22		
23	**ANOVA table**	**SS** **DF** **MS** **F (DFn, DFd)** **P value**
24	Treatment (between columns)	4760 2 2380 F (2, 14) = 22.57 P<0.0001
25	Residual (within columns)	1476 14 105.4
26	Total	6236 16
27		
28	**Data summary**	
29	Number of treatments (columns)	3
30	Number of values (total)	17

National Institute of Allergy and Infectious Diseases

Try It!

~ 5 mins

- Try the same process of unmatched one-way ANOVA in Prism
- Try to run the R codes
- Compare the results of Prism and R

National Institute of Allergy and Infectious Diseases

NIAID

Repeated measures one-way ANOVA example:

This example is provided by Prism. The four columns represent four sequential treatments. Each row represents a different subject (or a different set of matched data).

Repeated-measures ANOVA compares the means of three or more matched groups.

Table format: Column		Group A	Group B	Group C	Group D
		Control	Treatment 1	Treatment 2	Treatment 3
		Y	Y	Y	Y
1	GS	54	43	78	111
2	JM	23	34	37	41
3	HM	45	65	99	78
4	DR	31	33	36	35
5	PS	15	25	30	26

NIH National Institute of Allergy and Infectious Diseases

NIAID

Check if it follows normal distribution

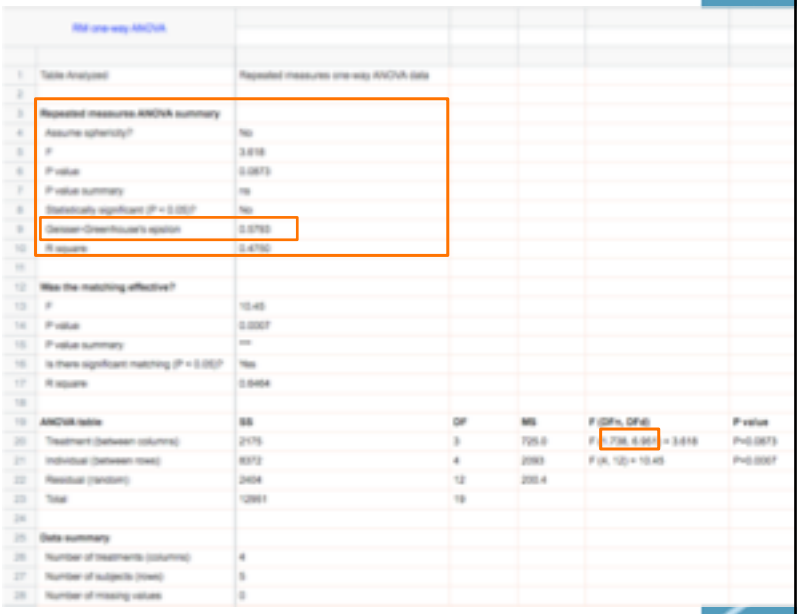
And the answer is YES!

Normality and Lognormality Tests Tabular results		A	B	C	D
		Control	Treatment 1	Treatment 2	Treatment 3
		Y	Y	Y	Y
1	**Test for normal distribution**				
2	**Anderson-Darling test**				
3	A2*	N too small	N too small	N too small	N too small
4	P value				
5	Passed normality test (alpha=0.05)?				
6	P value summary				
7					
8	**D'Agostino & Pearson test**				
9	K2	N too small	N too small	N too small	N too small
10	P value				
11	Passed normality test (alpha=0.05)?				
12	P value summary				
13					
14	**Shapiro-Wilk test**				
15	W	0.9646	0.8908	0.8317	0.8854
16	P value	0.8393	0.3613	0.1434	0.3345
17	Passed normality test (alpha=0.05)?	Yes	Yes	Yes	Yes
18	P value summary	ns	ns	ns	ns
19					
20	**Kolmogorov-Smirnov test**				
21	KS distance	0.1649	0.2519	0.3320	0.2858
22	P value	>0.1000	>0.1000	0.0750	>0.1000
23	Passed normality test (alpha=0.05)?	Yes	Yes	Yes	Yes
24	P value summary	ns	ns	ns	ns
25					
26	**Number of values**	5	5	5	5

NIH National Institute of Allergy and Infectious Diseases

NIAID

- The P value answers this question: If all the populations really have the same mean, what is the chance that random sampling would result in means as far apart (or more so) as observed in this experiment?
- If the overall P value is large: the data does not give you any reason to conclude that the means differ.
- If the overall P value is small: it is unlikely that the differences you observed are due to random sampling. You can reject the idea that all the populations have identical means.



1	Table Analyzed	Repeated measures one-way ANOVA data				
2						
3	Repeated measures ANOVA summary					
4	Assume sphericity?	No				
5	F	3.618				
6	P value	0.0673				
7	P value summary	ns				
8	Statistically significant (P < 0.05)?	No				
9	Geisser-Greenhouse's epsilon	0.5793				
10	R square	0.4750				
11						
12	Was the matching effective?					
13	F	10.45				
14	P value	0.0007				
15	P value summary	***				
16	Is there significant matching (P < 0.05)?	Yes				
17	R square	0.6464				
18						
19	ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
20	Treatment (between columns)	2175	3	725.0	F (3, 12) = 3.618	P=0.0673
21	Individual (between rows)	8372	4	2093	F (4, 12) = 10.45	P=0.0007
22	Residual (random)	2404	12	200.4		
23	Total	12951	19			
24						
25	Data summary					
26	Number of treatments (columns)	4				
27	Number of subjects (rows)	5				
28	Number of missing values	0				

National Institute of Allergy and Infectious Diseases

```
> summary(model1,multivariate=F)

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

            Sum Sq num Df Error SS den Df F value  Pr(>F)
(Intercept)  44086      1   8371.7      4 21.0643 0.01011 *
Trials        2175      3   2404.3     12  3.6184 0.04549 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Mauchly Tests for Sphericity

       Test statistic p-value
Trials       0.057644 0.19416


Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

       GG eps Pr(>F[GG])
Trials 0.57928    0.08727 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

          HF eps Pr(>F[HF])
Trials 0.9856654 0.04648165
```

National Institute of Allergy and Infectious Diseases

NIAID

Try It!

~ 5 mins

- Try the same process of Repeated measures one-way ANOVA
- Try to run the R codes
- Compare the results of Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

Outline

1. Statistical Testing Process
   - 1.1. Formulate Null and Alternative hypotheses
   - 1.2. Calculate the appropriate test statistics
   - 1.3. Compute the probability of observing the test statistic under the null hypothesis
   - 1.4. Make a statistical decision
   - 1.5. Make a biological conclusion
2. Common Statistical Tests
   - 2.1.Choose appropriate tests
   - 2.2.Two-sample T-test
   - 2.3.One-way ANOVA
   - **2.4.Two-way ANOVA**
   - 2.5.Multiple Comparisons
3. Application in Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

2.4.Two-way ANOVA

Two-way ANOVA, also called two-factor ANOVA, is used to evaluate simultaneously the effect of two grouping variables (A and B) on a response variable.

Two-way ANOVA test hypotheses

1. There is no difference in the means of factor A
2. There is no difference in means of factor B
3. There is no interaction between factors A and B

The alternative hypothesis for cases 1 and 2 is: the means are not equal.

The alternative hypothesis for case 3 is: there is an interaction between A and B.

NIH National Institute of Allergy and Infectious Diseases

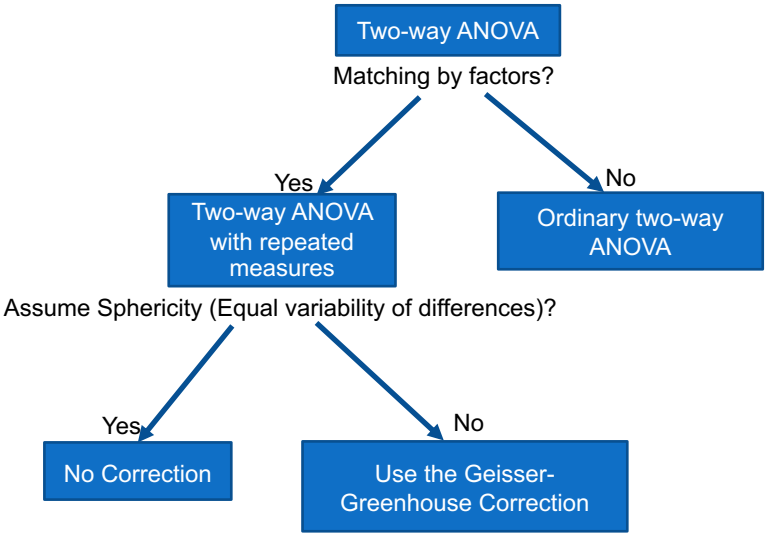
NIAID

One-Way vs Two-Way ANOVA Differences Chart

	One-Way ANOVA	Two-Way ANOVA
Definition	A test that allows one to make comparisons between the means of three or more groups of data.	A test that allows one to make comparisons between the means of three or more groups of data, where two independent variables are considered.
Number of Independent Variables	One.	Two.
What is Being Compared?	The means of three or more groups of an independent variable on a dependent variable.	The effect of multiple groups of two independent variables on a dependent variable and on each other.
Number of Groups of Samples	Three or more.	Each variable should have multiple samples.

NIH National Institute of Allergy and Infectious Diseases

NIAID



National Institute of Allergy and Infectious Diseases

NIAID

Ordinary Two-way ANOVA Example

The columns represent two cell lines. The rows represent two treatments. Within each treatment for each cell line, five replicate values are entered into subcolumns. This experiment has no matching or repeated measures. For this reason, it is OK that one of the values is missing

Table format: Grouped		Group A					Group B				
		Wild-type cells					GPP5 cell line				
		A:Y1	A:Y2	A:Y3	A:Y4	A:Y5	B:Y1	B:Y2	B:Y3	B:Y4	B:Y5
1	Serum starved	34	36	41		43	98	87	95	99	88
2	Normal culture	23	19	26	29	25	32	29	26	33	30

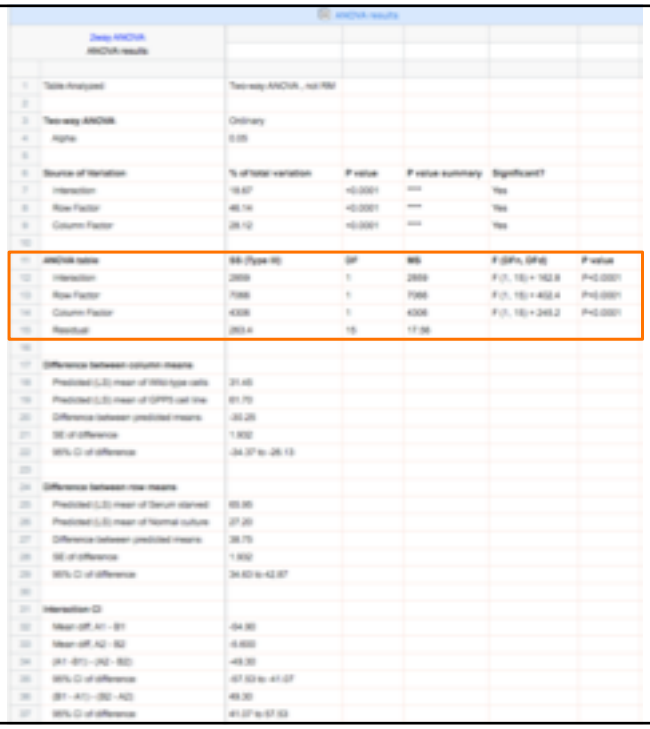
National Institute of Allergy and Infectious Diseases

NIAID

ANOVA table

The ANOVA table breaks down the overall variability between measurements (expressed as the sum of squares) into four components:

- *Interactions between row and column.*
- *Variability among columns.*
- *Variability among rows.*
- *Residual or error. Variation among replicates not related to systematic differences between rows and columns.*



National Institute of Allergy and Infectious Diseases

NIAID

Interaction P value

The null hypothesis is that there is no interaction between columns and rows. Often the test of interaction is the most important of the three tests. For example, If columns represent drugs and rows represent gender, then the null hypothesis is that the differences between the drugs are consistent for men and women.

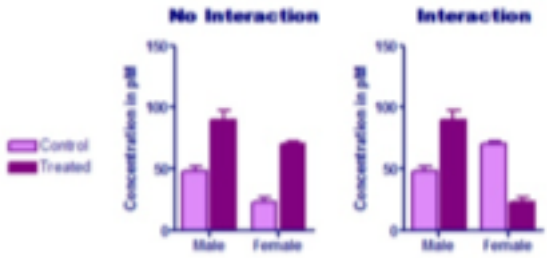
```
> summary(res.aov2)
             Df Sum Sq Mean Sq F value   Pr(>F)
Treatment     1   8276    8276   471.3 9.58e-13 ***
Cell          1   3917    3917   223.1 2.06e-10 ***
Treatment:Cell  1   2859    2859   162.8 1.86e-09 ***
Residuals    15    263      18
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1 observation deleted due to missingness
```

National Institute of Allergy and Infectious Diseases

NIAID

The graph on the left below shows no interaction. The treatment has about the same effect in males and females. The graph on the right, in contrast, shows a huge interaction. the effect of the treatment is completely different in males (treatment increases the concentration) and females (where the treatment decreases the concentration).

In this two-way ANOVA example, the test for interaction leads to statistically significant results, we probably won't learn anything of interest from the other two P values. In the example above, a statistically significant interaction means that the effect of the treatment (difference between treated and control) differs between cell groups.



National Institute of Allergy and Infectious Diseases

NIAID

Try It!

~ 5 mins

- Try the same process of Ordinary Two-way ANOVA
- Try to run the R codes
- Compare the results of Prism and R

National Institute of Allergy and Infectious Diseases

NIAID

Outline

1. Statistical Testing Process
   - 1.1. Formulate Null and Alternative hypotheses
   - 1.2. Calculate the appropriate test statistics
   - 1.3. Compute the probability of observing the test statistic under the null hypothesis
   - 1.4. Make a statistical decision
   - 1.5. Make a biological conclusion
2. Common Statistical Tests
   - 2.1.Choose appropriate tests
   - 2.2.Two-sample T-test
   - 2.3.One-way ANOVA
   - 2.4.Two-way ANOVA
   - **2.5.Multiple Comparisons**
3. Application in Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

2.5.Multiple Comparisons

Once an Analysis of Variance (ANOVA) test has been completed, the researcher may still need to understand subgroup differences among the different experimental and control groups. The subgroup differences are called *"pairwise"* differences. ANOVA does not provide tests of pairwise differences. When the researcher needs to test pairwise differences, follow-up tests called *post hoc* tests are required.

NIH National Institute of Allergy and Infectious Diseases

NIAID

Ordinary Two-way ANOVA Example (with multiple comparisons)

The columns represent two cell lines. The rows represent two treatments. Within each treatment for each cell line, five replicate values are entered into subcolumns. This experiment has no matching or repeated measures. For this reason, it is OK that one of the values is missing

Table format: Grouped		Group A					Group B				
		Wild-type cells					GPP5 cell line				
		A:Y1	A:Y2	A:Y3	A:Y4	A:Y5	B:Y1	B:Y2	B:Y3	B:Y4	B:Y5
1	Serum starved	34	36	41		43	98	87	96	99	88
2	Normal culture	23	19	26	29	25	32	29	26	33	30

Tukey multiple pairwise-comparisons

```
> TukeyHSD(res.aov2)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Y ~ Treatment + Cell + Treatment:Cell, data = two_way_anova)

$Treatment
                              diff      lwr      upr p adj
Serum Starved-Normal Culture 41.8 37.69614 45.90386     0

$Cell
                                    diff       lwr       upr p adj
Wild-Type cells-GPP5 cell line -28.71111 -32.81497 -24.60725     0

$`Treatment:Cell`
                                                                diff        lwr       upr     p adj
Serum Starved:GPP5 cell line-Normal Culture:GPP5 cell line     63.4  55.7614904  71.03851 0.0000000
Normal Culture:Wild-Type cells-Normal Culture:GPP5 cell line  -5.6 -13.2385096   2.03851 0.1936359
Serum Starved:Wild-Type cells-Normal Culture:GPP5 cell line    8.5   0.3981371  16.60186 0.0382505
Normal Culture:Wild-Type cells-Serum Starved:GPP5 cell line  -69.0 -76.6385096 -61.36149 0.0000000
Serum Starved:Wild-Type cells-Serum Starved:GPP5 cell line   -54.9 -63.0018629 -46.79814 0.0000000
Serum Starved:Wild-Type cells-Normal Culture:Wild-Type cells  14.1   5.9981371  22.20186 0.0007910
```

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value
Serum starved:Wild-type cells vs. Serum starved:GPP5 cell line	-54.90	-63.00 to -46.80	Yes	****	<0.0001
Serum starved:Wild-type cells vs. Normal culture:Wild-type cells	14.10	5.998 to 22.20	Yes	***	0.0008
Serum starved:Wild-type cells vs. Normal culture:GPP5 cell line	8.500	0.3981 to 16.60	Yes	*	0.0383
Serum starved:GPP5 cell line vs. Normal culture:Wild-type cells	69.00	61.36 to 76.64	Yes	****	<0.0001
Serum starved:GPP5 cell line vs. Normal culture:GPP5 cell line	63.40	55.76 to 71.04	Yes	****	<0.0001
Normal culture:Wild-type cells vs. Normal culture:GPP5 cell line	-5.600	-13.24 to 2.039	No	ns	0.1936

NIH National Institute of Allergy and Infectious Diseases

NIAID

Try It!

~ 5 mins

- Try the same process of Ordinary Two-way ANOVA (with multiple comparisons)
- Try to run the R codes
- Compare the results of Prism and R

NIH National Institute of Allergy and Infectious Diseases

NIAID

Two interesting facts about Multiple Comparisons

If the overall ANOVA finds a significant difference among groups, am I certain to find a significant post test?

If one-way ANOVA reports a P value of <0.05, you reject the null hypothesis that all the data come from populations with the same mean. In this case, it seems to make sense that at least one of the follow-up multiple comparisons tests will find a significant difference between pairs of means.

This is not necessarily true!!!

Because It is possible that the overall mean of group A and group B combined differs significantly from the combined mean of groups C, D and E.

NIH National Institute of Allergy and Infectious Diseases

NIAID

Two interesting facts about Multiple Comparisons

If the overall ANOVA finds no significant difference among groups, are the multiple comparisons test results valid?

You may find it surprising, but all the multiple comparisons tests offered by Prism are valid even if the overall ANOVA did not find a significant difference among means.

NIH National Institute of Allergy and Infectious Diseases

NIAID

Conclusions

- Statistical Testing methods are much more than these!
- The materials of this seminar will be updated on: Github - Statistical Testing
  fork us if you are interested in further update!
- If you have any specific statistical problem, please send email to
  bioinformatics@niaid.nih.gov
- Check our training schedule: http://www.eventzilla.net/user/NIAID_OCICB_BCBB
- Take the survey below and tell us what statistical topics you want to learn about:
  BCBB Statistical Training – Suggest a class!

NIH National Institute of Allergy and Infectious Diseases

NIAID

Thank You!

NIH National Institute of Allergy and Infectious Diseases

NIAID