the jungle of statistical tests

a survival guide

the jungle of statistical tests : a survival guide

1 THE HYPOTHESES

Running a power analysis saves you from the painful discovery that your study was underpowered all along.

Are you predicting a difference, direction, or relationship?

Be explicit about what you expect: a difference between means, an increase through time, or a correlation between two variables. If this is unclear, your analysis and conclusions may be inaccurate.

Is your hypothesis driven by prior evidence?

Ground your hypothesis in existing literature or solid preliminary data, not intuition alone. Even the shiniest p-value from a single isolated experiment won't save you, if every other study in the field shows the opposite.

What's your null hypothesis?

Write your hypothesis as a precise statement linking variables, not a vague research aim. If no possible result could prove you're wrong, it's not a scientific hypothesis, it's a belief.

"Gene X knock-out reduces cell viability by at least 20% compared to wild-type under standard culture conditions."

2 THE DATA

Have you visualised your raw data?

Plot individual data points to understand distribution, spread, and group overlap before applying any test. Not just the mean ± SEM. It can hide patterns that completely change interpretation.

Balanced designs are more robust to deviations from test assumptions. Think of it as statistical insurance.

Have you decided on data handling rules in advance?

Computing ratios, filtering out some data or normalising is not neutral. It can have a huge impact on the conclusions. Think carefully about whether these methods are being used effectively and honestly.

Most statistical tests can handle slight changes from a perfect bell-shaped curve, so a quick look at your histogram is usually enough.

Are assumptions even plausible?

Check assumptions using plots or simple diagnostics rather than assuming they hold. If they clearly fail, choose a different test instead of hoping reviewers won't notice.

3 THE HOLY P-VALUE

Every additional test you run increases the probability of false positives. If you're testing multiple groups or outcomes, you must correct for multiple comparisons.

When choosing a significance level, you have to consider a trade-off between Type I and Type II errors. The only way to reduce both types of error is to collect bigger samples.

P-values are not very reproducible

Don't treat your p-value like a gold medal. They are far more slippery than we'd like to admit and not very reproducible.

P-values vary according to the number of samples

With a large enough sample size, you can make almost anything statistically significant, including effects so tiny they are scientifically meaningless. More data doesn't always mean more truth.

P-values are just a tool among others

A p-value alone tells you surprisingly little. Pair it with effect sizes, confidence intervals, and context, or you're navigating with only half a map. Think of it as one instrument in an orchestra, not the entire symphony.

P-values bigger than 0.05 do not mean there is no effect

A p-value above 0.05 simply means your data didn't scream loud enough to cross an arbitrary threshold. Absence of statistical significance is not the same as evidence of absence.

4 THE TESTS

Are your replicates independent?

You cannot combine biological & technical replicates. This is called pseudoreplication. It is the same than asking the same person the same question ten times and calling it ten independent opinions.

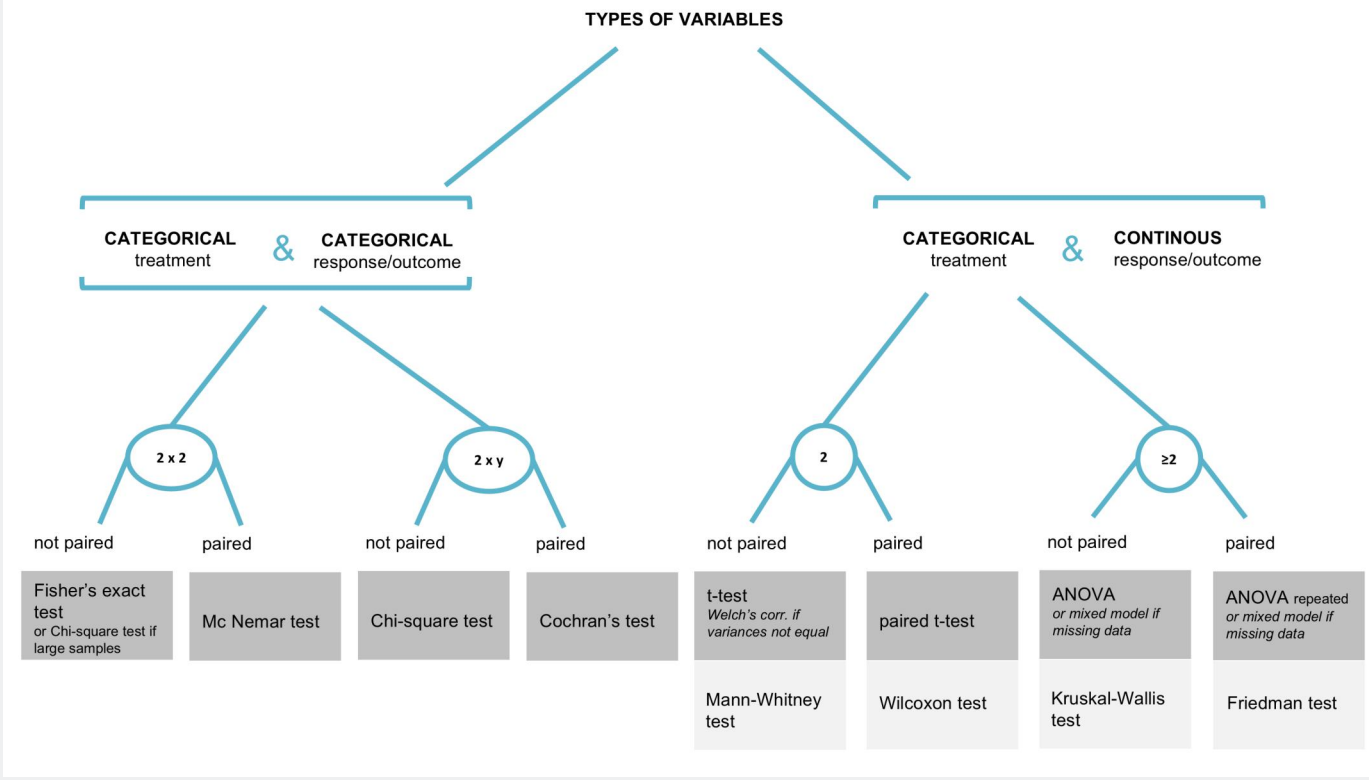
Can you justify the choice of the test?

Choosing a statistical test is not about picking the one your lab has always used. It depends on your research question, the structure of your data, and the assumptions you can reasonably make. If unsure, check out the decision tree on next page.

One-tailed or two-tailed?

Use a one-tailed test only if a result in the opposite direction would be completely irrelevant and uninterpretable. In practice, this is rare. When in doubt, use two-tailed.

Non-parametric tests are not inferior fallbacks. When your data are skewed or your sample size is small, non-parametric tests are the correct choice.



5 THE CONCLUSIONS

Statistical significance ≠ Proof

Rejecting the null hypothesis does not prove your alternative hypothesis. It tells you that the effect is unlikely to be zero, but not how big the effect is. Always report an effect size.

Negative results are results

A well-powered study that fails to reject Ho tells you something useful: the effect, if it exists, is smaller than you can detect with your current assay. Report it honestly. Publishing only positive results distorts the scientific literature.

The scope of your conclusion

Your conclusion must match the scope of your study. If you used male C57BL/6 mice, extrapolating to female mice, all mammals, or humans requires additional evidence.

A p-value below 0.05 is not a reason to stop thinking.
Same for the opposite.

your bookmark checklist

Cut on the black line, print and fold on the dotted line, then keep close by.

- ☐ is your hypothesis driven by prior evidence?
- ☐ do you know what's your null hypothesis?

- ☐ have you visualised your raw data?
- ☐ have you decided on data handling rules in advance?
- ☐ are assumptions even plausible?

- ☐ are your replicates independent?
- ☐ can you justify the choice of the test?
- ☐ one-tailed or two-tailed?

- ☐ aware that statistical significance is not proof?
- ☐ do you report effect size, context?
- ☐ do you report negative results?
- ☐ does your conclusion match the scope of your study?

"If you torture your data long enough,
they will tell you whatever you want to hear."

Mills, 1993