Experimental Design
XDASI Fall 2021

11/8/2021

Contents

Background Reading

- **Whitlock & Schluter, Chapter 14**

Which statistical test should I use?

Several basic methods for **hypothesis testing** are available that allow us to determine whether observed differences between two or more groups are statistically different. The choice of an appropriate test will depend on what kind of question you are interested in asking.

The diagram below illustrates how to choose the right test depending on your question. **Parametric** tests compare differences between samples based on model assumptions: t-tests, ANOVA, simple linear models, and linear models with mixed effects (i.e. interaction terms). **Non-parametric tests** enable comparisons when these assumptions do not hold. There are also tests for **correlations** and **categorical** data. The diagram below illustrates a decision tree
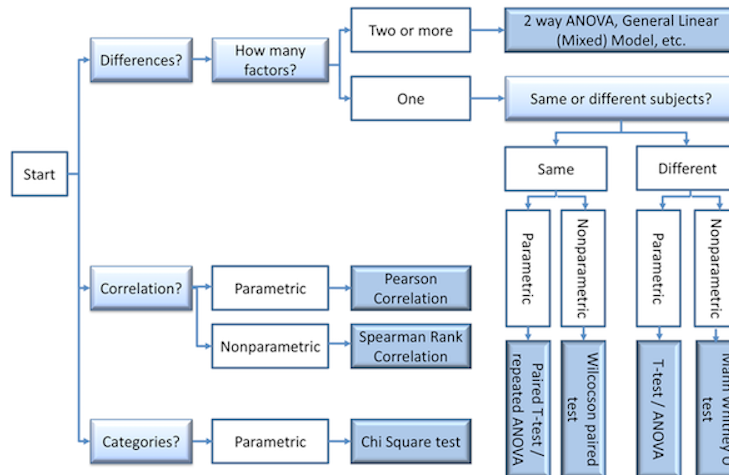
Figure 1: Decision tree for choosing a statistical test

Review: Hypothesis testing

- Type of data ⇒ Type of test

  - Categorical, Categorical-Numeric, Ordinal, Interval

- Comparing two samples ⇒ Normality, equal variance, skew

  - Graphical methods, statistical tests ⇒ Approach

- Approaches

  - Ignore violations of assumptions ⇒ Small deviations, large N (CLT)
    - ∗ Equal variances? ⇒ Type of test
  - Transformations ⇒ Adjust for moderate skew (vs extreme, outliers)
  - Nonparametric ⇒ Non-normal distributions, similar skew
  - Permutation tests ⇒ Resampling to generate null for test statistic

- Bootstrapping ⇒ Estimate sampling distribution of an estimate by resampling

  - Boostrap SE and CI's ⇒ Unknown distributions

Experimental Design

When designing an experiment we are always trying to reduce the chance of having extraneous factors influence our results. Variation that we cannot account for represents experimental ***noise***. This will reduce our ***power*** to detect true effects, as well as our ***precision*** in estimating the expectation value for an experimental outcome. It can also result in ***sampling bias***, if there is some correlation between extraneous variables and the response variable(s) we are trying to measure.

Therefore, we need to keep in mind a few simple concepts to reduce the effects of experimental noise:

⇒ **Reducing bias**

- Simultaneous controls
- Randomization
- Blinding

⇒ **Reducing sampling error**

- Replication
- Balance
- Blocking

⇒ **Experimental units**

- *Individuals* or *groups* of individuals assigned ***independently***

Reducing bias

Simultaneous control group

⇒ Only known variable is the treatment (match with similar conditions)

- Placebo
- "Sham" or "mock" treatment
- Avoid inadvertent experimental perturbation

Randomization

⇒ Break association (correlation) of confounding and explanatory variables by spreading variation more evenly

⇒ Account for any variation you can think of!

Suppose we are collecting plants from a plot, and we have four different treatments collected in the fashion described below.

A	A	B	B
A	A	B	B
C	C	D	D
C	C	D	D

Even though we have collected 4 sample of each, the layout isn't random, and we may be inadvertently introducing some bias due to the layout.

Unconscious bias There are many sources of **unconscious bias**, and a good experimental design is essential to avoid this. For example, we may pick flowers that emerge earliest, or the first 10 eggs that are laid, or colonies that are a little bigger than others, without realizing that this could bias our conclusions.

Blinding

- Conceal information about which units receive which treatment
    - Single blind: subjects only
    - Double blind: subjects and researchers

Reducing sampling error

⇒ Reduces **precision** and **power**

- Controlled conditions
    - Caveat ⇒ may limit generality
- Extreme treatments ⇒ increase effect sizes
    - Caveat ⇒ may be qualitatively different

Replication

Performing the same experiment, or simply collecting data several times on **independent** experimental units, improves the chances that our estimates are closer to the expected values.

For normally distributed data, the variation is inversely proportional to sample size. We see this in the calculation of **standard error**, where:

$$SE = \frac{s}{\sqrt{n}}$$

⇒ **As n, the number of samples, increases, the standard error will decrease.**

Pseudoreplication   Special care must be taken so that we don't confuse *replicates* with *pseudoreplicates*. Pseudo-replicates are when we take measurements on experimental units that are essentially the same sample. For example, if we sequence the same library twice, this is really a **technical replicate**, not a **biological replicate**. This is because we are really testing the variability of the sequencing run and not the variability among independent biological samples. Similarly, picking plants from a plot gives a greater chance that the conditions they experiience are very similar, and thus might not reflect the true variation within a larger population (e.g. a field or forest). Even though this will give us less variability, it is not a true representation of the biology.

⇒ **What other methods of sampling could result in pseudoreplicates?**

Balance

Treatments with the same / similar sample sizes reduce the influence of sampling error. Why is this?

Recall the formula for the standard error of the difference in sample means:

$$SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{s_p^2 (\frac{1}{n_1} + \frac{1}{n_2})}$$

It is easy to see that **the SEM is minimized when** $n_1 = n_2$. Try this yourself for $N = n_1 + n_2 = 20$.

Several related points are worth keeping in mind:

- Precision is limited by the precision of the smaller sample.
- Larger sample sizes are always better since they provide better estimates of sample means!
- Balanced designs are more robust to violations of assumptions of equal variance, since the variance of smaller samples is usually larger.

Blocking

Sometimes we can not avoid variation in external factors, so we should find a way to include them in our model. This will allow us to account for the influence of the factors in the variation of the data.

You've already learned that a good way to avoid ***bias*** is by performing randomized, paired or matched trials within multiple experimental groups, or **blocks**.

Blocking also minimizes the effects of extraneous variation by using groups that share common features. The general approach is:

- Intersperse random experimental units within blocks
- Repeat randomized experiments within blocks
- Evaluate differences between treatments only within blocks
- Accounting for variation will increase power and precision of estimates.

Latin Square Design One of the best ways to design an experiment is using a **Latin Square Design**. In this case all the different blocks are represented equally for all treatments. For the example above, it would look something like this:

A	B	D	C
D	A	C	B
C	D	B	A
B	C	A	D

Here our samples would be collected randomly and the factors that vary across rows or columns will be taken into account. For example, a field trial with different pesticide treatments may experience a gradient of water availability or variations in sunlight across the field.

In an ANOVA model in R (which we will cover soon), you simply provide the Blocking variable in a **formula**, so you can calculate the amount of variance that can be explained by this factor and it will be deducted from the variation explained from your treatment. This will provide you with a more proper estimation of how the variation in your data relates to the treatment.