This overview is organized in pairs of concepts, following chapters in the lecture notes.

1

Experimental vs. observational unit

- EU: the smallest unit to which a treatment is applied
- OU: the smallest unit from which a response is collected

We studied two cases: EU = OU and each EU contains several OU. Can you give examples for each case?

Treatment structure vs. plot structure

e.g. Each patient is a block and the plots (teeth) are structured in blocks

Treatment structure is only about the relations of the treatments to each other (for example, factorial designs). Plot structure is only about the relations of the plots to each other, regardless of the treatments (for example, blocks).

with this term, the oats experiments is consist of six blocks, each having three main plots, each of which is split into four subplots (varietis are the main plot factor and nitrogen is sub-plot factor)

Observational studies vs. experiments

Sometimes it is not possible to randomize treatments to experimental units. This is the setting in observational studies. Experiments always involve randomizing the treatments to the experimental units.

2

Know your data! vs. know your models!

Both a good overview over the data at hand (including the design) and a solid understanding of the statistical properties of our models are needed. Accordingly, plot your data before modeling.

Location vs. variability

Equality of variance, the different mean means shift of the location

Treatments can affect the mean, but also the variance of the outcome variable. Be aware of this and look for this when plotting the data.

Parametric vs. nonparametric methods

The classical parametric methods assume that the data (actually, the residuals) come from a normal distribution and all have the same variance.

Signal vs. noise

The overall F test is based on dividing the mean between sum of squares by the mean within sum of squares. It essentially compares the strength of the signal (MSB) to the noise level (MSW).

`lm(e)` vs. `aov`

Linear (mixed) models or calculating sums of squares: both approaches are possible and yield the same results in many simple cases. For missing/unbalanced data, the former approaches are often a bit better.

Numerator vs. denominator degrees of freedom

In the F test, we have numerator degrees of freedom (related to: how many terms are tested?) and denominator degrees of freedom (related to the residual degrees of freedom).

Observed values vs. fitted values

Comparing these two yields the residuals (observed values minus fitted values). We use these for model diagnostics.

True vs. estimated standard deviation

We use the root mean square error (RMSE) to estimate the unknown standard deviation of the error terms.

Explained vs. unexplained proportion of the variance

The coefficient of determination R^2 is the proportion of the variance of the dependent variable explained by the model.

3

A vs. B vs. C vs. ...

After a significant F test, the aim is often to conduct further comparisons. This is dictated by the aim of the experiment. Comparing all treatment means pairwisely is one example. In all of these tests, be aware and know how to account for the multiple testing problem. We studied:

- All pairwise comparisons with t or Wilcoxon tests.
- Tukey's HSD. If normality is given, I would use this over pairwise t tests. Do not use in case of problems with normality. Often used with a compact letter display.

- Treatments vs. control. All treatments are compared to the control treatment.

Treatment contrasts vs. sum contrasts

The former choose a reference level to which things are compared; in the latter setting, the comparison is to the grand mean (average of all the treatments). No adjustment for multiple testing in both cases.

4

Ideal vs. reality

Model assumptions (normality, equal variances) are not always fulfilled and should be checked as a routine part of each analysis. Some simpler cases of what to do in case of problems are discussed in a separate handout.

Good points vs. bad points

Be aware of the influence that a single outlier may have on the results, especially for small sample sizes.

5

Blocked vs. completely randomized designs

Plot structure may often by accounted for by collecting similar experimental units in blocks. Treatments are then randomized only within blocks, as opposed to the CRD, where randomization is across all the plots. In the RCBD, each treatment is observed exactly one time per block.

Good blocks vs. bad blocks

Good blocks all have the same size (same number of plots) and are sufficiently large to apply each treatment at least once per block. Both rules may be broken, but then life becomes more complicated.

Fixed vs. random effect models

Simple decision rule that often works: if you are interested in exactly the levels that occurred in the study, treat as fixed effect. If the levels are a random sample from a bigger population, treat as random effect.

6

ML vs. REML estimation

These are two methods to estimate parameters in mixed models. ML is useful to compare models with different fixed effects (use for model selection), REML gives better variance estimates (use for final model). In this module, we mainly studied models with a random intercept for the blocks.

Including vs. excluding random effects in fitted values

To investigate the model fit, you should include the random effects; if you want to compare fixed effects, you may exclude them.

Tests vs. confidence intervals

While tests of the model's coefficients only inform you about the significance (model summary), confidence intervals for the coefficients give an easier to understand statement about the precision of the estimation.

`glht` vs. `emmeans`

Technically slightly different ways of getting different treatment comparisons. Often both produce similar results.

ICC vs. R^2

While in mixed models, there is no classical R^2, some alternatives exist.

RCBD: Parametric vs. nonparametric analysis

The sample size in RCBD is often really small; in such cases, the Friedman test offers a good nonparametric approach that does not rely on normality.

GRCBD: block $\times$ treatment interaction vs. no interaction

In GRCBD, we can model block $\times$ treatment interactions to test (with the likelihood ratio test) whether we need to account for different treatment effects depending on the blocks.

7

Interaction vs. no interaction

The first question to ask in factorial designs with factors A and B: does the effect of A depend on the level of B? Often visualized with interaction plots.

Should I stay or should I go

A significant interaction term must not be removed from the model. A non-significant interaction effect may be removed if it is the aim to have a simple model. Effects involved in higher-order interaction terms should not be removed from the model.

Pooling levels vs. keeping them separate

Comparing two factors must be done carefully in case their interaction is significant. One option is to compare the levels of A separately for each level of B.

Balanced vs. unbalanced designs

In unbalanced designs, we have to distinguish between sequential (Type I) and marginal F tests (Type II). Most of the time, the latter should be fine. Also Type III tests are often useful.

Equal vs. unequal group variances

With sufficient data, it is possible to estimate separate variances for subsets of the data (for example by the levels of one factor). This is often very useful to deal with heteroskedasticity problems.

8

Classical ANOVA tables vs. linear (mixed) models

For balanced, homoskedastic data, the choice does not matter. For more complex settings, the linear (mixed) model is more flexible. Both approaches can deal with nested designs.

9

Complete vs. incomplete block designs

In incomplete block designs, the blocks are too small to accommodate all treatments. If you have to use such a design (because larger blocks are not feasible or would have an unreasonably high variance), try to use a balanced incomplete block design.

Adjusted vs. unadjusted data

Because the data are not balanced, you should adjust for block effects when comparing treatments. Otherwise you risk to confound block effects with treatment effects.

10

To split or not to split?

Split plot designs are very popular in case one of the factors has levels which are harder to change. This is then the main plot factor, and its effects are estimated less precisely than the effects of the sub plot factor, which is randomized within the main plots only.

Ordered vs. unordered factors

Equidistant numeric factors may be treated as ordered factors, which allows to simplify the model often. For example, they may only have a linear effect.

11

With or without you?

Covariates measured during the experiment may often help to explain the values of the response and should be used in the model.

12

One vs. two blocking factors

Row-column designs (such as Latin squares) may be used in case the design requires two blocking factors.

Block effects vs. treatment effects

In factorial experiments with blocks, we can use confounded designs to make estimation possible in situations with very small data sets and some treatment interactions which may be neglected.

Fractional factorial experiments vs. factorial experiments

In case some interactions are zero, we may use fractions of factorial designs, which can massively reduce the resources used for the experiment. Some contrasts will be aliased as a result.

13

Power vs. sample size

You should always calculate the required sample size before conducting an experiment. This is often neglected, resulting in underpowered hypothesis tests (most often).

Design D_1 vs. Design D_2

To compare the statistical properties of designs, you can also use the relative efficiency to help you choose the better design. This can also be done with power calculations.

Planning vs. doing

Always carefully elaborate the protocol before conducting an experiment. You would also not just go to the lab and start mixing stuff and see how it works during the experiment. A careful protocol can save you very much work.

作者：宁大爷链接：https://www.zhihu.com/question/57802751/answer/154646233来源：知乎著作权归作者所有。商业转载请联系作者获得授权，非商业转载请注明出处。
首先题主要知道什么叫假设检验。别人给了我们一个假设(Null Hypothesis)，我们需要根据某种准则(Criterion)，来判断我们做的这个假设到底是真的(Not Reject)还是假的(Reject)。这种准则具体在计算机上就是你看到的那个p值(p-value)，如果它小于0.05就拒绝假设，大于0.05就不拒绝假设。
题主需要自己看书学的部分是“某种准则是什么”，它的学名叫“假设检验原理”。如果题主想真正彻底搞明白假设检验(Hypothesis Testing)到底是什么东西，而不是一知半解只知道做题的话，一定要看这个，绝对是值得的。其次题主要知道什么叫参数检验(Parametric test)什么叫非参数检验(Nonparametric test)。所谓参数检验，意思是说样本的总体是已知的某个分布(Distribution)，这个分布是必须在确定了具体的参数(Parameter)后才能最终确定下来。我们掌握着各种关于这个分布的信息，进而我们就可以通过某种方法来做假设检验。所谓非参数检验，意思是说样本的总体不是已知的某个分布，我们只能根据基于数据/观测(Data/Observation)本身的某种方法来做假设检验。题主需要自己看书学的部分是“分布的信息是什么”和“某种方法是什么”，它们的学名叫“常见分布”和“服从某分布的统计量的构造&拒绝域的构造”
最后针对题主列举的这些检验，我可以大致说说它们针对的都是几样本问题，具体的(原)假设是什么。
参数检验:
Independent sample t-test(独立样本t检验)单样本问题，均值为给定值
Paired sample t-test(成对样本t检验)两样本问题，两总体均值相等ANOVA/Analysis of Variance(方差分析)多样本问题，多总体均值相等
非参数检验:
Wilcoxon Signed Rank test(威尔科克森符号秩检验)单样本问题，对称总体的均值为给定值
Mann-Whitney test(曼惠特尼U统计量检验)两样本问题，两个同分布的总体是完全相同的(平移量为0)
Kruskal-Wallis test(好像没有中文名)多样本问题，多个同分布的总体是完全相同的(平移量为0)