

Install.packages(rstatix)

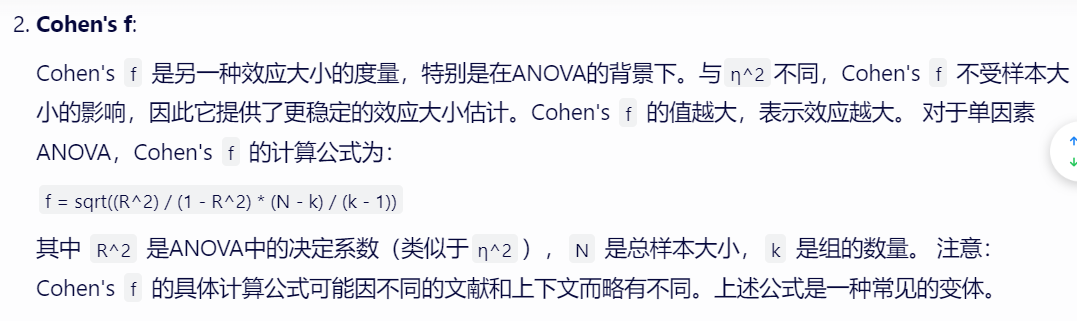
library(rstatix) # 真的好用

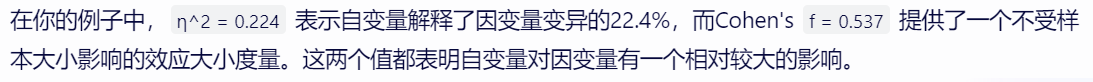
eta\_squared(model)

library(pwr) # 真的好用

pwr.anova.test() # 计算anova的power

library(stats)中的power.anova.test()也可以用，但是要提供between.var和within.var，比上面的pwr.anova.test()要麻烦





Bartlett's test of homogeneity of variances（巴特利特方差齐性检验）是一种用于检验多个独立样本是否来自具有相同方差的总体的一种非参数方法。它通常用于在方差分析（ANOVA）之前检查各组之间的方差是否相等，因为ANOVA的一个基本假设就是各组之间的方差相等（方差齐性）。

Homoscedasticity

data\_list <- list(group1 = group1, group2 = group2, group3 = group3)

# 执行Bartlett方差齐性检验

bartlett\_result <- bartlett.test(data\_list)

Or

bartlett.test(dif~Treatment, data = dbnew)

代替了

Equality of variance:

• Can be checked visually (plot(anova\_model, 1)) or by running a suitable hypothesis test

课堂笔记：

ANOVA tests the null hypothesis i.e.

“There is no difference between any of the groups”

or

“The group does not influence the response”

Assumptions for ANOVA

• Independent random sampling (1. We can assume it at once. 2.Cannot be assessed from data set itself. We have to believe that this is true given the description of the experiment itself. )

• Normality of residuals (distances from group mean)

plot (model, 2)

• Equality of Variances

plot(model, 1)

bartlett.test(pain ~ treatment, data = trial)

确保子组中有足够的观测值进行方差检验（通常至少 5 个）

bartlett test is not right for two-way ANOVA

When use for one-way ANOVA, highly recommend all groups are of equal size.

If group size vary too much, can also use Levene's Test(use median)

leveneTest(weight\_gain ~ genotype \* diet, data = mouse)

可以做two-way ANOVA

Solution: Tukey's HSD test

• Honestly Signifcant Difference

• This runs multiple comparisons, with the appropriate corrections of p values to account for multiple testing

When to use ANOVA?

1. More than one population
2. More than 1 predictive variable (factor)

1-way ANOVA → 1 factor (e.g. effect of 3 doses of a drug on heart rate)

2-way ANOVA → 2 factors (e.g. effect of age and sex on salary)

3-way ANOVA → 3 factors (e.g. effect of age, sex and education on salary)

Interacton

If more than one factor is included, then the response to one factor may be affected by the other factor(s). This is called an interaction.

• H0: There is no effect of class attendance or previous grades on course performance

• HA: At least one of those factors (class attendance or previous grades) influences course performance.

Additional hypothesis if test for interaction

• H0: There is no interaction between class attendance and previous grades

• HA: There is an interaction between class attendance and previous grades

文心一言给出的答案是

Aov(a~b\*c)

H0:

变量b对变量a没有显著影响。

变量c对变量a没有显著影响。

变量b和变量c的交互作用对变量a没有显著影响。

HA:

至少有一个假设不成立，即：

HA\_b: 变量b的至少一个水平对应的a的均值与其他水平对应的均值显著不同。

HA\_c: 变量c的至少一个水平对应的a的均值与其他水平对应的均值显著不同。

HA\_bc: 变量b和变量c的交互作用对a的均值有显著影响。

来自problem set notes

Q-Q plot:

For the two bonus questions, the normal-QQ plot is maybe the most difficult to understand. Remind yourself of the concept of quantile: the x quantile of a dataset is the value such that x fraction of your dataset is below that value. For instance, the 0.5 quantile is the median. The QQ plot asks the question: For each of the residuals in your dataset, what quantile are they in? This is the y axis, labelled “sample quantiles”. Now, for each of the residuals, if they were truly normally distributed, what quantile would they be in? This is the x axis, labelled “theoretical quantiles”.

residual vs fitted

You can take each group mean as the fitted value (x axis) and for each data point, just plot the residual on the y axis.

文心一言

ANOVA（方差分析）的样本大小并没有一个固定的“适合”范围