1. **η^2 (eta-squared)**:

η^2 是效应大小的一个度量，特别是在方差分析（ANOVA）中。它表示由自变量解释的因变量方差的比例。 η^2 的值范围从0到1，其中0表示没有效应，1表示完全效应（即因变量的所有变异都可以由自变量解释）。然而，需要注意的是， η^2 可能会受到样本大小的影响，因此有时可能会给出偏高的效应大小估计。 计算公式（对于单因素ANOVA）：

```
η^2 = SS_between_groups / (SS_between_groups + SS_within_groups)
```

其中 SS_between_groups 是组间平方和， SS_within_groups 是组内平方和。

```
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

2. **Cohen's f**:

Cohen's f 是另一种效应大小的度量，特别是在ANOVA的背景下。与 η^2 不同，Cohen's f 不受样本大小的影响，因此它提供了更稳定的效应大小估计。Cohen's f 的值越大，表示效应越大。 对于单因素ANOVA，Cohen's f 的计算公式为：

```
f = sqrt((R^2) / (1 - R^2) * (N - k) / (k - 1))
```

其中 R^2 是ANOVA中的决定系数（类似于 η^2 ）， N 是总样本大小， k 是组的数量。 注意：Cohen's f 的具体计算公式可能因不同的文献和上下文而略有不同。上述公式是一种常见的变体。

在你的例子中， η^2 = 0.224 表示自变量解释了因变量变异的22.4%，而Cohen's f = 0.537 提供了一个不受样本大小影响的效应大小度量。这两个值都表明自变量对因变量有一个相对较大的影响。

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