

Understanding Data and Statistical Design (60117)

Lab 4: Two factor experiments

This lab is marked from 21.

Please submit via Canvas.

Due by the conclusion of the lab class

In this week's lab we look at two way ANOVA F -tests.

QUESTION 1 [21 marks]

In this question we analyse the effect of sugar content and chocolate type on the flavour of chocolate chip cookies. The variables we consider are summarised in the table below.

Name	Type	Description
<i>score</i>	numerical	flavour rating of cookie: 1-10
<i>sugar</i>	factor	sugar content: 0 (0.5 cups), 1 (0.375 cups), 2 (0.25 cups)
<i>chip</i>	factor	chocolate type: 1 (milk), 2 (semi-sweet), 3 (dark)

The sample data consists of 11 observations of *score* for each combination of levels of *sugar* and *chip* (data available in lab4.csv on Canvas) for a total of 99 observations.

The statistical model for this experiment is

$$score_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk}, \quad i \in \{1,2,3\}, \quad j \in \{1,2,3\}, \quad k \in \{1,2, \dots, 11\},$$

where

- $score_{ijk}$ is flavour rating of the k -th cookie with $sugar = i$ and $chip = j$
- μ is global mean *score*
- α_i is the treatment effect on *score* of $sugar = i$
- β_j is the treatment effect on *score* of $chip = j$
- δ_{ij} is the interaction effect on *score* of $sugar = i$ and $chip = j$
- ϵ_{ijk} is the random effect on *score* of k -th cookie with $sugar = i$ and $chip = j$.

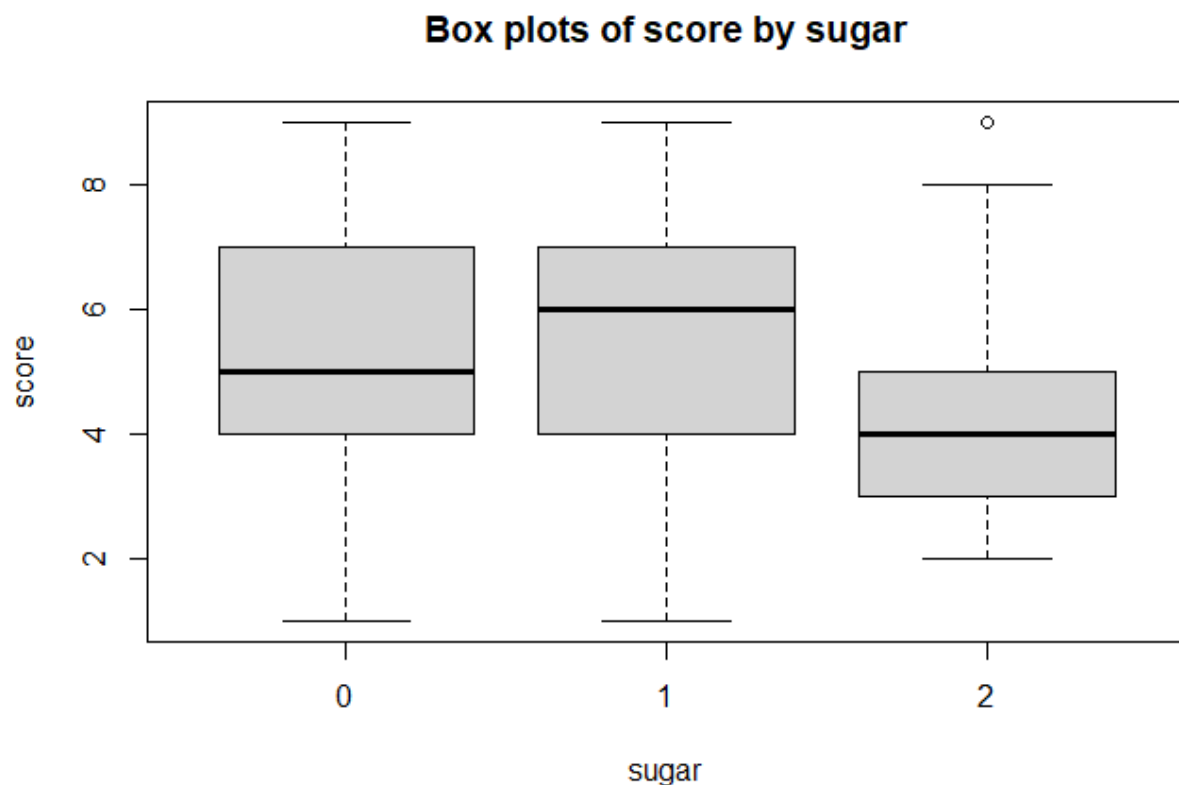
The components of the 3×3 factorial design are:

- **factor A** – sugar content (variable *sugar*) which has 3 **levels**: 0 (0.5 cups), 1 (0.375 cups), 2 (0.25 cups)
- **factor B** – chocolate type (variable *chip*) which has 3 **levels**: 1 (milk), 2 (semi-sweet), 3 (dark)
- **treatments** – the 9 combinations of levels of the factors
- **experimental units** – the 9 groups of 11 cookies each prepared with a different treatment
- **measurement units** – the 99 cookies used in the experiment
- **response variable** – flavour rating (variable *score*).

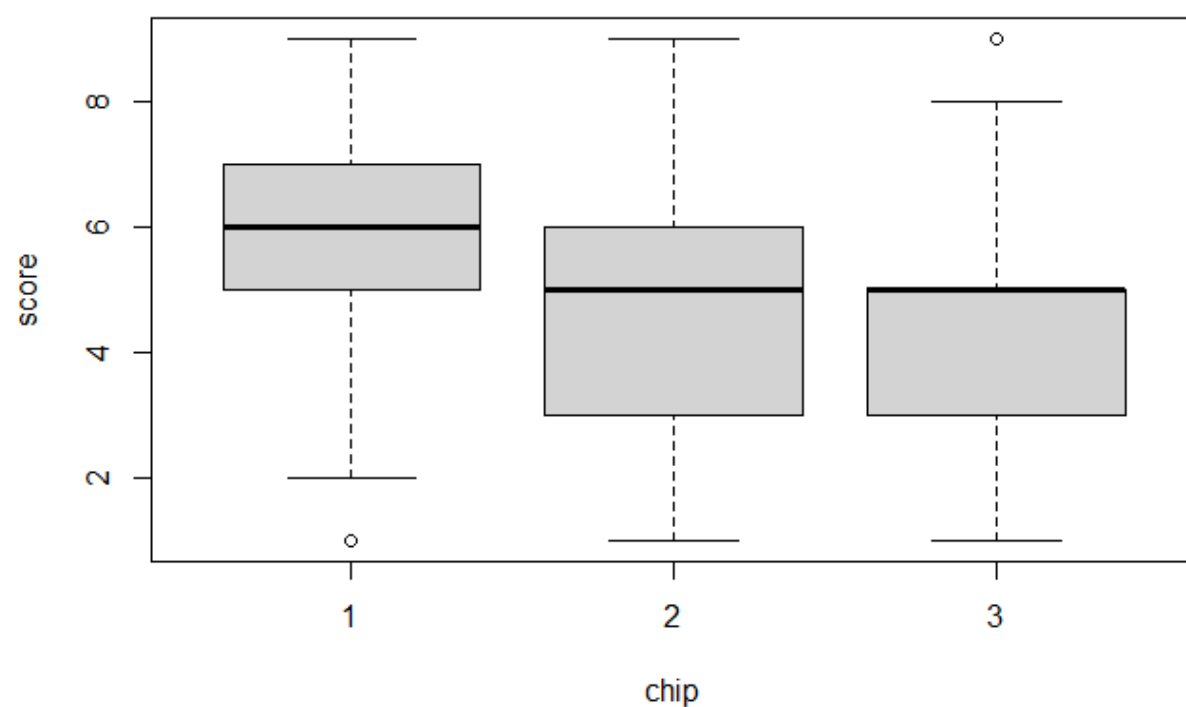
Note that the response variable *score* is a discrete variable, so we know it cannot be normally distributed. Irrespective of this, we will apply *F*-tests to the data and determine from the residual analysis whether application of these tests seems appropriate.

Plots

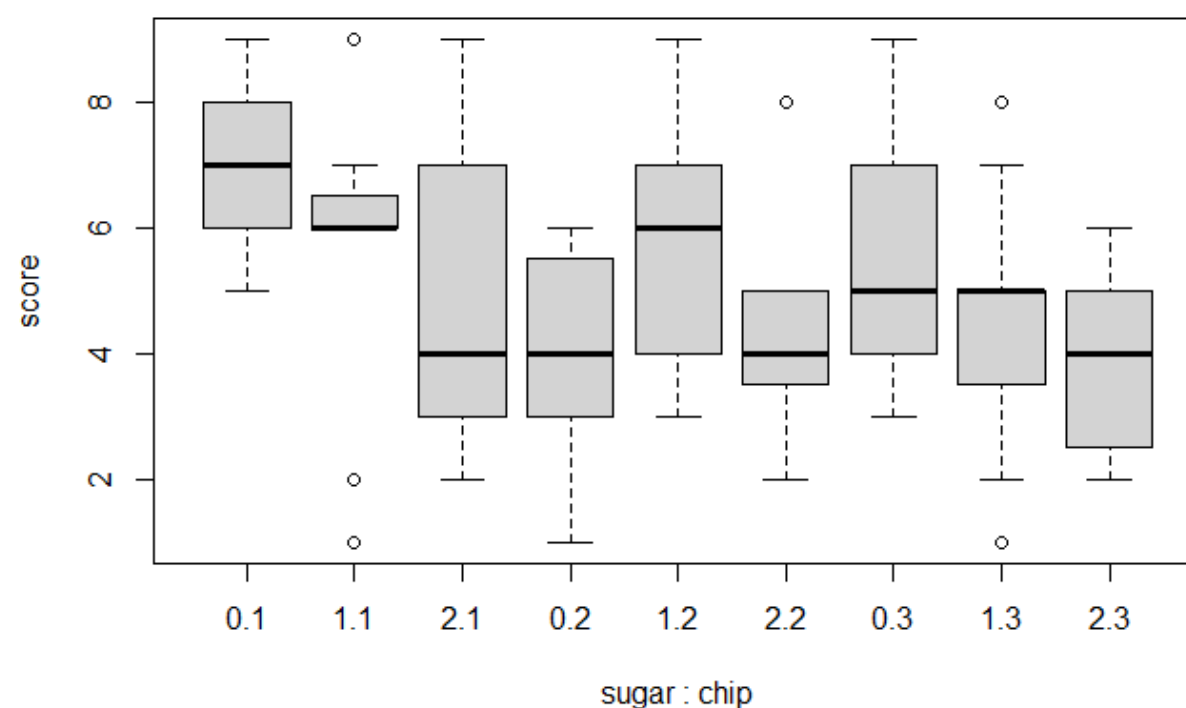
To start we produce some boxplots of *score* by levels of factors *sugar* and *chip* and combinations of these.



Box plots of score by chip



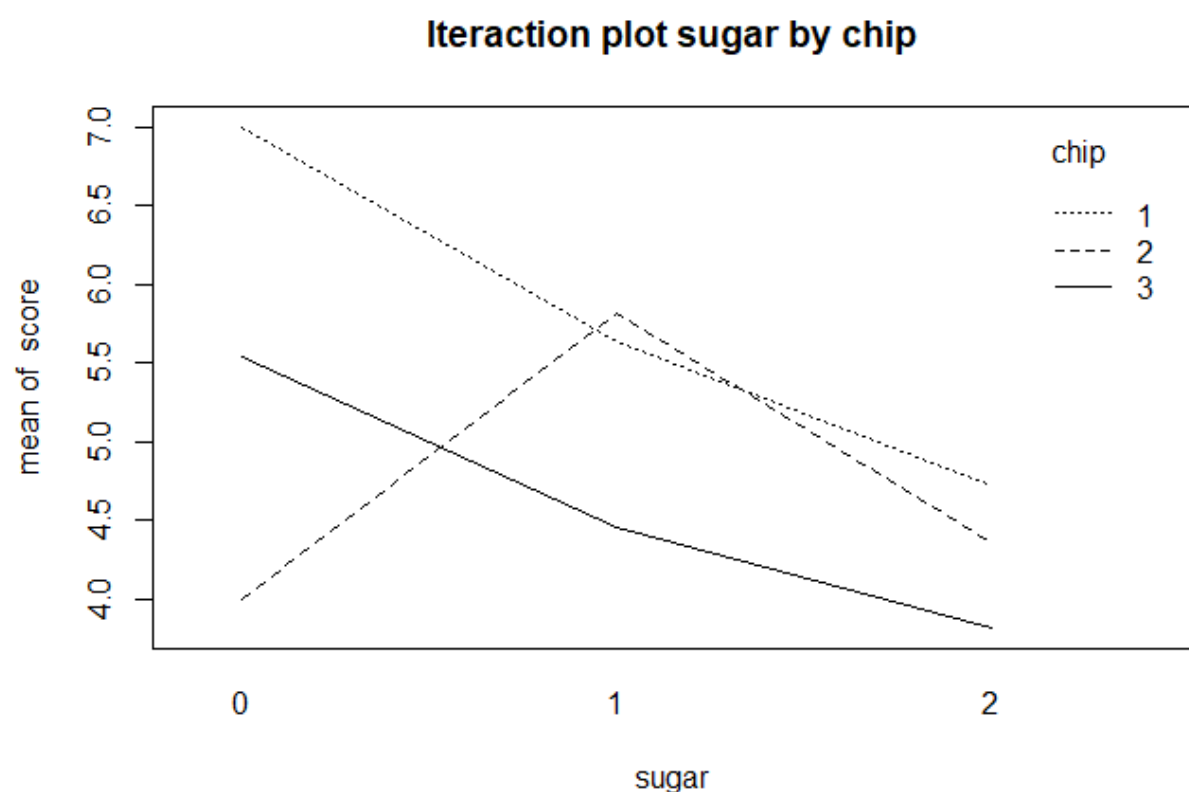
Box plots of score by sugar and chip

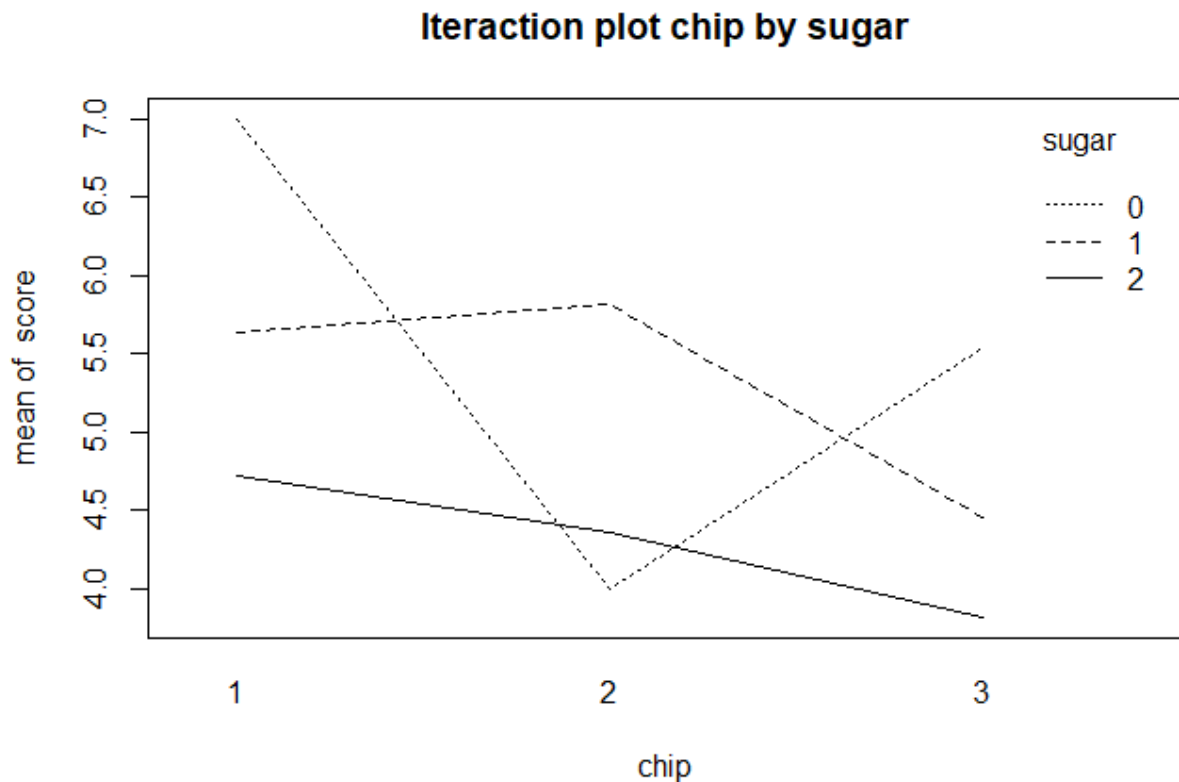


The first chart show the between-treatment variation is dominated by the within-treatment variation, so it is not clear from the chart whether the factor *sugar* will turn out to have a statistically-significant effect on *score*. Similar comments can be made about the second chart and the effect of the factor *chip* on *score*.

The third chart does show some large between-treatment variation for some combinations of levels of factors *sugar* and *chip*, so it appears likely that a significant interaction effect between these factors will be found.

We can investigate interaction effects using the interaction plots below.





The non-parallel trace lines in both plots suggest the interaction between factors could be significant.

- (a) Referring to the interaction plots above, rank the average flavour ratings of cookies made with the 3 types of chocolate using 0.5 cups of sugar **[3 marks]**.

Analysing the plot, the chip factor 1 has the highest mean score, followed by the factor 3 and last factor 2.

ANOVA and F-tests

Perform two-factor ANOVA (with interaction) and associated F -tests to determine the significance of the two factors.

- (b) Using significance level $\alpha = 0.05$, document the interaction term F -test from the two way ANOVA. Write down the null and alternative hypotheses, the test statistic and associated p-value, the test decision (with reason) and a conclusion using a minimum of mathematical language **[3 marks]**.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sugar	2	27.7	13.828	3.726	0.0279	*
chip	2	27.9	13.949	3.759	0.0270	*
sugar:chip	4	38.3	9.571	2.579	0.0426	*
Residuals	90	334.0	3.711			

F-test: 2.579

P_value: 0.0426

Ho: There is no significance difference between the groups of the interaction between sugar:chips

Ha: There is significance difference between the groups of the interaction between sugar:chips

Test decision: We have strong evidence to reject the null hypotheses, since the $p_value < \alpha$

Conclusion: There is a least one of the groups with a significance difference.

(c) Consider again the interaction F -test from (b). Write down R code that takes as input the interaction test statistic and returns the p-value [3 marks].

`pf(q = 2.579, df1=4, df2=90, lower.tail = FALSE)`

```
> pf(q = 2.579, df1=4, df2=90, lower.tail = FALSE)
[1] 0.04261623
```

Tukey post hoc analysis

From the results of the F -tests we know that both factors are significant, as is their interaction. Now we perform post-hoc analysis for each factor. Due to the interaction term, we need to do this for each level of the other factor.

(d) Using significance level $\alpha = 0.05$, perform Tukey post-hoc analysis for factor *sugar* by each level of *chip*. Summarise the findings [3 marks].

```
chip = 1:
contrast      estimate      SE df lower.CL upper.CL t.ratio p.value
sugar0 - sugar1  1.364 0.821 90   -0.594   3.321   1.660 0.2262
sugar0 - sugar2  2.273 0.821 90    0.315   4.230   2.767 0.0187
sugar1 - sugar2  0.909 0.821 90   -1.048   2.867   1.107 0.5124

chip = 2:
contrast      estimate      SE df lower.CL upper.CL t.ratio p.value
sugar0 - sugar1 -1.818 0.821 90   -3.776    0.139  -2.213 0.0743
sugar0 - sugar2 -0.364 0.821 90   -2.321    1.594  -0.443 0.8977
sugar1 - sugar2  1.455 0.821 90   -0.503    3.412   1.771 0.1853

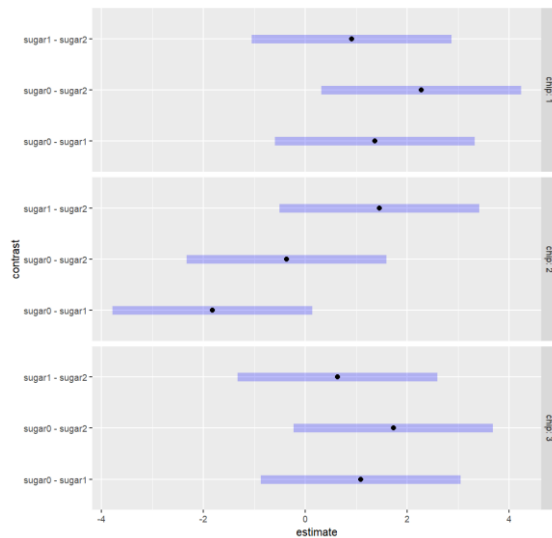
chip = 3:
contrast      estimate      SE df lower.CL upper.CL t.ratio p.value
sugar0 - sugar1  1.091 0.821 90   -0.867    3.048   1.328 0.3834
sugar0 - sugar2  1.727 0.821 90   -0.230    3.685   2.103 0.0950
sugar1 - sugar2  0.636 0.821 90   -1.321    2.594   0.775 0.7195

Confidence level used: 0.95
```

There is a significant difference between one group for the factor 1.

For the factor 1, the contrast between sugar0-2 has a p_value of 0.0187

For the rest there is no strong evidence to reject the null hypothesis.
This can be appreciated in the following graph:



(e) Using significance level $\alpha = 0.05$, perform Tukey post-hoc analysis for factor *chip* by each level of *sugar*. Summarise the findings [3 marks].

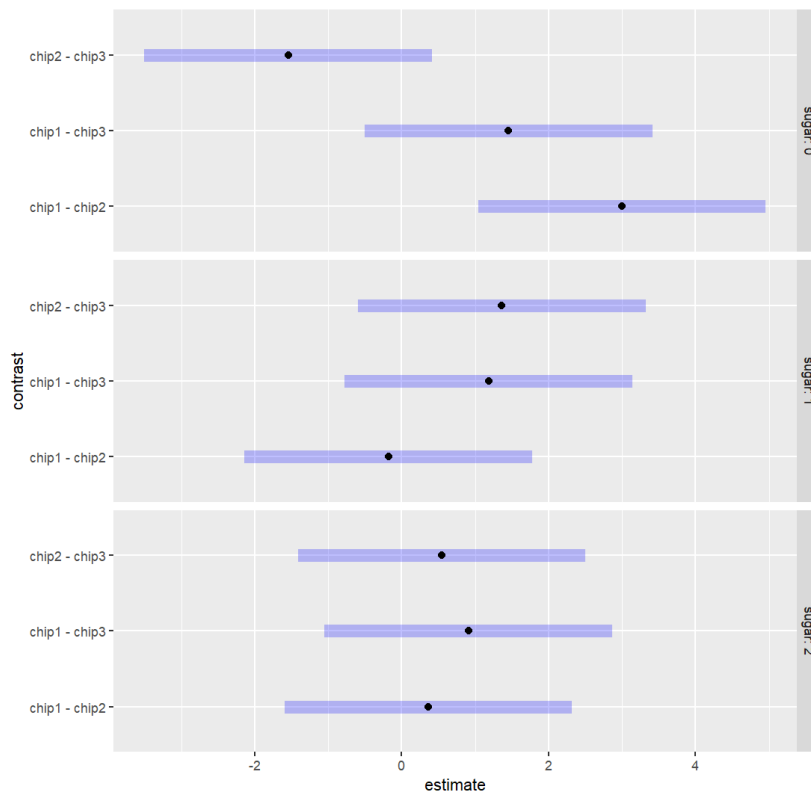
```
sugar = 0:
contrast      estimate      SE df lower.CL upper.CL t.ratio p.value
chip1 - chip2    3.000 0.821 90    1.042    4.958   3.652 0.0013
chip1 - chip3    1.455 0.821 90   -0.503    3.412   1.771 0.1853
chip2 - chip3   -1.545 0.821 90   -3.503    0.412  -1.881 0.1500

sugar = 1:
contrast      estimate      SE df lower.CL upper.CL t.ratio p.value
chip1 - chip2   -0.182 0.821 90   -2.139    1.776  -0.221 0.9734
chip1 - chip3    1.182 0.821 90   -0.776    3.139   1.439 0.3255
chip2 - chip3    1.364 0.821 90   -0.594    3.321   1.660 0.2262

sugar = 2:
contrast      estimate      SE df lower.CL upper.CL t.ratio p.value
chip1 - chip2    0.364 0.821 90   -1.594    2.321   0.443 0.8977
chip1 - chip3    0.909 0.821 90   -1.048    2.867   1.107 0.5124
chip2 - chip3    0.545 0.821 90   -1.412    2.503   0.664 0.7849

Confidence level used: 0.95
```

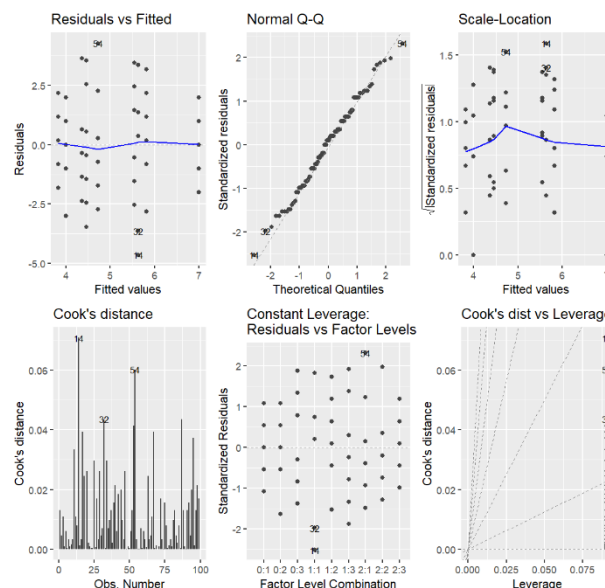
There is a significant difference between one group for the factor 0.
For the factor 0, the contrast between chip1-2 has a p_value of 0.013
For the rest there is no strong evidence to reject the null hypothesis.
This can be appreciated in the following graph:



Assumptions

The conclusions drawn from the F -tests rely on certain assumptions being satisfied.

- (f) Using diagnostic plots, determine if the assumptions of normality, independence and constant variance appear to have been met and from this decide if the conclusions drawn in (b)-(e) are sound [3 marks].



Residual vs Fitted: It doesn't seem to have a pattern and it appears to be randomly spread, indicating constant variance.

Normality: The Q-Q plot suggest a normal distribution since almost all dots lie in the line, however both tails have some points out the line, suggesting that a normality test is necessary

Scale-Location: As the residual vs fitted, there is no pattern, suggesting that there is a constant variance for the error.

Cook's distance plot: Suggest that there may be some observations interacting strongly with the model results.

Constant leverage plot: It does not seem to be observations with strong interaction with model (outliers).

After reviewing the assumptions using the autoplot, the drawing conclusions from b-e sounds good, however it's better to conduct a normality test like the Shapiro test.

- (g) Using significance level $\alpha = 0.05$, test if the residuals are normally distributed. Write down the null and alternative hypotheses, the test statistic and associated p-value, the test decision (with reason) and a conclusion using a minimum of mathematical language [3 marks].

```
Shapiro-Wilk normality test
data:  res
W = 0.99083, p-value = 0.7373
```

P_value: 0.7373

H₀: The distribution of the sample is normal

H_a: The distribution of the sample is not normal

Test decision: We don't have evidence to reject the null hypotheses, since the $p_value > \alpha$

Conclusion: We can't treat the sample distribution as normal.