

TP N°4----ANALYSE DE VARIANCE (ANOVA) A UN FACTEUR

I. Enoncé 1:

Voici les résultats d'une étude qui mesure le score de l'iq (indice d'intelligence) sur un groupe d'étudiant selon leur spécialité. (1=étudiant en physique, 2=étudiant en maths, 3=étudiant en chimie) :

Groupe score_iq

1,00	44,00
1,00	40,00
1,00	44,00
1,00	39,00
1,00	25,00
1,00	37,00
1,00	31,00
1,00	40,00
1,00	22,00
1,00	34,00
1,00	39,00
1,00	20,00
1,00	39,00
1,00	42,00
1,00	41,00
2,00	36,00
2,00	40,00
2,00	37,00
2,00	35,00
2,00	39,00
2,00	40,00
2,00	36,00
2,00	38,00
2,00	24,00
2,00	27,00
2,00	29,00
2,00	24,00
2,00	45,00
2,00	44,00
2,00	44,00
3,00	52,00
3,00	50,00
3,00	51,00
3,00	52,00
3,00	45,00
3,00	49,00
3,00	47,00
3,00	46,00
3,00	47,00
3,00	47,00
3,00	46,00
3,00	45,00
3,00	50,00
3,00	47,00
3,00	49,00

TAF :

- a. Formulez le problème (qu'est-ce qu'on cherche de cette étude ?)
- b. Variable dépendante :
- c. Facteur :
- d. Effectuez une ANOVA pour valider vos hypothèses.

- i. Lecture de données : `score_iq <- read.csv2("D:/COURSES_FSTBM/MASTER_ID_1/AD_R_2021/score_iq.txt", sep="")`
- ii. Préciser le facteur : `data <- data.frame(score=score_iq$score_iq, group=factor(score_iq$Groupe))`
`levels(data$group)`
- iii. Stats descriptives : `data %>% group_by(group) %>% get_summary_stats(score, type = "mean_sd")`
- iv. Visualisation: `ggboxplot(data, x = "group", y = "score")`

```

plot(score ~ group, data=data)
v. Outliers : data %>% group_by(group) %>% identify_outliers(score)
vi. Normalité:
data %>% group_by(group) %>% shapiro_test(score)
ggqqplot(data, "score", facet.by = "group")
vii. Homogénéité des variances:
bartlett.test(data$score ~ data$group)
data %>% levene_test(score ~ group)
viii. ANOVA :
rs.aov <- Anova(lm(score~group, data=data))
anova_summary(rs.aov)
rs.aov <- aov(score ~ group, data )
summary(rs.aov)
#
rs.aov <- data %>% anova_test(score ~ group)
rs.aov
#
rs.aovw <- data %>% welch_anova_test(score ~ group)
rs.aovw
ix. Test de Tukey:
rs.tk <- data %>% tukey_hsd(score ~ group)
rs.tk
rs.tk2 <- rs.tk %>% add_xy_position(x = "group")
ggboxplot(data, x = "group", y = "score") + stat_pvalue_manual(rs.tk2, hide.ns = TRUE) + labs(
  subtitle = get_test_label(rs.aov, detailed = TRUE),caption = get_pwc_label(rs.tk2))
x. Normalité des résidus:
# Construire le modèle linéaire
mdl <- lm(score ~ group, data = data)

# Créer un QQ plot des résidus
ggqqplot(residuals(mdl))

# Calculer le test de normalité de Shapiro-Wilk
shapiro_test(residuals(mdl))
xi. En cas de non normalité
rs.krl <- data %>% kruskal_test(score ~ group)
rs.krl

# taille de l'effet
#Les valeurs d'interprétation couramment utilisées dans la littérature sont : 0,01- < 0,06 (petit
effet), 0,06 - < 0,14 (effet modéré) et >= 0,14 (effet important).
data %>% kruskal_effsize(score ~ group)

#test de Dunn post-hoc
rs.dnt <- data %>% dunn_test(score ~ group, p.adjust.method = "bonferroni")
rs.dnt

rs.dnt2 <- rs.dnt %>% add_xy_position(x = "group")
ggboxplot(data, x = "group", y = "score") + stat_pvalue_manual(rs.dnt2, hide.ns = TRUE) + labs(
  subtitle = get_test_label(rs.krl, detailed = TRUE),caption = get_pwc_label(rs.dnt2))

# ou bien test de Wilcoxon
rs.wxt <- data %>% wilcox_test(score ~ group, p.adjust.method = "bonferroni")
rs.wxt

rs.wxt2 <- rs.wxt %>% add_xy_position(x = "group")
ggboxplot(data, x = "group", y = "score") + stat_pvalue_manual(rs.wxt2, hide.ns = TRUE) + labs(
  subtitle = get_test_label(rs.krl, detailed = TRUE),caption = get_pwc_label(rs.wxt2))

```

xii. En cas de non homogénéité:

```
rs.ght <- data %>% games_howell_test(score ~ group)
rs.ght
rs.ght2 <- rs.ght %>% add_xy_position(x = "group")
ggboxplot(data, x = "group", y = "score") + stat_pvalue_manual(rs.ght2, hide.ns = TRUE) + labs(
  subtitle = get_test_label(rs.aovw, detailed = TRUE), caption = get_pwc_label(rs.ght2))
```

ou two sample t-test

```
rs.tt <- data %>% pairwise_t_test(score ~ group, pool.sd = FALSE, p.adjust.method =
  "bonferroni")
```

```
rs.tt
```

```
rs.tt2 <- rs.ght %>% add_xy_position(x = "group")
ggboxplot(data, x = "group", y = "score") + stat_pvalue_manual(rs.tt2, hide.ns = TRUE) + labs(
  subtitle = get_test_label(rs.aovw, detailed = TRUE), caption = get_pwc_label(rs.tt2))
```

II. Enoncé 2:

Les résultats d'une étude menée sur un groupe d'individus pour mesure le score de réaction par rapport au genre et la dose de drogue administrée, sont fournis dans le tableau : genre (1=homme, 2=femme), dose (1=faible, 2=forte).

TAF : Mêmes questions que (I).

Individu	genre	dose	score
1,00	1,00	1,00	6,00
2,00	1,00	1,00	6,00
3,00	1,00	1,00	3,00
4,00	1,00	1,00	5,00
5,00	1,00	1,00	6,00
6,00	1,00	1,00	4,00
7,00	1,00	1,00	5,00
8,00	1,00	1,00	4,00
9,00	1,00	1,00	4,00
10,00	1,00	1,00	5,00
11,00	1,00	1,00	4,00
12,00	1,00	1,00	3,00
13,00	1,00	2,00	6,00
14,00	1,00	2,00	8,00
15,00	1,00	2,00	7,00
16,00	1,00	2,00	8,00
17,00	1,00	2,00	6,00
18,00	1,00	2,00	8,00
19,00	1,00	2,00	8,00
20,00	1,00	2,00	6,00
21,00	1,00	2,00	7,00
22,00	1,00	2,00	8,00
23,00	1,00	2,00	6,00
24,00	1,00	2,00	7,00
25,00	2,00	1,00	2,00
26,00	2,00	1,00	5,00
27,00	2,00	1,00	2,00
28,00	2,00	1,00	4,00
29,00	2,00	1,00	5,00
30,00	2,00	1,00	7,00
31,00	2,00	1,00	4,00

<u>32,00</u>	<u>2,00</u>	<u>1,00</u>	<u>1,00</u>
<u>33,00</u>	<u>2,00</u>	<u>1,00</u>	<u>2,00</u>
<u>34,00</u>	<u>2,00</u>	<u>1,00</u>	<u>7,00</u>
<u>35,00</u>	<u>2,00</u>	<u>1,00</u>	<u>4,00</u>
<u>36,00</u>	<u>2,00</u>	<u>1,00</u>	<u>,00</u>
<u>37,00</u>	<u>2,00</u>	<u>2,00</u>	<u>2,00</u>
<u>38,00</u>	<u>2,00</u>	<u>2,00</u>	<u>3,00</u>
<u>39,00</u>	<u>2,00</u>	<u>2,00</u>	<u>4,00</u>
<u>40,00</u>	<u>2,00</u>	<u>2,00</u>	<u>,00</u>
<u>41,00</u>	<u>2,00</u>	<u>2,00</u>	<u>,00</u>
<u>42,00</u>	<u>2,00</u>	<u>2,00</u>	<u>1,00</u>
<u>43,00</u>	<u>2,00</u>	<u>2,00</u>	<u>2,00</u>
<u>44,00</u>	<u>2,00</u>	<u>2,00</u>	<u>2,00</u>
<u>45,00</u>	<u>2,00</u>	<u>2,00</u>	<u>4,00</u>
<u>46,00</u>	<u>2,00</u>	<u>2,00</u>	<u>3,00</u>
<u>47,00</u>	<u>2,00</u>	<u>2,00</u>	<u>6,00</u>
<u>48,00</u>	<u>2,00</u>	<u>2,00</u>	<u>3,00</u>