EX 4

roi hezkiyahu

17 3 2022

Q.1

שאלה 1

בהינתן מודל ניתוח שונות עם אפקט מקרי אחד עבור מערך לא מאוזן,

$$Y_{ij} = \mu + a_i + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2) \quad (\epsilon_{ij}\ iid) \quad a_i \sim N(0, \sigma_a^2)$$

עבור $i = 1, ..., I, j = 1, ..., n_i$,

א. חשבו כיצד מתפלג הממוצע הכללי ($\bar{Y}$).

ב. חשבו כיצד מתפלג הממוצע של כל קבוצה ($\bar{Y}_i$).

ג. חשבו למה שווה $E(SSA)$.

ד. חשבו למה שווה $E(SSE)$.

רמז: במערך לא מאוזן $SSA = \sum_{i=1}^{I} n_i(\overline{Y}_{i\cdot} - \overline{Y})^2$ ו- $SSE = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y_i})^2$

a

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{n_i} \mu + \alpha_i + \varepsilon_{ij} = \mu + \frac{1}{N}\sum_{i=1}^{I} n_i\alpha_i + \bar{\varepsilon}$$

$$E(\bar{Y}) = E(\mu + \frac{1}{N}\sum_{i=1}^{I} n_i\alpha_i + \bar{\varepsilon}) = \mu + \frac{1}{N}\sum_{i=1}^{I} n_i E(\alpha_i) + E(\bar{\varepsilon}) = \mu$$

$$V(\bar{Y}) = V(\mu + \frac{1}{N}\sum_{i=1}^{I} n_i\alpha_i + \bar{\varepsilon}) = V(\frac{1}{N}\sum_{i=1}^{I} n_i\alpha_i) + V(\bar{\varepsilon}) = \frac{1}{N^2}\sum_{i=1}^{I} n_i^2 V(\alpha_i) + \frac{\sigma_\varepsilon^2}{N} = \frac{\sum_{i=1}^{I} n_i^2}{N^2}\sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{N}$$

because α_i are iid and independent of all the errors, we get that $\bar{Y}$ distributes normally:

$$\bar{Y} \sim N(\mu, \frac{\sum_{i=1}^{I} n_i^2}{N^2}\sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{N})$$

b

$$\bar{Y_i} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n_i}\sum_{j=1}^{n_i} \mu + \alpha_i + \varepsilon_{ij} = \mu + \alpha_i + \bar{\varepsilon}_i$$

$$E(\bar{Y_i}) = \mu + E(\alpha_i) + E(\bar{\varepsilon}_i) = \mu$$

$$V(\bar{Y_i}) = V(\mu + \alpha_i + \bar{\varepsilon}_i) = V(\alpha_i) + V(\bar{\varepsilon}_i) = \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{n_i}$$

because α_i are iid and independent of all the errors, we get that $\bar{Y_i}$ distributes normally:

$$\bar{Y_i} \sim N(\mu, \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{n_i})$$

$$(1) : SSA = \sum_{i=1}^{I} n_i(\bar{Y}_i - \bar{Y})^2$$

$$(2) : (\bar{Y}_i - \bar{Y})^2 = (\mu + \alpha_i + \bar{\varepsilon}_i - (\mu + \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j + \bar{\varepsilon}))^2 = ((\alpha_i - \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j) + (\bar{\varepsilon}_i - \bar{\varepsilon}))^2 =$$

$$= (\alpha_i - \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j)^2 + (\bar{\varepsilon}_i - \bar{\varepsilon})^2 + 2(\bar{\varepsilon}_i - \bar{\varepsilon})(\alpha_i - \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j)$$

$$(3) : E((\bar{Y}_i - \bar{Y})^2) = E((\alpha_i - \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j)^2 + (\bar{\varepsilon}_i - \bar{\varepsilon})^2 + 2(\bar{\varepsilon}_i - \bar{\varepsilon})(\alpha_i - \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j)) =$$

$$= E((\alpha_i - \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j)^2) + E((\bar{\varepsilon}_i - \bar{\varepsilon})^2) + 2E((\bar{\varepsilon}_i - \bar{\varepsilon})(\alpha_i - \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j)) = E((\alpha_i - \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j)^2) + E((\bar{\varepsilon}_i - \bar{\varepsilon})^2)$$

$$(4) : E((\alpha_i - \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j)^2) = E^2(\alpha_i - \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j) + V(\alpha_i - \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j) = 0 + V(\alpha_i) + V(\frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j) - 2Cov(\alpha_i, \frac{1}{N}\sum_{j=1}^{I} n_j\alpha_j) =$$

$$= \sigma_\alpha^2 + \frac{\sum_{j=1}^{I} n_j^2}{N^2}\sigma_\alpha^2 - 2\frac{n_i}{N}\sigma_\alpha^2 = (1 + \frac{\sum_{j=1}^{I} n_j^2}{N^2} - 2\frac{n_i}{N})\sigma_\alpha^2$$

$$(5) : E((\bar{\varepsilon}_i - \bar{\varepsilon})^2) = V(\bar{\varepsilon}_i - \bar{\varepsilon}) = V(\bar{\varepsilon}_i) + V(\bar{\varepsilon}) - 2cov(\bar{\varepsilon}_i, \bar{\varepsilon}) = (\frac{1}{n_i} - \frac{1}{N})\sigma_\varepsilon^2 = \frac{I-1}{N}\sigma_\varepsilon^2$$

$$(6) : E(SSA) = \sum_{i=1}^{I} n_i E(\bar{Y}_i - \bar{Y})^2 = \sum_{i=1}^{I} n_i((1 + \frac{\sum_{j=1}^{I} n_j^2}{N^2} - 2\frac{n_i}{N})\sigma_\alpha^2 + \frac{I-1}{N}\sigma_\varepsilon^2) = (N + \frac{\sum_{j=1}^{I} n_j^2}{N} - 2\frac{\sum_{i=1}^{I} n_i^2}{N})\sigma_\alpha^2 + (I-1)\sigma_\varepsilon^2 =$$

$$= (N - \frac{\sum_{i=1}^{I} n_i^2}{N})\sigma_\alpha^2 + (I-1)\sigma_\varepsilon^2$$

$$(7) : SSE = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y})^2 = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(\mu + \alpha_i + \varepsilon_{ij} - (\mu + \frac{1}{N}\sum_{k=1}^{I} n_k\alpha_k + \bar{\varepsilon}))^2 = \sum_{i=1}^{I}\sum_{j=1}^{n_i}[(\alpha_i - \frac{1}{N}\sum_{k=1}^{I} n_k\alpha_k) + (\varepsilon_{ij} - \bar{\varepsilon})]^2 =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{n_i}[(\alpha_i - \frac{1}{N}\sum_{k=1}^{I} n_k\alpha_k)^2 + (\varepsilon_{ij} - \bar{\varepsilon})^2 + 2(\alpha_i - \frac{1}{N}\sum_{k=1}^{I} n_k\alpha_k)(\varepsilon_{ij} - \bar{\varepsilon})]$$

$$(8) : E(SSE) = E(\sum_{i=1}^{I}\sum_{j=1}^{n_i}[(\alpha_i - \frac{1}{N}\sum_{k=1}^{I} n_k\alpha_k)^2 + (\varepsilon_{ij} - \bar{\varepsilon})^2 + 2(\alpha_i - \frac{1}{N}\sum_{k=1}^{I} n_k\alpha_k)(\varepsilon_{ij} - \bar{\varepsilon})]) =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{n_i}[E[(\alpha_i - \frac{1}{N}\sum_{k=1}^{I} n_k\alpha_k)^2] + E[(\varepsilon_{ij} - \bar{\varepsilon})^2] + 2E(\alpha_i - \frac{1}{N}\sum_{k=1}^{I} n_k\alpha_k)(\varepsilon_{ij} - \bar{\varepsilon})] =$$

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}[V(\alpha_i - \frac{1}{N}\sum_{k=1}^{I} n_k\alpha_k) + V(\varepsilon_{ij} - \bar{\varepsilon})$$

$$(9) : V(\varepsilon_{ij} - \bar{\varepsilon}) = V(\varepsilon_{ij}) + V(\bar{\varepsilon}) - 2cov(\varepsilon_{ij}, \bar{\varepsilon}) = (1 + \frac{1}{N} - \frac{2}{N})\sigma_\varepsilon^2 = \frac{N-I}{N}\sigma_\varepsilon^2$$

$$(10) : V(\alpha_i - \frac{1}{N}\sum_{k=1}^{I} n_k\alpha_k) = \sigma_\alpha^2 + \frac{\sum_{k=1}^{I} n_k^2}{N^2}\sigma_\alpha^2 - 2\frac{n_i}{N}\sigma_\alpha^2$$

$$(11) : E(SSE) = \sum_{i=1}^{I}\sum_{j=1}^{n_i}[\frac{N-I}{N}\sigma_\varepsilon^2 + \sigma_\alpha^2 + \frac{\sum_{k=1}^{I} n_k^2}{N^2}\sigma_\alpha^2 - 2\frac{n_i}{N}\sigma_\alpha^2] = (N-I)\sigma_\varepsilon^2 + N\sigma_\alpha^2 - \frac{\sum_{k=1}^{I} n_k^2}{N^2}\sigma_\alpha^2$$

Q.2

שאלה 2

כיתבו פונקציה המחשבת רווח סמך עבור סטיית התקן של האפקט המקרי (σ_a) בשיטת Satterthwaite's procedure.
הניחו מודל ניתוח שונות עם אפקט מקרי אחד (וללא אפקטים נוספים), ומדגם מאוזן.
הפונקציה צריכה לקבל חמישה ארגומנטים:

- n (מספר התצפיות בכל קבוצה)
- I (מספר הקבוצות)
- MSA
- MSE
- α

ולהחזיר רווח סמך ברמת ביטחון $(1-\alpha)100\%$ עבור סטיית התקן של האפקט המקרי (σ_a).

```
s_p <- function(n,I,MSA,MSE,alpha){
  N <- n*I
  c_s <- c(1/n,-1/n)
  Ms <- c(MSA,MSE)
  df <- c(I-1,N-I)
  sigma_hat <- c_s %*% Ms
  df_sigma = (sigma_hat)^2/ sum(c_s^2*Ms^2/df)
  u_q <- qchisq(1-alpha/2,df_sigma)
  l_q <- qchisq(alpha/2,df_sigma)
  ci <- c(df_sigma*sigma_hat/u_q,df_sigma*sigma_hat/l_q)
  return(ci)
}
```

Q.3

שאלה 3

בשאלה זו נשתמש בנתונים בשם feedback_df_bi הזמינים כקובץ csv במודל.

אלו נתונים מאותו מחקר בו השתמשתם בשיעור, אך הפעם המשתנה התלוי performance מכיל מדד אחר ושמו emg_bi. המדד הזה מכיל תוצאות מדידה של פוטנציאל חשמלי הנמדד על ידי מכשיר אלקטרומיוגרף (EMG) בשריר הזרוע הדו ראשי (biceps brachii muscles). הנבדקים בקבוצת הביקורת לא קיבלו עידוד לפני מדידת הפוטנציאל החשמלי שלהם, ואילו אלו שבקבוצות הטיפול קיבלו חיזוק חיובי ("You are looking great.") או שלילי ("You are not trying.") . בהמשך יעניין אותנו לבחון, למשל, האם יש השפעה לחיזוק חיובי או שלילי על הפוטנציאל החשמלי הנמדד בשריר הזרוע הדו ראשי.

העמודות הנדרשות לנו בשאלה זו הן -

- id - מספר מזהה ייחודי למשתתף
- performance - (נומרי) emg bi הציון המתקבל במדד
- feedback- (פידבק חיובי / פידבק שלילי / ביקורת) האם קיבל פידבק

בינתיים נרצה להשתמש בקבוצת הביקורת בלבד. תוכלו ליצור את תת המדגם הזה למשל כך,

```
feedback_bi_control = feedback_bi[(feedback_bi$feedback=="no feedback"),]
```

א. נרצה לדעת האם בתוך קבוצת הביקורת, השונות של התיפקוד (performance) בין משתתפים (id) שונה מאפס. בדקו זאת בעזרת ניתוח שונות עם אפקט מקרי אחד.

I. מדוע נעדיף ניתוח שונות עם אפקט מקרי על פני אפקט קבוע?

II. דווחו את התוצאה שקיבלתם. מה המסקנה עבור $\alpha = 0.05$?

III. חשבו אומדים לכל אחד משלושת הפרמטרים של המודל.

ב. בנו רווח סמך ברמת סמך של 95% עבור התוחלת (μ) בעזרת הנוסחאות שהוצגו בכיתה.

ג. בנו רווח סמך ברמת סמך של 95% עבור שונות השגיאות (σ_ϵ^2) בעזרת הנוסחאות שהוצגו בכיתה.

ד. בנו רווח סמך ברמת סמך של 95% עבור שונות האפקט המקרי (σ_a^2) בעזרת Satterthwaite's procedure (תוכלו להיעזר לשם כך בפונקציה שכתבתם בשאלה 2)

ה. בנו רווח סמך ברמת סמך של 95% עבור ICC ($\rho = \frac{\sigma_a^2}{\sigma_a^2+\sigma_\epsilon^2}$) בעזרת הנוסחאות שהוצגו בכיתה.

ו. הפונקציה confint מקבלת מודל סטטיסטי ומחשבת רווחי סמך לפרמטרים של המודל. הפעילו את הפונקציה על מודל ניתוח השונות עם האפקט המקרי, ופרטו מה התוצאות המתקבלות. האם רווחי הסמך שקיבלתם מ confint זהים לרווחי הסמך שחישבתם באופן ידני?

```
#imports
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------------------------------- tidyverse 1.3.0 --
```

```
## <U+221A> ggplot2 3.3.2     <U+221A> purrr   0.3.4
## <U+221A> tibble  3.0.3     <U+221A> dplyr   1.0.2
## <U+221A> tidyr   1.1.2     <U+221A> stringr 1.4.0
## <U+221A> readr   1.3.1     <U+221A> forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(glue)
```

```
##
## Attaching package: 'glue'
```

```
## The following object is masked from 'package:dplyr':
##
##     collapse
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

a

```
feed_back <- read.csv("feedback_df_bi.csv")
feed_back_cont <- feed_back %>%
  select(id,performance,feedback) %>%
  filter(feedback == "no feedback")
rand_a <- lmer(performance ~ (1|id) , data = feed_back_cont)
summary(rand_a)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: performance ~ (1 | id)
##    Data: feed_back_cont
##
## REML criterion at convergence: 1913.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3784 -0.3585 -0.0520  0.2818  6.4500
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 68.69    8.288
##  Residual             67.29    8.203
## Number of obs: 264, groups:  id, 22
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   89.503      1.838    48.7
```

```
n <-  12
I <- 22
N <- n*I
sigma_a <- 68.69
sigma_e <- 67.29
mu <- 89.503
MSE <-  sigma_e
MSA <- n*sigma_a+MSE
f_stat <- MSA/MSE
glue("pvalue is: {pf(f_stat,I-1,N-I,lower.tail = FALSE)} thus we will reject the null with a confidance level of 95%")
```

```
## pvalue is: 1.36577270543185e-29 thus we will reject the null with a confidance level of 95%
```

```
glue("the estimators are:
     mu = {mu}
     sigma alpha = {sigma_a}
     sigma epsilon = {sigma_e}")
```

```
## the estimators are:
## mu = 89.503
## sigma alpha = 68.69
## sigma epsilon = 67.29
```

נעדיף ניתוח שונות עם אפקט מקרי על פני קבוע מכיוון ויש שונות בין אנשים שונים ללא קשר לקבוצת איכות/ביקורת בה הם נמצאים

b

```
ci_mu <- round(mu +c(-1,1)*qt(0.975,I-1)*sqrt(MSA/N),3)
glue("ci at a confidance level of 95% for mu is:({ci_mu[1]},{ci_mu[2]})")
```

```
## ci at a confidance level of 95% for mu is:(85.681,93.325)
```

c

```
alpha <- 0.05
ci_sigma_e <- round(MSE*(N-I)/c(qchisq(1-alpha/2,N-I),qchisq(alpha/2,N-I)),3)
glue("ci at a confidance level of 95% for sigma epsilon is:({ci_sigma_e[1]},{ci_sigma_e[2]})")
```

```
## ci at a confidance level of 95% for sigma epsilon is:(56.743,81.094)
```

d

```
ci_sigma_a = round(s_p(n,I,MSA,MSE,alpha),3)
glue("ci at a confidance level of 95% for sigma alpha is:({ci_sigma_a[1]},{ci_sigma_a[2]})")
```

```
## ci at a confidance level of 95% for sigma alpha is:(39.188,150.452)
```

e

```
L_U <- ((MSA/MSE)/c(qf(1-alpha/2,I-1,N-I),qf(alpha/2,I-1,N-I)) -1)/n
L <- L_U[1]
U <- L_U[2]
ci_ICC <- round(c(L/(L+1),U/(U+1)),3)
glue("ci at a confidance level of 95% for ICC is:({ci_ICC[1]},{ci_ICC[2]})")
```

```
## ci at a confidance level of 95% for ICC is:(0.354,0.689)
```

f

```
confint(rand_a)
```

```
## Computing profile confidence intervals ...
```

```
##                 2.5 %    97.5 %
## .sig01       5.981416 11.444573
## .sigma       7.522899  8.991571
## (Intercept) 85.824293 93.181073
```

הרווח סמך לתוחלת זההה, אך הרווחי סמך לסיגמאות שונות וזאת מיכוון והרווח סמך כאן הינו לסטיית תקן ואילו אנו חישבנו רווח סמך לשונות

נשים לב שאם ניקח את הקצוות של הרווחי סמך של פונקציה זו ונעלה אותם בריבוע נקבל ביטוי קרוב לביטוי שלנו ועל כן רווחי הסמך שקולים.