**Repeated Measures Practical**

Please complete all tasks this exercise.

|  |
| --- |
| In this exercise we will use the following R functions, for which you should briefly read the documentation:  purrr::map\_df()  dplyr::group\_split()  lme4::glmer()  jtools::summ()  You may need to install the following packages:  install.packages("jtools")  install.packages("lme4")  install.packages("skimr")  install.packages("tidyverse")  library(jtools)  library(lme4)  library(skimr)  library(tidyverse) |

1. We are going to analyse the Skin data – as described briefly in the lecture notes. The file is called skin.csv and consists of 4 columns.

id - ID number

resp - response coded 0=(excellent/good), 1=(fair/poor)

treat – treatment coded 0=Treatment, 1=Placebo

time - in weeks

The aim of the analysis is to identify whether treatment improves the skin condition.

1. Familiarise yourself with the dataset using commands such as skimr::skim() and dplyr::glimpse(), and check if there are any missing data.

*# Define file path to practical directory (EDIT THIS)*

dir <- "/Users/robertfletcher/Documents/phd/projects"

*# Define practical directory (DO NOT EDIT THIS)*

prac <- "advanced\_biostats/09\_hierarchical\_data"

*# Read data*

skin <- readr::read\_csv(glue::glue("{dir}/{prac}/data/skin.csv"))

*# Inspect data*

print(skin, n = 10)

skimr::skim(skin)

dplyr::glimpse(skin)

*# Inspect missing data*

skin |>

purrr::map\_df(\(x) sum(is.na(x)))

> skimr::skim(skin)

── Data Summary ────────────────────────

Values

Name skin

Number of rows 288

Number of columns 4

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Column type frequency:

factor 3

numeric 1

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Group variables None

── Variable type: factor ───────────────────────────────────────────────────────────────

skim\_variable n\_missing complete\_rate ordered n\_unique top\_counts

1 resp 0 1 FALSE 2 Exc: 223, Fai: 65

2 treat 0 1 FALSE 2 Pla: 144, Tre: 144

3 time 0 1 FALSE 4 3: 72, 6: 72, 9: 72, 12: 72

── Variable type: numeric ──────────────────────────────────────────────────────────────

skim\_variable n\_missing complete\_rate mean sd p0 p25 p50 p75 p100 hist

1 id 0 1 36.5 20.8 1 18.8 36.5 54.2 72 ▇▇▇▇▇

> dplyr::glimpse(skin)

Rows: 288

Columns: 4

$ id <dbl> 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10,…

$ resp <fct> Fair/poor, Excellent/good, Excellent/good, Excellent/good, Excellent/good, Excellent/good, Excellent/good, Excellent/go…

$ treat <fct> Placebo, Placebo, Placebo, Placebo, Placebo, Placebo, Placebo, Placebo, Placebo, Placebo, Placebo, Placebo, Placebo, Pl…

$ time <fct> 3, 6, 9, 12, 3, 6, 9, 12, 3, 6, 9, 12, 3, 6, 9, 12, 3, 6, 9, 12, 3, 6, 9, 12, 3, 6, 9, 12, 3, 6, 9, 12, 3, 6, 9, 12, 3,…

> skin |>

+ purrr::map\_df(\(x) sum(is.na(x)))

# A tibble: 1 × 4

id resp treat time

<int> <int> <int> <int>

1 0 0 0 0

1. Tabulate the responses by time and treatment using this pre-written function (just copy and paste it into your Rscript):

*# Define function to tabulate*

tabulate <- function(.data, .var1, .var2) {

tab <- .data |>

dplyr::group\_by({{ .var1 }}, {{ .var2 }}) |>

dplyr::summarise(n = n(), .groups = "drop") |>

tidyr::pivot\_wider(

id\_cols = {{ .var1 }}, names\_from = {{ .var2 }}, values\_from = n

) |>

dplyr::rowwise() |>

dplyr::mutate(total = sum(dplyr::c\_across(`3`:`12`))) |>

dplyr::ungroup() |>

dplyr::add\_row(resp = "total") |>

dplyr::mutate(

dplyr::across(

tidyselect:::where(is.numeric),

~ dplyr::case\_when(resp == "total" ~ sum(., na.rm = TRUE), TRUE ~ .)

),

dplyr::across(

tidyselect:::where(is.numeric),

~ paste0(., " (", round(. / (sum(.) / 2) \* 100, digits = 1), "%)")

)

)

var1 <- deparse(substitute(.var1))

var2 <- deparse(substitute(.var2))

chi <- chisq.test(table(.data[[var1]], .data[[var2]]))

res <- list(tab, chi)

return(res)

}

*# Tabulate for the entire data*

skin |>

tabulate(resp, time)

[[1]]

# A tibble: 3 × 6

resp `3` `6` `9` `12` total

<chr> <chr> <chr> <chr> <chr> <chr>

1 Excellent/good 39 (54.2%) 59 (81.9%) 55 (76.4%) 70 (97.2%) 223 (77.4%)

2 Fair/poor 33 (45.8%) 13 (18.1%) 17 (23.6%) 2 (2.8%) 65 (22.6%)

3 total 72 (100%) 72 (100%) 72 (100%) 72 (100%) 288 (100%)

[[2]]

Pearson's Chi-squared test

data: table(.data[[var1]], .data[[var2]])

X-squared = 39.321, df = 3, p-value = 1.484e-08

*# Split by treatment*

skin |>

dplyr::group\_split(treat) |>

purrr::map(\(x) tabulate(x, resp, time))

[[1]]

[[1]][[1]]

# A tibble: 3 × 6

resp `3` `6` `9` `12` total

<chr> <chr> <chr> <chr> <chr> <chr>

1 Excellent/good 26 (72.2%) 32 (88.9%) 32 (88.9%) 35 (97.2%) 125 (86.8%)

2 Fair/poor 10 (27.8%) 4 (11.1%) 4 (11.1%) 1 (2.8%) 19 (13.2%)

3 total 36 (100%) 36 (100%) 36 (100%) 36 (100%) 144 (100%)

[[1]][[2]]

Pearson's Chi-squared test

data: table(.data[[var1]], .data[[var2]])

X-squared = 10.368, df = 3, p-value = 0.01568

[[2]]

[[2]][[1]]

# A tibble: 3 × 6

resp `3` `6` `9` `12` total

<chr> <chr> <chr> <chr> <chr> <chr>

1 Excellent/good 13 (36.1%) 27 (75%) 23 (63.9%) 35 (97.2%) 98 (68.1%)

2 Fair/poor 23 (63.9%) 9 (25%) 13 (36.1%) 1 (2.8%) 46 (31.9%)

3 total 36 (100%) 36 (100%) 36 (100%) 36 (100%) 144 (100%)

[[2]][[2]]

Pearson's Chi-squared test

data: table(.data[[var1]], .data[[var2]])

X-squared = 32.071, df = 3, p-value = 5.056e-07

What differences do you observe between treatment arms and over time?

Fewer positive responses over time and in the treat=0 group.

Response improves over time.

1. Fit a logistic model to the responses, with fixed terms for treatment and time, and a random intercept term to account for the correlation within individuals using the command:

*# Fit logistic model to responses, with fixed terms for treatment and time,*

*# and random intercept term to account for the correlation within individuals*

fit <-

lme4::glmer(

resp ~ treat + time + (1 | id), data = skin, family = "binomial"

)

jtools::summ(fit, confint = TRUE, digits = 3, exp = TRUE)

Interpret the estimated parameters.

> summ(fit, confint = TRUE, digits = 3, exp = TRUE)

MODEL INFO:

Observations: 288

Dependent Variable: resp

Type: Mixed effects generalized linear regression

Error Distribution: binomial

Link function: logit

MODEL FIT:

AIC = 255.199, BIC = 269.851

Pseudo-R² (fixed effects) = 0.296

Pseudo-R² (total) = 0.535

FIXED EFFECTS:

-------------------------------------------------------

exp(Est.) 2.5% 97.5% z val.

----------------- ----------- ------- -------- --------

(Intercept) 0.873 0.332 2.294 -0.275

treat1 5.056 1.855 13.783 3.168

time 0.700 0.614 0.799 -5.310

-------------------------------------------------------

RANDOM EFFECTS:

---------------------------------

Group Parameter Std. Dev.

------- ------------- -----------

id (Intercept) 1.301

---------------------------------

Grouping variables:

--------------------------

Group # groups ICC

------- ---------- -------

id 72 0.340

--------------------------

Important to interpret parameter estimates correctly. Interpret as odds-ratios

For TREAT: OR= 5.06 (95% CI=1.86,13.78). This can be interpreted as “given an individuals underlying odds of a poor/fair response, being in the placebo group increases these odds by a factor of 5.”

For TIME: OR=0.70 (0.61,0.80). This can be interpreted as

“given as individuals underlying odds of a poor/fair response, a change of a day led to a 30% decrease in the odds of a poor/fair outcome”

We have assumed that the effect of time is linear and is also the same for both the treatment and placebo arms (on the logit scale). We should investigate to see if there is evidence of non-linearity and an interaction.

*# Explore non-linear association with time, and interaction between time and*

*# treatment*

skin <- skin |>

dplyr::mutate(time = as.factor(time))

fit\_nl <-

lme4::glmer(

resp ~ treat \* time + (1 | id), data = skin, family = "binomial"

) |>

jtools::summ(confint = TRUE, digits = 3, exp = TRUE)

*# Add interaction term*

fit\_int <-

lme4::glmer(

resp ~ treat \* time + (1 | id), data = skin, family = "binomial"

) |>

jtools::summ(confint = TRUE, digits = 3, exp = TRUE)

What do you conclude?

MODEL INFO:

Observations: 288

Dependent Variable: resp

Type: Mixed effects generalized linear regression

Error Distribution: binomial

Link function: logit

MODEL FIT:

AIC = 246.267, BIC = 268.245

Pseudo-R² (fixed effects) = 0.386

Pseudo-R² (total) = 0.634

FIXED EFFECTS:

-------------------------------------------------------

exp(Est.) 2.5% 97.5% z val.

----------------- ----------- ------- -------- --------

(Intercept) 0.317 0.127 0.791 -2.462

treat1 5.879 1.935 17.859 3.125

time6 0.133 0.047 0.370 -3.859

time9 0.216 0.084 0.554 -3.186

time12 0.011 0.002 0.064 -4.939

-------------------------------------------------------

RANDOM EFFECTS:

---------------------------------

Group Parameter Std. Dev.

------- ------------- -----------

id (Intercept) 1.491

---------------------------------

Grouping variables:

--------------------------

Group # groups ICC

------- ---------- -------

id 72 0.403

We **cannot** formally test for the non-linear relationship using a likelihood ratio test. However, the estimates suggest it is unlikely to be a linear relationship. Note also that the association between treat and outcome has increased to an OR=5.9.

> summ(fit\_int, confint = TRUE, digits = 3, exp = TRUE)

MODEL INFO:

Observations: 288

Dependent Variable: resp

Type: Mixed effects generalized linear regression

Error Distribution: binomial

Link function: logit

MODEL FIT:

AIC = 249.600, BIC = 282.567

Pseudo-R² (fixed effects) = 0.346

Pseudo-R² (total) = 0.631

FIXED EFFECTS:

---------------------------------------------------------

exp(Est.) 2.5% 97.5% z val.

------------------- ----------- ------- -------- --------

(Intercept) 0.242 0.081 0.719 -2.552

treat1 9.994 2.126 46.984 2.915

time6 0.234 0.055 1.003 -1.956

time9 0.234 0.055 1.004 -1.955

time12 0.044 0.004 0.435 -2.671

treat1:time6 0.335 0.047 2.384 -1.092

treat1:time9 0.734 0.111 4.843 -0.321

treat1:time12 0.089 0.003 2.373 -1.444

---------------------------------------------------------

RANDOM EFFECTS:

---------------------------------

Group Parameter Std. Dev.

------- ------------- -----------

id (Intercept) 1.596

---------------------------------

Grouping variables:

--------------------------

Group # groups ICC

------- ---------- -------

id 72 0.436

----------------------------------------------------

We do not have enough evidence to suggest an interaction between TREAT and TIME. It is difficult (not impossible) to perform a multivariate wald test after glmer because the covariance matrix needs a bit of manipulation (beyond the scope of the teaching today).