**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 (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)))

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)

*# Split by treatment*

skin |>

dplyr::group\_split(treat) |>

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

What differences do you observe between treatment arms and 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.

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?