

# Day6 exercise solutions

Ali Movasati

Oct. 21st, 2024

```
# Set global code chunk options
knitr::opts_chunk$set(warning = FALSE)

# load required libraries
library(skimr)
library(ggplot2)
library(ggpubr)
library(magrittr)
library(dplyr)
library(tibble)
library(lme4)

# define functions
`%notin%` <- Negate(`%in%`)
```

## Problem 1

```
# read in the data

hearing <- read.table(file = "/Users/alimos313/Documents/studies/phd/university/courses/stat-modelling/
```

1.A)

```
# print descriptive
skim(hearing)
```

Table 1: Data summary

Name	hearing
Number of rows	96
Number of columns	3
Column type frequency:	
character	1
numeric	2
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ListID	0	1	5	5	0	4	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
SubjectID	0	1	12.50	6.96	1	6.75	12.5	18.25	24	
Hearing	0	1	28.31	8.37	14	20.00	30.0	34.00	48	

```
table(hearing$ListID)
```

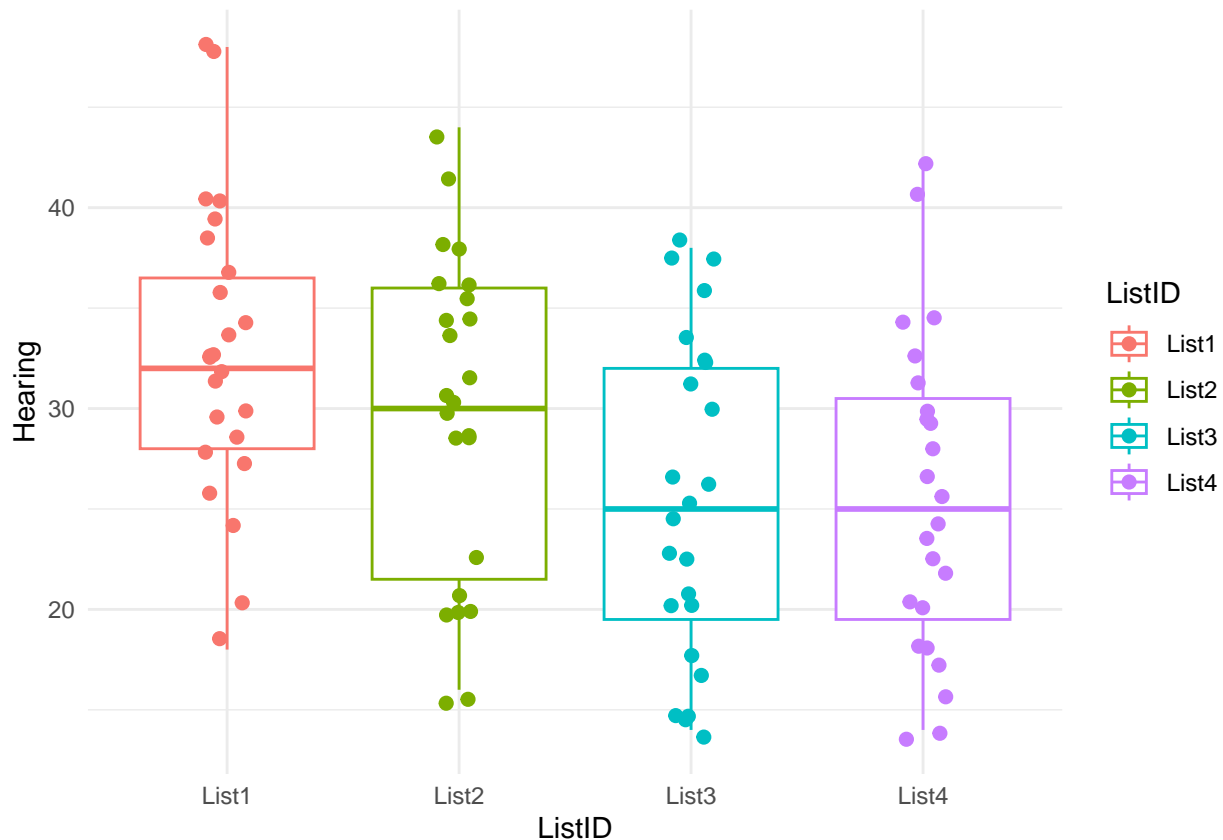
```
##
## List1 List2 List3 List4
##    24    24    24    24
```

```
# prepare data
```

```
hearing %<>% mutate(ListID = as.factor(ListID), SubjectID = as.factor(SubjectID))
```

```
# plot a box-plot for visualization
```

```
hearing %>% ggplot(aes(x = ListID, y = Hearing, color = ListID)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, aes(color = factor(ListID)), size = 2) +
  theme_minimal()
```



## 1.B)

```
lm_simple <- lm(Hearing ~ 1 + ListID, data = hearing)

sum_model_simple <- summary(lm_simple)

print(sum_model_simple)

##
## Call:
## lm(formula = Hearing ~ 1 + ListID, data = hearing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.7500  -5.5833  -0.2083   6.3333  16.4167
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    32.750      1.612   20.315 < 2e-16 ***
## ListIDList2     -3.083      2.280   -1.352  0.17955
## ListIDList3     -7.500      2.280   -3.290  0.00142 **
## ListIDList4     -7.167      2.280   -3.144  0.00225 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.898 on 92 degrees of freedom
## Multiple R-squared:  0.1382, Adjusted R-squared:  0.1101
## F-statistic: 4.919 on 3 and 92 DF,  p-value: 0.00325
```

« comments »

Only 11.01% of variability in hearing measures are explained by different lists

We have enough evidence to state that the mean hearing score for List 3 and 4 are different than list 1, while for list 2 we cannot state that!

## 1.C)

```
# fit the mixed model
lm_mixed <- lmer(Hearing ~ 1 + ListID + (1|SubjectID), data = hearing)

sum_model_mixed <- summary(lm_mixed)

print(sum_model_mixed)

## Linear mixed model fit by REML ['lmerMod']
## Formula: Hearing ~ 1 + ListID + (1 | SubjectID)
## Data: hearing
##
## REML criterion at convergence: 635.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.86533 -0.56158 -0.01092  0.63222  2.69167
##
```

```

## Random effects:
##   Groups      Name      Variance Std.Dev.
##   SubjectID (Intercept) 26.04    5.103
##   Residual              36.33    6.027
## Number of obs: 96, groups: SubjectID, 24
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   32.750      1.612  20.315
## ListIDList2   -3.083      1.740  -1.772
## ListIDList3   -7.500      1.740  -4.311
## ListIDList4   -7.167      1.740  -4.119
##
## Correlation of Fixed Effects:
##              (Intr) LsIDL2 LsIDL3
## ListIDList2 -0.540
## ListIDList3 -0.540  0.500
## ListIDList4 -0.540  0.500  0.500

# fit the model without ListID (null model)
lm_mixed_null <- lmer(Hearing ~ (1 | SubjectID), data = hearing)

sum_model_mixed_null <- summary(lm_mixed_null)

# Likelihood ratio test
anova(lm_mixed_null, lm_mixed)

## refitting model(s) with ML (instead of REML)

## Data: hearing
## Models:
## lm_mixed_null: Hearing ~ (1 | SubjectID)
## lm_mixed: Hearing ~ 1 + ListID + (1 | SubjectID)
##              npar      AIC      BIC  logLik deviance Chisq Df Pr(>Chisq)
## lm_mixed_null    3 674.22 681.91 -334.11   668.22
## lm_mixed         6 657.70 673.09 -322.85   645.70 22.52  3 5.083e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

1.D)

## Problem 2

```

# read in the data

termite <- read.table(file = "/Users/alimos313/Documents/studies/phd/university/courses/stat-modelling

```

2.A)

2.B)

2.C)

2.D)

2.E)

2.F)