Day5 exercise solutions

Ali Movasati

Oct. 14th, 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)


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

Problem 1

```
# read in the data

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

1.A)

```
salary %<>% mutate(size = factor(districtSize)) %>%
            select(-districtSize)
```

1.B)

- Numerical summary

```
# summary of dataset
skim(salary)
```

Table 1: Data summary

Name	salary
Number of rows	325
Number of columns	4
Column type frequency:	
character	1

factor	1
numeric	2
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
District	0	1	3	24	0	325	0

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
size	0	1	FALSE	3	1: 223, 2: 68, 3: 34

Variable type: numeric

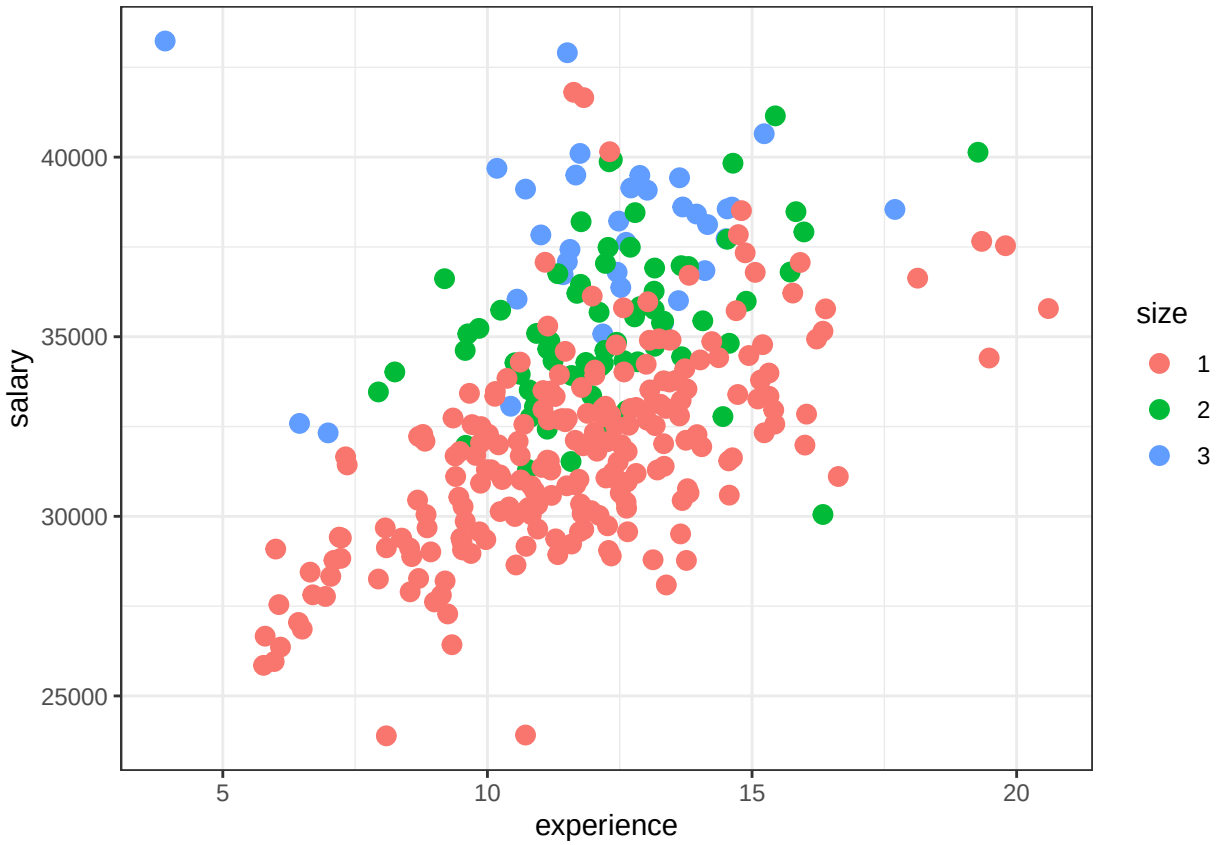
skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
salary	0	1	33168.33	3412.77	23889.50	30847.70	32867.50	35296.70	43232.6	
experience	0	1	11.86	2.55	3.91	10.44	11.97	13.33	20.6	

The average salary in USD is 3.3168327×10^4 and in CHF is 2.8856445×10^4!

- Graphical summary

```
# visualize the dataset


## scatter plot
salary %>%
    ggplot(aes(x = experience, y = salary, color = size)) +
    geom_point(size = 3) +
    theme_bw()
```

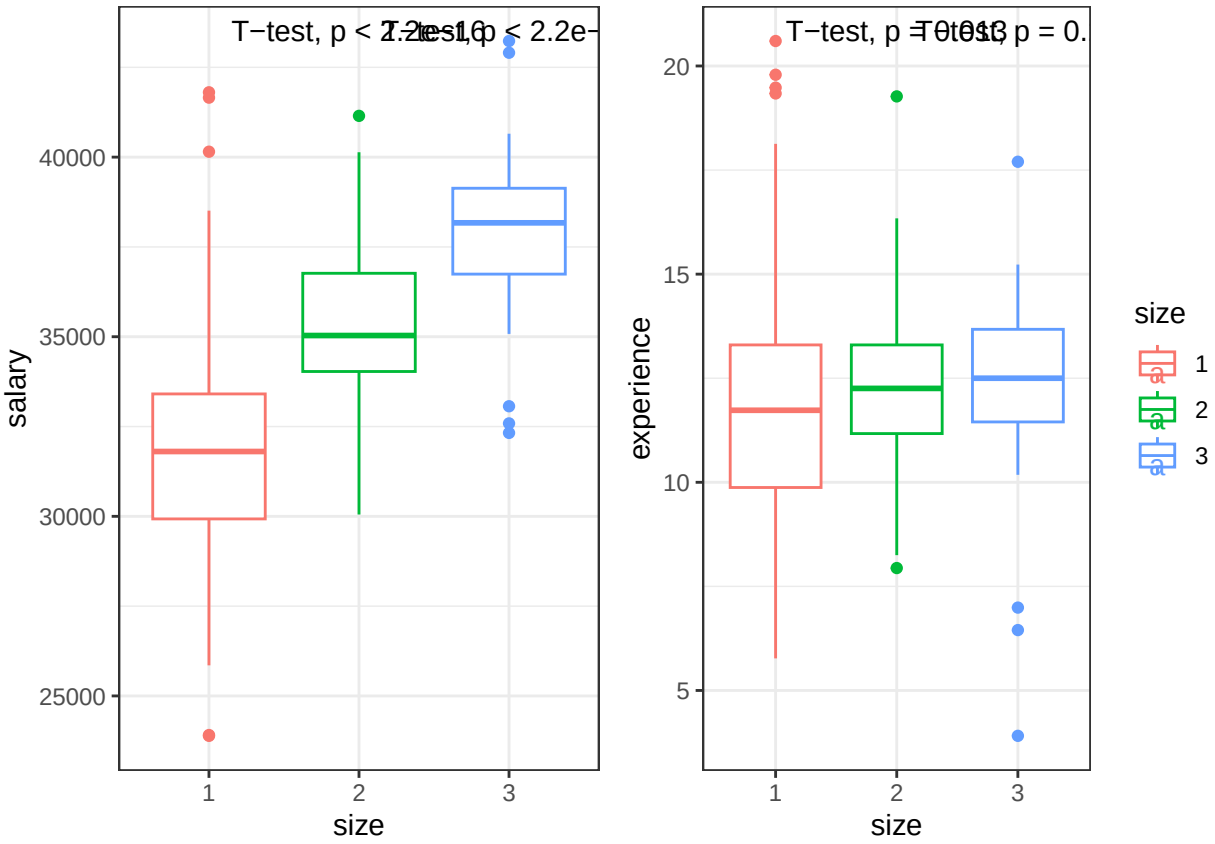



```
## boxplots


bxp1 <- salary %>%
    ggplot(aes(x = size, y = salary, color = size)) +
    geom_boxplot() +
    stat_compare_means(ref.group = "1", method = "t.test") +
    theme_bw() +
    theme(legend.position = "none")

bxp2 <- salary %>%
    ggplot(aes(x = size, y = experience, color = size)) +
    geom_boxplot() +
    stat_compare_means(ref.group = "1", method = "t.test") +
    theme_bw()

# Arrange the plots side by side
grid.arrange(bxp1, bxp2, ncol = 2)
```



1.C)

1.D)

1.E)

1.F)

1.G)

Problem 2

2.A)

2.B)

2.C)

2.D)

2.E)