Chapter 3 Workshop

Table of contents

Dataset `Prestige`

We will be using a well-known dataset called `Prestige` from the `car` R package. This dataset deals with prestige ratings of Canadian Occupations. The `Prestige` dataset has 102 rows and 6 columns. The observations are occupations.

This data frame contains the following columns:

- **education** - Average education of occupational incumbents, years, in 1971.

- **income** - Average income of incumbents, dollars, in 1971.

- **women** - Percentage of incumbents who are women.

- **prestige** - Pineo-Porter prestige score for occupation, from a social survey conducted in the mid-1960s.

- **census** - Canadian Census occupational code.

- **type** - Type of occupation. A factor with levels: bc, Blue Collar; prof, Professional, Managerial, and Technical; wc, White Collar. (includes four missing values).

Exercise 3.1

For a standard normal variable z , obtain the area between -1.8 and 2.1.

```
pnorm(2.1, mean=0, sd=1) - pnorm(-1.8, mean=0, sd=1)
```

Note that the mean=0, sd=1 are the defaults for pnorm function, so don't need to be specified.

```
pnorm(2.1) - pnorm(-1.8)
```

Exercise 3.2

Plot the `prestige` scores data as a histogram and show the theoretical normal curve fitted to the data.

```
library(tidyverse)
library(car)

Prestige |>
  ggplot() +
  aes(prestige) +
  geom_histogram(aes(y=after_stat(density)), bins=10) +
  stat_function(fun = dnorm,
                args=list(mean=mean(Prestige$prestige),
                          sd=sd(Prestige$prestige)),
                geom = "line")
```

or

```
library(car)

hist(Prestige$prestige, probability=T)

curve(dnorm(x,
            mean(Prestige$prestige),
            sd(Prestige$prestige)),
      add= T, lty=2)
```

Exercise 3.3

Obtain the normal quantile plot and test for the normality for `prestige` scores data.

```r
Prestige |>
  ggplot() +
  aes(sample=prestige) +
  stat_qq() +
  stat_qq_line()

# or
qqnorm(Prestige$prestige)
qqline(Prestige$prestige)

shapiro.test(Prestige$prestige)

ks.test(Prestige$prestige,
        "pnorm",
        mean(Prestige$prestige),
        sd(Prestige$prestige) )
```

Exercise 3.4

Examine the fit of non-normal distributions for `prestige` scores data.

```r
library(fitdistrplus)

m1 <- fitdist(Prestige$prestige, "lnorm")

plot(m1)
```

```r
library(fitdistrplus)

m2 <- fitdist(Prestige$prestige, "gamma")

plot(m2)
```

```r
library(fitdistrplus)

m3 <- fitdist(Prestige$prestige, "weibull")

plot(m3)
```

```r
descdist(Prestige$prestige)
```

More graphing examples are here (R codes file).