Quiz 3

Anna Miletova, 89231151

2025-05-22

```r
library(pwr)

my.data <- read.table("data.txt", header = TRUE, sep = "\t")

x <- 1
z <- 5

my.data <- my.data[unique(c(seq(x,nrow(my.data),by=10),seq(z,nrow(my.data),by=10))),]
```

Task 1. Check shoe size difference for males and females

```r
t.test(Shoe.size ~ Sex, data = my.data, var.equal = TRUE, conf.level = 0.99)
```

```
##
##  Two Sample t-test
##
## data:  Shoe.size by Sex
## t = -16.7, df = 103, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group female and group male is not equal to
## 99 percent confidence interval:
##  -5.883141 -4.285189
## sample estimates:
## mean in group female   mean in group male
##             38.74342             43.82759
```

```r
pwr.t.test(d = 1.5 / 2, n = 25, sig.level = 0.01, type = "two.sample")
```

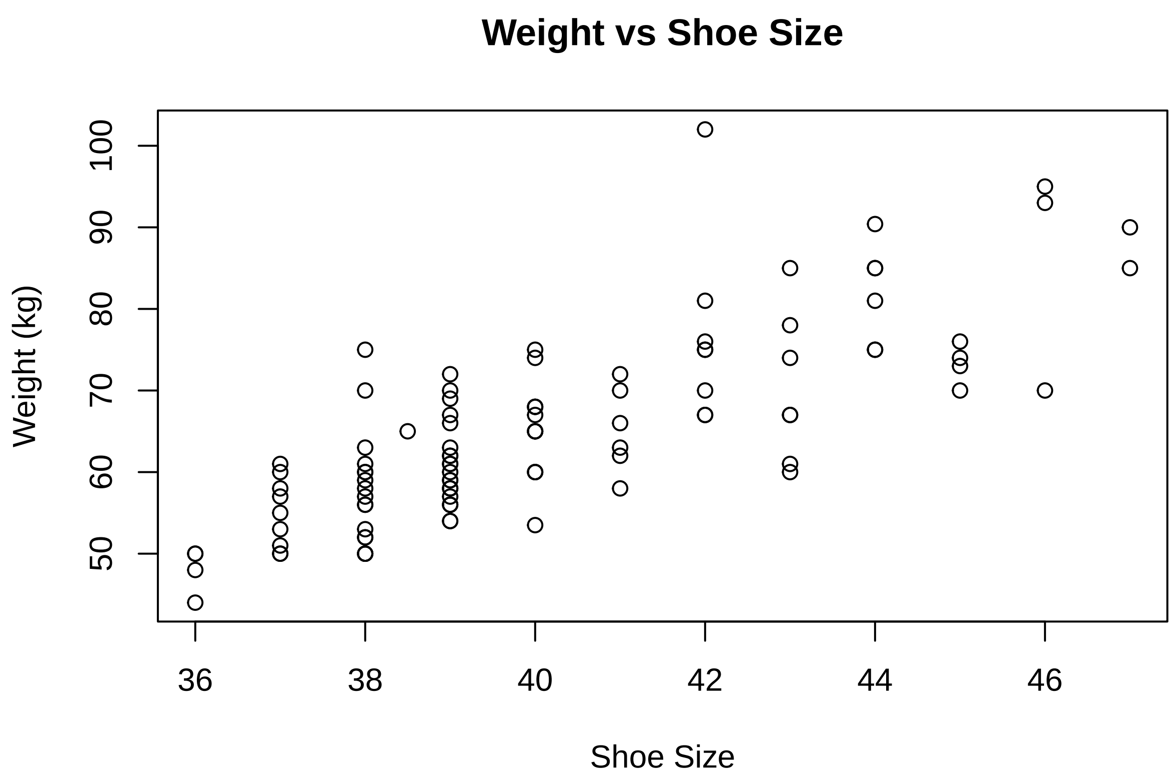
```
##
##      Two-sample t test power calculation
##
##               n = 25
##               d = 0.75
##       sig.level = 0.01
##           power = 0.4937293
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Since p-value is significantly less than 0.01, we reject H0 null hypothesis and therefore shoe size for male and female is significantly different.

The calculated power of the two-sample t-test is approximately 0.49, which is below the commonly recommended threshold of 0.80. This means that there is a less than 50% chance of correctly detecting a true difference in shoe size between males and females if such a difference exists.

Task 2. Graph of Weight vs Shoe Size

```
plot(x=my.data$Shoe.size, y=my.data$Weight,
     xlab = "Shoe Size", ylab = "Weight (kg)",
     main = "Weight vs Shoe Size")
```



The plot reveals a positive trend — as shoe size increases, weight also tends to increase. However, the relationship is not perfectly linear, and there is considerable spread in the data for each shoe size, suggesting moderate variability.

Task 3. Weight and Sport Hours Per Week correlation

```
cor.test(my.data$Weight, my.data$Sport..hours.per.week., method = "spearman")
```

```
## Warning in cor.test.default(my.data$Weight, my.data$Sport..hours.per.week., :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  my.data$Weight and my.data$Sport..hours.per.week.
## S = 132105, p-value = 0.002341
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.2952879
```

Since p-value < 0.05, we reject the H0 null hypothesis. The value $= 0.295$ indicates a moderate positive relationship. This suggests that, in this sample, students who report spending more time doing sports also tend to weight more.

There is a statistically significant monotonic relationship between weight and sport hours per week in the population.

Task 4. Pet owners ratio check

To check if the ratio is different than dog 41%, cat 33%, fish 10%, other 16%.

```r
dog_count <- sum(my.data$PetDog == "Yes")
cat_count <- sum(my.data$PetCat == "Yes")
fish_count <- sum(my.data$PetFish == "Yes")
other_count <- sum(my.data$PetOther == "Yes")
observed <- c(dog_count, cat_count, fish_count, other_count)
total_count <- sum(observed)

expected.proportions <- c(0.41, 0.33, 0.10, 0.16)
expected <- total_count * expected.proportions

chisq.test(observed, p = expected.proportions)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  observed
## X-squared = 0.30993, df = 3, p-value = 0.9582
```

Since the p-value is much greater than 0.05, we fail to reject the null hypothesis. There is no significant difference between the observed pet ratio and the expected one.