Please, You save it at the form of .Rmd

THE DATA: file employee.csv

The file employee.csv is a database of administrative employees, restated according to the database provided by Andrew F. Siegel.1[1] Thus gender is coded 0 for female and 1 for male. Two progressive level training courses are offered from time to time on a voluntary basis. The training variable counts the number of training sessions carried out by the considered employee. The experience variable gives the number of years of experience in the position held. The annual salary ($) is stored in the salary variable.

Exercise 1

We will focus on Student's t distribution in this exercise.

a. Check that the sample size is n = 71 observations.

b. Noting $T_{d.d.l.}$ a random variable following a Student's t law at d.d.l. degrees of freedom, calculate the probability P ($T_{69} > 0$) [2]. Comment.

c. Find the point $q_{0.8}$ such that P ($T_{69} < q_{0.8}$) = 0.8 (80% of the observations are below).

d. Find the point $t^{\star}_{69}$ such that P ($|T_{69}| \geq t^{\star}_{69}$) = α where α = 5%.

Exercise 2

Consider experience as the explanatory variable and salary as the explained variable.

a. Calculate the mean of each of these variables as well as their respective standard deviation.

b. Use a scatter plot to represent the relationship between these two variables and describe the relationship between them.

c. Calculate and interpret their correlation coefficient. Comment against the scatter plot.

d. Give the equation of the regression line that connects the two variables and plot it on the scatter plot.

e. Calculate the standard error S_{b1} of the slope b1 of the regression line.

f. Deduce the 95% confidence interval of the slope b1.

g. Test at the 5% threshold if the slope is significantly different from 0. Interpret the result.

h. How much of the variability in wages can be explained by the fact that some employees have more experience than others?

i. What annual salary could be expected from an employee with 8 years of experience? And an employee with 3 years of experience?

j. Examine the regression conditions based on the residuals.

k. Create employee_f a base that includes all employees and employee_m the rest of the sample. Calculate the slope of the regression in each subgroup and comment.

---

[1] - Practical Business Statistics, 6th ed. (Elsevier, 2012).

[2] - Hint: use the lower.tail argument.