

Compulsory exercise 1: Group 3

TMA4268 Statistical Learning V2023

Helle Villmones Haug, Hjalmar Jacob Vinje and Sanna Baug Warholm

08 February, 2023

Problem 1

For this problem you will need to include some LaTeX code. Please install latex on your computer and then consult Compulsor1.Rmd for hints how to write formulas in LaTeX.

a)

b)

c)

d)

e)

Problem 2

a)

i)

The `lm()` function creates a variable to be estimated which is multiplied with the binary variable of `rankAsstProf` and `rankProf`. The interpretation of the number corresponding to the variables is that holding all other variables constant going from a `AsstProf` to `AssocProf` is associated with an increase in salary of 12,907.6 and going from `AsstProf` to `Prof` is associated with an increase in salary of 45,066.0 holding all other variables constant.

ii)

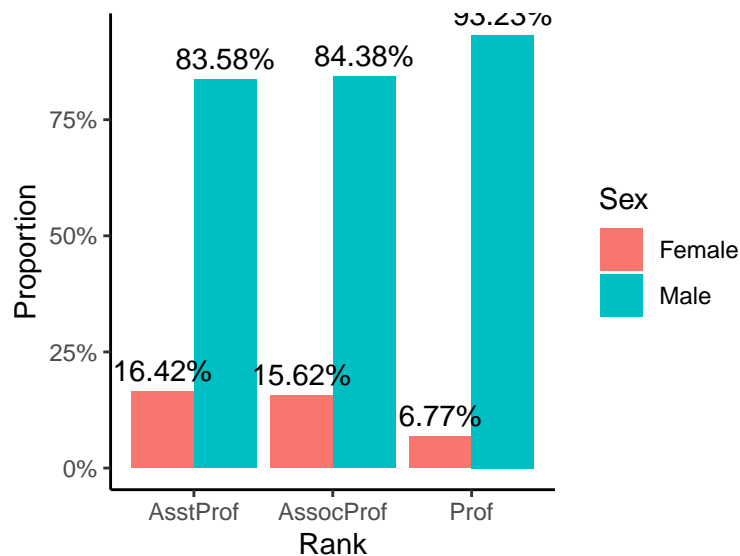
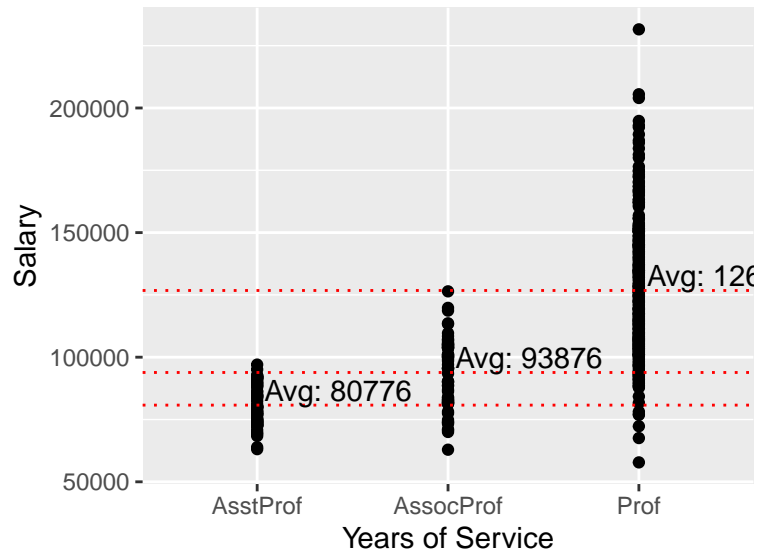
To test the whole categorical variable `Rank`, you can use an ANOVA test to compare the means of salary for each category of `Rank`.

```
      Df    Sum Sq   Mean Sq F value Pr(>F)
rank      2 1.432e+11 7.162e+10   128.2 <2e-16 ***
Residuals 394 2.201e+11 5.586e+08
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This shows that rank as a whole has an impact on salary at the 99% confidence level.

b)



The first graph shows that professors earn more than assistant professors and associate professors. The second shows that women make up a higher proportion of AsstProf than Prof.

This is the reason why the regression with sex as the only covariate shows that sex is statistically significant in terms of salary, but it is no longer significant when controlling for rank and the other covariates. ## c)

Problem 3

Problem 4