

# Howework 4

J Lacasa

2024-10-22

Read the data in the chunk below. The data can be also found [here](https://raw.githubusercontent.com/jlacasa/stat705_fall2024/main/classes/data/prostate.csv).

The data come from a study that aims to “*examine the relationship of preoperative ambulatory serum prostate specific antigen (PSA) to cancer volume, capsular penetration into the periprostatic fat, seminal vesicle invasion, margin-positive tissue planes, lymph node metastases, Gleason grade of the prostate cancer, prostate weight and amount of benign prostatic hyperplasia (BPH) within the radical prostatectomy specimen*”.

Please provide all your code in order to make your results reproducible and submit an html od pdf file to canvas by October 31st.

```
url <- "https://raw.githubusercontent.com/jlacasa/stat705_fall2024/main/classes/data/prostate.csv"
dd_prostate <- read.csv(url)
names(dd_prostate)
```

```
## [1] "lcavol" "lweight" "age"      "lbph"      "svi"      "lcp"      "gleason"
## [8] "pgg45"  "lpsa"     "train"
```

## 1. Statistical model

Design a statistical model that describes log PSA (`lpsa` column in the dataframe) as a function of at least 3 of the predictor variables, and provide the estimates (i.e., the values for your  $\hat{\beta}$ ). Fit the statistical model using only the data that contain `train = TRUE`, and evaluate the model with the data that contain `train = FALSE`.

Please mention:

1. The variable selection criteria you use.
2. The estimates (i.e., the values for your  $\hat{\beta}$ ) you obtained for your final model.

*Note: there are many ways to complete this task. As long as it is correctly justified, you can pick any option we studied in class.*

## 2. Results

Interpret and describe your results in 3 sentences. For example: mention the relationship between the different predictors and the response `lpsa` (e.g., positive or negative), mention the relative importance of the different predictors to explain changes in `lpsa` (e.g., variable  $x_1$  showed a stronger relationship to `lpsa` than  $x_2$ ).