Statictical Modeling and Advanced Regression Analyses

R Tutorials

Holger Sennhenn-Reulen
Northwest German Forest Research Institute (NW-FVA), Germany.

October 14, 2024

Contents

1 Software

We use the statistical software environment *R* (R Core Team, 2024), and R add-on packages *ggplot2* (Wickham, 2016), and *arm* (Gelman & Su, 2024).

This document is produced using *Quarto* (Allaire et al., 2024).

1.1 Organize R Session

```
rm(list = ls())
library("ggplot2")
```

2 Linear Regression Model

Data are simulated according to the equations given in the lecture slides[1]:

```
set.seed(123)
N <- 500
df <- data.frame(x_1 = runif(n = N),
                 x_2 = runif(n = N))
(beta_0 <- rnorm(n = 1, mean = 1, sd = .1))
```

```
[1] 0.9398107
```

```
(beta_x_1 <- rnorm(n = 1, mean = 1, sd = .1))
```

```
[1] 0.9006301
```

```
(beta_x_2 <- rnorm(n = 1, mean = -.5, sd = .1))
```

```
[1] -0.3973215
```

```
(sigma <- rgamma(n = 1, shape = 1, rate = 4))
```

```
[1] 0.293026
```

[1] For two covariates `x_1` and `x_2`.

```
df$mu <- beta_0 + beta_x_1 * df$x_1 + beta_x_2 * df$x_2
df$y <- df$mu + rnorm(n = N, mean = 0, sd = sigma)
```

```
ggplot(data = df, aes(x = x_1, y = x_2)) +
  geom_point()
```

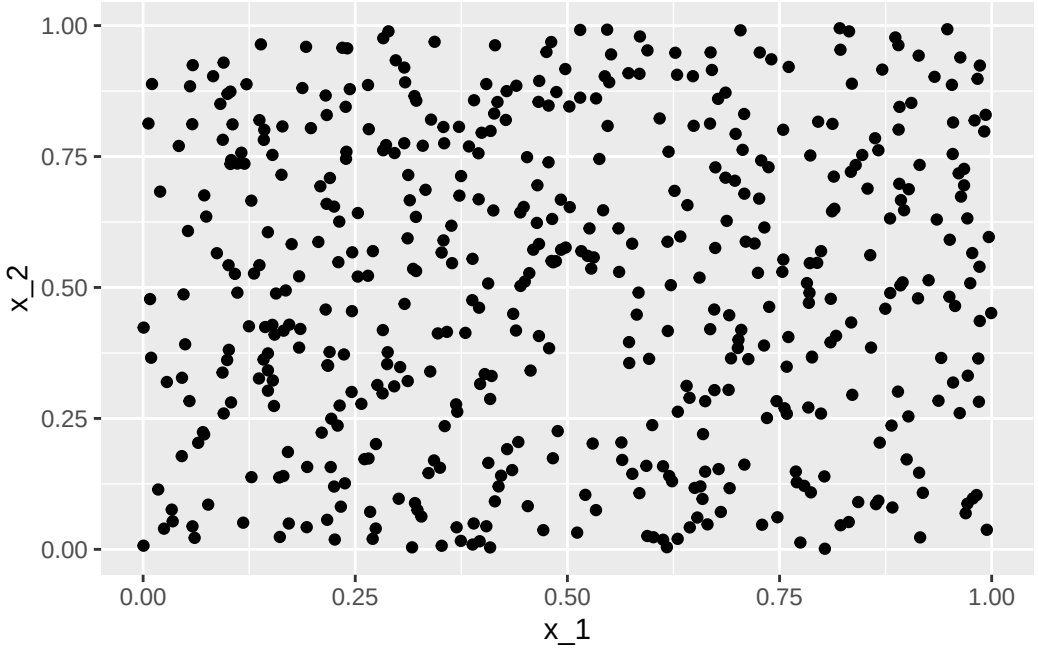

Figure 1: Scatterplot of the two simulated covariates `x_1` and `x_2` - each from the uniform distribution between 0 and 1.

References

Allaire, J. J., Teague, C., Scheidegger, C., Xie, Y., & Dervieux, C. (2024). *Quarto (Version 1.4.553)*. https://doi.org/10.5281/zenodo.5960048

Gelman, A., & Su, Y.-S. (2024). *Arm: Data analysis using regression and multilevel/hierarchical models*. https://CRAN.R-project.org/package=arm

R Core Team. (2024). *R: A Language and Environment for Statistical Computing (Version 4.4.1)*. R Foundation for Statistical Computing.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org