Fundamentals of Statistical Modeling (VT21)

Andrea Discacciati
Karolinska Institutet
Stockholm, Sweden

Lab 1

{{1}}

Exercise 0 (local macros)

In this course, we'll make extensive use of Stata's `local macros`. If they're new to you, get familiar with the help of the code below. Copy and paste it in a new do-file and run it one chunk at a time.

The Stata command `display` (abbreviated with `di`) prints strings in the Results window and can also be used as a hand calculator (`help display`).

```
* Chunk 1
local a = "Fundamentals of statistical modeling."
di "`a'"

* Chunk 2
local a = "statistical modeling"
di "Fundamentals of `a'."

* Chunk 3
local a = "Fundamentals of"
local b = "statistical"
local c = "modeling."
di "`a' `b' `c'"

* Chunk 4
local a = "Fundamentals"
local b = "`a' of"
local c = "`b' statistical"
local d = "`c' modeling."
di "`d'"

* Chunk 5
di exp(2) * exp(-2)

* Chunk 6
local a = "exp(2)"
local b = "exp(-2)"
di `a' * `b'

* Chunk 7
sysuse auto, clear
describe
```

```
local yvar = "weight"
local xvars = "length mpg"
local xvars2 = "price"
regress `yvar' `xvars' `xvars2'
```

Exercise 1

Some of the distributions we'll be working with in this course are:

- Normal: `https://en.wikipedia.org/wiki/Normal_distribution`
- Skew-normal: `https://en.wikipedia.org/wiki/Skew_normal_distribution`
- Gamma: `https://en.wikipedia.org/wiki/Gamma_distribution`
- Beta: `https://en.wikipedia.org/wiki/Beta_distribution`
- Exponential: `https://en.wikipedia.org/wiki/Exponential_distribution`
- Chi-squared: `https://en.wikipedia.org/wiki/Chi-squared_distribution`
- Weibull: `https://en.wikipedia.org/wiki/Weibull_distribution`
- Bernoulli: `https://en.wikipedia.org/wiki/Bernoulli_distribution`
- Negative binomial: `https://en.wikipedia.org/wiki/Negative_binomial_distribution`

Get familiar with the distributions above.

- Are they useful to model continuous or discrete variables?
- What's their support?
- By how many parameters are they parametrized?
- What does their shape look like?
- Which Stata functions implement their probability mass / density functions? See `help density_functions`.

Exercise 2

Load the data

{{2}}

We assume that $f(y_n)$ is normal. Estimate the parameters μ and σ. Use the `normalden()` function.

{{3}}

Estimate again the parameters μ and σ, but this time constrain the parameter σ to be positive by replacing $\sigma = \exp(\theta)$. You'll get now an estimate of θ. Recover the MLE of σ. (Note that the MLE $\hat{\sigma}$ hasn't changed – see slide 33 – but this "trick" improves computational stability).

{{4}}

Plot the estimated density $\hat{f}(y_n)$ over the sample histogram.

{{5}}

Plot the model-based estimated CDF $\hat{F}(y_n)$ (see slide 38) over the empirical CDF (see slide 20). The function `normal()` returns the CDF of a standard normal distribution.

{{6}}

Exercise 3

Assume that $f(y_g)$ is (2-parameter) gamma. Estimate the parameters α and β using the `gammaden()` function. Fix the location parameter g (the third argument of the `gammaden()` function), equal to 0.

{{7}}

Plot the estimated density $\hat{f}(y_g)$ over the sample histogram.

{{8}}

Exercise 4

We assume that $f(y_b)$ is beta. Estimate the parameters α and β using the `betaden()` function.

{{9}}

Plot the estimated density $\hat{f}(y_b)$ over the sample histogram. (OBS: We need to be careful when plotting the estimated density function close to the boundaries of the distribution's support).

{{10}}

Exercise 5

We assume that $f(y_e)$ is exponential. Estimate the parameter λ. You'll need to code the density of an exponential distribution yourself: $f(y; \lambda) = \lambda \exp(\lambda y)$

{{11}}

Plot the estimated density $\hat{f}(y_e)$ over the sample histogram.

{{12}}

Exercise 6

We assume that $f(y_c)$ is chi-squared. Estimate the parameter k using the `chi2den()` function.

{{13}}

Plot the estimated density $\hat{f}(y_c)$ over the sample histogram.

{{14}}