

# Statistical test

**Student:** Mirjam Pergar

**Proposer and mentor:** dr. asist. Gregor Šega

20th April 2017

# Description of the project

- 1 Generate

$$X_{1,1}, X_{1,2}, X_{1,3}, \dots, X_{1,n}$$

$$\vdots$$

$$X_{r,1}, X_{r,2}, X_{r,3}, \dots, X_{r,n}$$

for a i.i. normally distributed random variable  $X$  and some  $r, n$ .

- 2 Calculate sample variances  $s_1^2, \dots, s_r^2$ .
- 3 Test homoscedacity for significance level  $\alpha$ , for some  $\alpha$ .
- 4 Use the proposed statistical test:

$$F = \frac{\sum_{i=1}^r s_{(i)}^2 \frac{2i-1}{r}}{\sum_{i=1}^r s_i^2}$$

# Assumptions

- $X \sim N(0, 1)$
- $n = \{2, 3, 4, \dots, 9, 10, 12, 14, \dots, 18, 20, 25, 30, 40, 60\}$
- $r = \{2, 3, 4, \dots, 9, 10, 15, 20\}$
- $\alpha = \{1\%, 5\%, 10\%\}$

# Assumptions

- $X \sim N(0, 1)$
- $n = \{2, 3, 4, \dots, 9, 10, 12, 14, \dots, 18, 20, 25, 30, 40, 60\}$
- $r = \{2, 3, 4, \dots, 9, 10, 15, 20\}$
- $\alpha = \{1\%, 5\%, 10\%\}$

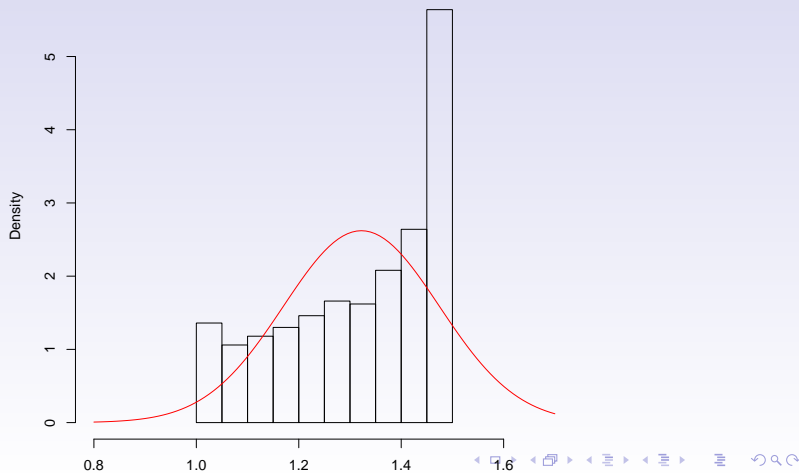
- (a) How many times should we generate the random variables?
- (b) How long will this take?

# Loops

$n = 2, r = 2$

$1e+3$  loops

Histogram of F

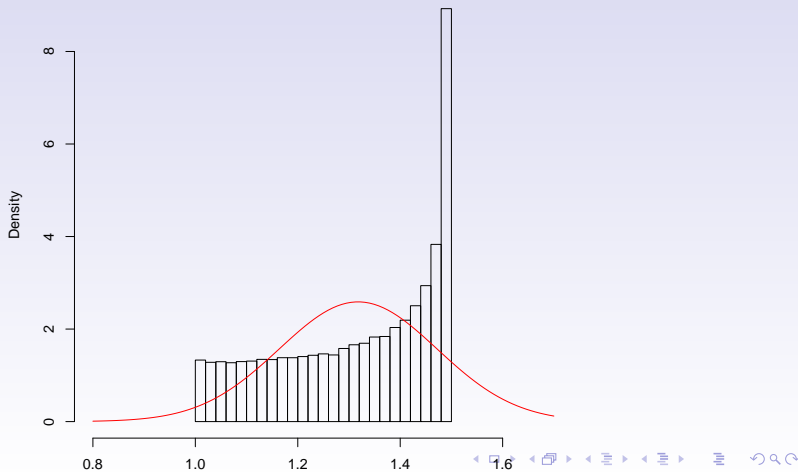


# Loops

$n = 2, r = 2$

$1e+5$  loops

Histogram of F

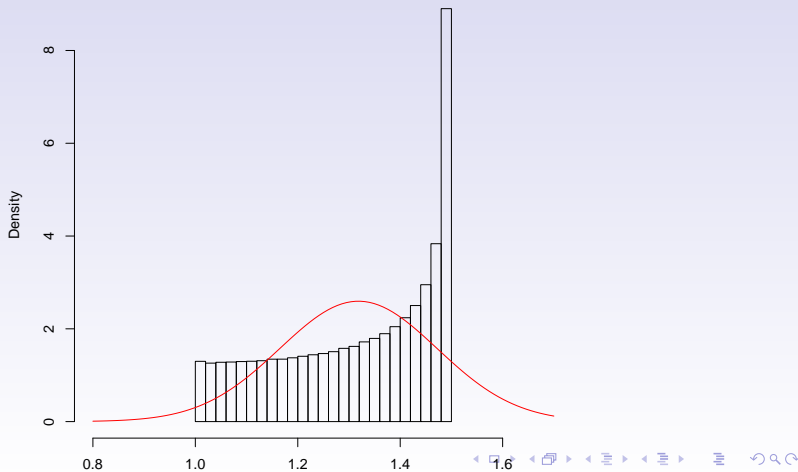


# Loops

$n = 2, r = 2$

$1e+6$  loops

Histogram of F

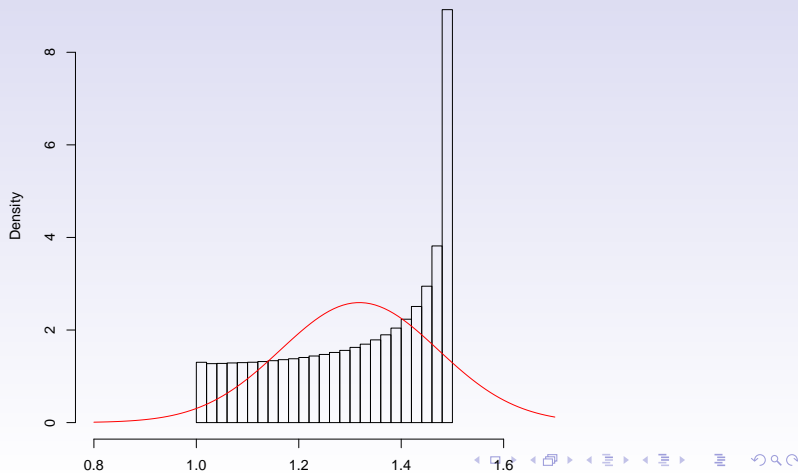


# Loops

$$n = 2, r = 2$$

*1e+7 loops*

Histogram of F





- $n = 2, r = 2$ : 15,06s
- $n = 2, r = 20$ : 19,22s
- $n = 10, r = 20$ : 69,25s
- $n = \{2, \dots, 10\}, r = \{2, \dots, 10\}$ : 2123,33s – 35min
- $n = 60, r = 20$ : 21,21min

One table can be generated in  $\sim 3 - 4h$ .

**Main goal:** Generate three tables

**Work done so far:** Generated all the tables

**Plans:**

- Compare the test to well known tests:
  - Bartlett's test
  - Levene's test
  - Brown–Forsythe test

## Plans:

- Compare the test to well known tests:
  - Bartlett's test:

$$T = \frac{(nr - k) \ln s_p^2 - \sum_{i=1}^r (n - 1) \ln s_i^2}{1 + \frac{1}{3(r-1)} \left( \frac{r}{n-1} - \frac{1}{nr-r} \right)},$$

where  $s_p^2$  is the pooled variance  $s_p^2 = \sum_{i=1}^r \frac{(n-1)s_i^2}{nr-r}$ .

- Levene's test :

$$W = \frac{(nr - r) \sum_{i=1}^r n(\bar{Z}_{i.} - \bar{Z}_{..})^2}{(r - 1) \sum_{i=1}^r \sum_{j=1}^n (z_{ij} - \bar{Z}_{i.})^2},$$

where  $Z_{ij} = |X_{ij} - \bar{X}_j|$  ( $\bar{X}_j$  is the mean of group  $j$ ).

- Brown-Forsythe test:

$$F = \frac{(nr - r) \sum_{i=1}^r n(\tilde{z}_{i.} - \tilde{z}_{..})^2}{(r - 1) \sum_{i=1}^r \sum_{j=1}^n (z_{ij} - \tilde{z}_{.j})^2},$$

where  $z_{ij} = |X_{ij} - \tilde{X}_j|$  ( $\tilde{X}_j$  is the median of group  $j$ ).

## Plans:

- Compare the test to well known tests:
  - Bartlett's test
  - Levene's test
  - Brown–Forsythe test
- Improve the test (if time permits)