

Анализа података у биологији

Марија Цупарић

Осми час



Претпоставимо да имамо популацију и на њој дефинисану случајну величину X чији су средња вредност m и дисперзија σ^2 непознати.

Тачкаста оцена параметра σ^2 је $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

Оцена S^2 има пожељне особине, тј. $ES^2 = \sigma^2$ и DS^2 тежи нули када n тежи бесконачности.

Пример Претпоставимо да студент мери температуру кључања (y °C) одређене течности и добио је резултате 102.5, 101.7, 103.1, 100.9, 100.5 и 102.2 на 6 различитих узорака течности. На основу овог узорка, добија се да је оцена за дисперзију температуре кључања:

$$\begin{aligned} s^2 &= \frac{1}{5} (102.5^2 + 101.7^2 + 103.1^2 + 100.9^2 + 100.5^2 + 102.2^2) - \frac{6}{5} 101.8167^2 \\ &= 0.9615213 \end{aligned}$$

Интервал поверења за параметар дисперзије

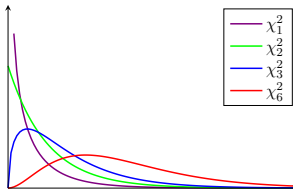
Видели смо да је узорачка дисперзије S^2 логична оцена дисперзије σ^2 . Која је расподела од S^2 ?

Теорема: Нека је $X_1, \dots, X_n$ узорак обима n из нормалне расподеле са параметрима μ и σ^2 . Тада $\frac{(n-1)S_n^2}{\sigma^2}$ има χ^2 расподелу са $n - 1$ степени слободе.

Шта је χ^2 расподела?



Хи-квадрат расподела са ν степени слободe, $X \in \chi^2_\nu$, $\nu \in \mathbb{N}$



- ★ Има један параметар ν , број степени слободe
- ★ Непрекидна је и несиметрична, а узима ненегативне вредности
- ★ Математичко очекивање је ν , а дисперзија 2ν

Теорема: Нека је $X_1, \dots, X_n$ узорак обима n из стандардне нормалне расподеле. Тада $\sum_{k=1}^n X_k^2$ има χ^2_n расподелу.



Читање таблица χ^2 расподеле

χ^2 расподела - вредности x такве да је $P\{X_\nu^2 < x\} = p$

	p											
ν	0.001	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995	0.999
1	0.000	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879	10.828
2	0.002	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597	13.816
3	0.024	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838	16.266
4	0.091	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860	18.467
5	0.210	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750	20.515
6	0.381	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548	22.458
7	0.598	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278	24.322
...	...	...	...	...	...	...	...	...	...	...	...	...
20	5.921	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997	45.315
21	6.447	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401	46.797
22	6.983	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796	48.268
...	...	...	...	...	...	...	...	...	...	...	...	...

- ★ Тражимо x такво да је $P\{X_5^2 < x\} = 0.05$, из таблице: $x = 1.145$.
- ★ Тражимо x такво да је $P\{X_{20}^2 > x\} = 0.025$, онда је $P\{X_{20}^2 \leq x\} = 0.975$, па је $x = 34.170$



Случајна величине X_ν^2 има χ_ν^2 расподелу

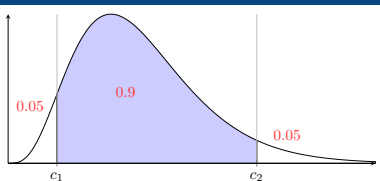
- ★ Вероватноћу догађаја $P\{X_\nu^2 < x\}$ рачунамо функцијом `pchisq(x, df=ν)`.

```
R > pchisq(4.25,df=3)
0.7642966
```

- ★ Вредност x за коју је $P\{X_\nu^2 < x\} = p$ рачунамо функцијом `qchisq(p, df=ν)`.

```
R > qchisq(0.90,df=7)
12.01704
```

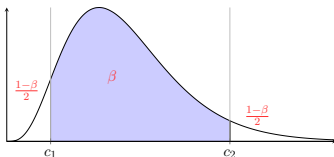
Пример Ако X има χ^2 са 9 степени слободe, одредити c_1 и c_2 тако да важи $P\{c_1 < X < c_2\} = 0.9$ и $P\{X < c_1\} = P\{X > c_2\}$.



Видимо да је површина графика лево од c_1 једнака 0.05 па c_1 налазимо као `qchisq(0.05, df = 9)`, а то је 3.325113. С друге стране, површина графика лево од c_2 је 0.95 па c_2 налазимо као `qchisq(0.95, df = 9)`, а то је 16.91898.

Нека је $X_1, \dots, X_n$ случајни узорак обима n из нормалне расподеле с параметрима μ и σ^2 .

- ★ случајна величина: $T_n = \frac{(n-1)S_n^2}{\sigma^2} \in \chi_{n-1}^2$
- ★ расподела није симетрична па поред тога да је $P\{c_1 < T_n < c_2\}$ мора се додати још неки услов, а то је најчешће да је интервал „површински симетричан”



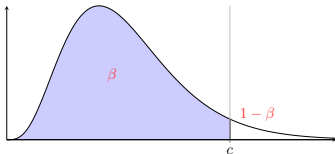
- ★ треба одредити константе c_1 и c_2 тако да $P\{T_n < c_1\} = \frac{1-\beta}{2}$ и $P\{T_n > c_2\} = \frac{1-\beta}{2}$

$$\star c_1 < T_n < c_2 \iff c_1 < \frac{(n-1)S_n^2}{\sigma^2} < c_2 \iff \frac{(n-1)S_n^2}{c_2} < \sigma^2 < \frac{(n-1)S_n^2}{c_1}$$

Двострани интервал поверења је $(\frac{(n-1)S_n^2}{c_2}, \frac{(n-1)S_n^2}{c_1})$.

Једнострани интервали поверења

- ★ треба одредити константу c тако да $P\{T_n < c\} = \beta$

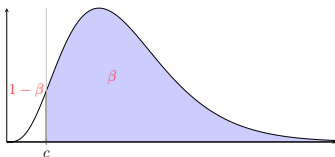


$$★ c = F_{\chi_{n-1}^2}^{-1}(\beta)$$

$$★ T_n < c \iff \sigma^2 > \frac{(n-1)S_n^2}{c}$$

Интервал поверења је $(\frac{(n-1)S_n^2}{c}, \infty)$

- ★ треба одредити константу c тако да $P\{T_n > c\} = \beta$



$$★ c = F_{\chi_{n-1}^2}^{-1}(1 - \beta)$$

$$★ T_n > c \iff \sigma^2 < \frac{(n-1)S_n^2}{c}$$

Интервал поверења је $(0, \frac{(n-1)S_n^2}{c})$



МАТ
Универзитет
Математички

КАТЕДРА ЗА ВЕРОВАЈНОСТ

Пример Власник музичке продавнице жели да процени старост својих купаца. Случајан узорак од 31-ог купца има средњу вредност 23 године са узорачком стандардним одступањем 5 година. Одредити 95% интервал поверења за дисперзију броја година.

Двострани интервал поверења је $(\frac{(n-1)S_n^2}{c_2}, \frac{(n-1)S_n^2}{c_1})$.

★ $c_1 = F_{\chi_{30}^2}^{-1}(\frac{1-0.95}{2}) = 16.79077$, $c_2 = F_{\chi_{30}^2}^{-1}(\frac{1+0.95}{2}) = 46.97924$

$qchisq(0.025, 30)$, $qchisq(0.975, 30)$

$\implies \left(\frac{30 \cdot 25}{46.97924}, \frac{30 \cdot 25}{16.79077} \right) = (15.9645, 44.6674)$.

Интервал поверења за стандардно одступање се добија узимањем корена леве и десне границе претходног интервала.

Ако нам треба само горње ограничења броја година, можемо одредити једнострани интервал поверења $(0, \frac{(n-1)S_n^2}{c})$, где је $c = F_{\chi_{n-1}^2}^{-1}(1 - \beta)$.

Односно, $c = F_{\chi_{30}^2}^{-1}(1 - 0.95) = 18.49266$ ($qchisq(0.05, 30)$)

$\implies \left(0, \frac{30 \cdot 25}{18.49266} \right) = (0, 40.55663)$

Тестирање хипотеза у вези са параметром дисперзије

хипотезе:

- ★ $H_0(\sigma^2 = \sigma_0^2)$ против $H_1(\sigma^2 \neq \sigma_0^2)$ (двострани тест)
- ★ $H_0(\sigma^2 \geq \sigma_0^2)$ против $H_1(\sigma^2 < \sigma_0^2)$ (једнострани леви)
- ★ $H_0(\sigma^2 \leq \sigma_0^2)$ против $H_1(\sigma^2 > \sigma_0^2)$ (једнострани десни)

тест статистика: $T_n = \frac{(n-1)S_n^2}{\sigma_0^2}$ при H_0 има χ_{n-1}^2

p -вредност теста:

- ★ код једностраног левог теста је површина лево од вредности t_0 коју статистика T_n узме у узорку;
- ★ код једностраног десног теста је површина десно од вредности t_0 коју статистика T_n узме у узорку;
- ★ код двостраног теста је **приближно** двострука површина лево или десно од вредности t_0 , у зависности од тога која од тих површина је мања.

Како је S_n^2 оцена од σ_0^2 , велике вредности тест статистике упућују на то да је $\sigma^2 > \sigma_0^2$, док мале вредности тест статистике упућују на то да је $\sigma^2 < \sigma_0^2$.

Пример Унутрашњи притисак тениских лоптица је нормално расподељен са средњом вредношћу 28 и дисперзијом 0.25. Проверава се да ли нови процес производње има мању дисперзију притиска, јер ће онда лоптице имати униформније одскакање. Узет је узорак од 25 вредности: 28.20, 27.31, 28.68, 27.98, 27.99, 28.04, 27.47, 28.57, 28.12, 28.75, 28.36, 27.96, 28.30, 28.29, 28.40, 27.46, 27.99, 27.94, 27.76, 27.91, 27.59, 27.71, 28.60, 27.91, 27.85. Са нивоом значајности 0.05, испитати да ли је произвођач у праву.

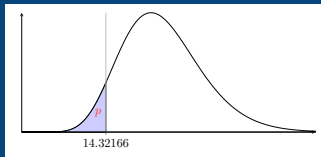
Поставимо хипотезе: $H_0(\sigma^2 = 0.25)$ $H_1(\sigma^2 < 0.25)$.

Тест статистика: $T_n = \frac{(n-1)S_n^2}{\sigma_0^2}$ при H_0 има χ_{24}^2

Реализована вредност тест статистике: $\frac{(n-1)s_n^2}{\sigma_0^2} = \frac{24 \cdot 0.15}{0.25} = 14.32166$.

p —вредност теста:

$$\begin{aligned} P_{H_0}\{T_n \leq 14.32166\} &= F_{\chi_{24}^2}(14.32166) \\ &= 0.06 \end{aligned}$$



Пошто је p —вредност већа од α , немамо доказа да одбацимо H_0 , па сма-трамо да нове лопте немају мањи притисак од стандардних.

У R-у за тестирање се може користити:

 `varTest()` % пакет `EnvStats`



```
x<-c(28.20, 27.31, 28.68, 27.98, 27.99, 28.04, 27.47, 28.57, 28.12,
      28.75, 28.36, 27.96, 28.30, 28.29, 28.40, 27.46, 27.99, 27.94,
      27.76, 27.91, 27.59, 27.71, 28.60, 27.91, 27.85)
(length(x)-1)*var(x)/0.25
qchisq(0.05,24)
pchisq(14.32166,24)
varTest(x,alternative = "less", sigma.squared=0.25)
```

Results of Hypothesis Test

```
-----
Null Hypothesis:           variance = 0.25
Alternative Hypothesis:    True variance is less than 0.25
Test Name:                 Chi-Squared Test on Variance
Estimated Parameter(s):    variance = 0.149184
Data:                      x
Test Statistic:            Chi-Squared = 14.32166
Test Statistic Parameter:  df = 24
P-value:                   0.06095146
95% Confidence Interval:   LCL = 0.0000000
                           UCL = 0.2585432
```

Пример Унутрашњи притисак тениских лоптица је нормално расподељен са средњом вредношћу 28 и дисперзијом 0.25. Проверава се да ли нови процес производње даје другачију дисперзију притиска од старе. Узет је узорак од 25 вредности: 28.20, 27.31, 28.68, 27.98, 27.99, 28.04, 27.47, 28.57, 28.12, 28.75, 28.36, 27.96, 28.30, 28.29, 28.40, 27.46, 27.99, 27.94, 27.76, 27.91, 27.59, 27.71, 28.60, 27.91, 27.85. Са нивоом значајности 0.05, испитати да ли је произвођач у праву.

Поставимо хипотезе: $H_0(\sigma^2 = 0.25)$ $H_1(\sigma^2 \neq 0.25)$.

Тест статистика: $T_n = \frac{(n-1)S_n^2}{\sigma_0^2}$ при H_0 има χ_{24}^2

Реализована вредност тест статистике: $\frac{(n-1)s_n^2}{\sigma_0^2} = \frac{24 \cdot 0.15}{0.25} = 14.32166$.

p —вредност теста је приближно:

$$2P_{H_0}\{T_n \leq 14.32166\} = 2F_{\chi_{24}^2}(14.32166) \\ = 0.12$$

R varTest(x, sigma.squared=0.25)

p-value: 0.1219029

Пошто је p —вредност већа од α , немамо доказа да одбацимо H_0 , па сматрамо да нове лопте немају мањи притисак од стандардних.

Статистичко закључивање у моделу са нормалном расподелом

Урађено до сада: $X \in \mathcal{N}(m, \sigma^2)$

- ★ тачкаста и интервална оцена за m , тестирање у вези са параметром m
- ★ тачкаста и интервална оцена за σ^2 , тестирање у вези са параметром σ^2

Следи статистичко закључивање за:

- ▷ случај независних узорака из популација $X \in \mathcal{N}(m_1, \sigma_1^2)$ и $Y \in \mathcal{N}(m_2, \sigma_2^2)$
 - ★ поређење параметара средњих вредности
 - ★ поређење параметара дисперзија
- ▷ случај спарених узорака, односно за узорак из популације $D = X - Y \in \mathcal{N}(m_D, \sigma_D^2)$, $m_D = m_1 - m_2$

Статистичко закључивање за два независна узорка

Нека су дати независни узорци $X_1, \dots, X_n$ из популације чије обележје X има нормалну $\mathcal{N}(m_1, \sigma_1^2)$ расподелу и $Y_1, \dots, Y_k$ из популације чије обележје Y има нормалну $\mathcal{N}(m_2, \sigma_2^2)$ расподелу.

Разликујемо случајеве:

★ оцењивање и тестирање хипотеза у вези са параметаром $m_1 - m_2$

- (1) када су σ_1 и σ_2 познати
- (2) када су σ_1 и σ_2 непознати и једнаки
- (3) када су σ_1 и σ_2 непознати и неједнаки

★ оцењивање и тестирање хипотеза у вези са параметаром $\frac{\sigma_1^2}{\sigma_2^2}$

С обзиром да нека статистичка закључивања о средњој вредности зависе од закључивања о дисперзији, прво ћемо се бавити закључивањима у вези са дисперзијама независних узорака.

Поређење дисперзија два нормалне популације

Пример Две куглашице, Марси и Пег, покушавају да се изборе за место у тиму. Обе имају просечну оцену од 160. Пошто је конзистентност важна у куглању, куглашица са најмањом варијабилношћу резултата је најкориснија за тим. Нека X означава резултат који постиже Марси, а Y резултат који постиже Пег. Желимо да упоредимо дисперзије њихових резултата.

На основу мерења, добијено је

X	163	160	155	160	157	159	166	159	166	169	157	160	165	159		
Y	165	177	140	166	170	168	170	163	146	155	146	167	171	163	153	140

За почетак одредимо узорачке дисперзије:

$$s_X^2 = \frac{1}{13} \sum_{i=1}^{14} (x_i - \bar{x})^2 = 16.68681, \quad s_Y^2 = \frac{1}{15} \sum_{i=1}^{16} (y_i - \bar{y})^2 = 137.8667.$$

С обзиром да је разлика између узорачких дисперзија велика, можемо рећи да постоји разлика и између теоријских дисперзија.

Како да знамо шта је велика разлика? Како да знамо да ли је статистички значајна?

Морамо користити интервале поверења и тестирање хипотеза.

- ★ Потребна нам је случајна величина чију расподелу знамо (односно статистика чију расподелу знамо при H_0)
- ★ Када нешто поредимо, можемо поредити разлику ($S_X^2 - S_Y^2$) са нулом или количник ($\frac{S_X^2}{S_Y^2}$) са јединицом

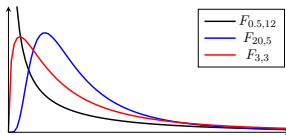
Теорема: Нека су дати независни узорци $X_1, \dots, X_n$ из популације чије обележје X има нормалну $\mathcal{N}(m_1, \sigma_1^2)$ расподелу и $Y_1, \dots, Y_k$ из популације чије обележје Y има нормалну $\mathcal{N}(m_2, \sigma_2^2)$ расподелу. Тада

$$\frac{\frac{S_n^2(X)}{\sigma_1^2}}{\frac{S_k^2(Y)}{\sigma_2^2}} \in \mathcal{F}_{n-1, k-1},$$

где је $\mathcal{F}_{n-1, k-1}$ Фишерава расподела са параметрима $n-1$ и $k-1$.

Шта је Фишерава расподела?

Фишерава расподела са параметрима ν_1 и ν_2 , $X \in \mathcal{F}_{\nu_1, \nu_2}$



- ★ Има два параметра ν_1 и ν_2
- ★ Непрекидна је и несиметрична, а узима ненегативне вредности
- ★ Ако $X \in \mathcal{F}_{\nu_1, \nu_2}$, онда $\frac{1}{X} \in \mathcal{F}_{\nu_2, \nu_1}$



Читање таблица Фишерове $\mathcal{F}_{\nu_1, \nu_2}$ расподеле

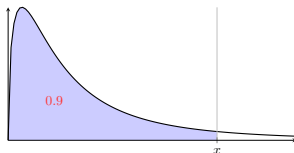


Таблица Фишерове расподеле, $p = 0.90$

$\nu_2 \backslash \nu_1$	2	3	4	5	6	7	8	10	12	15
1	49.5	53.6	55.8	57.2	58.2	59.1	59.7	60.5	61.0	61.5
2	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.39	9.41	9.43
3	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.23	5.22	5.20
4	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.92	3.90	3.87
5	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.30	3.27	3.24
6	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.94	2.90	2.87
7	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.70	2.67	2.63
8	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.54	2.50	2.46
...										

- ★ Тражимо x такво да је $P\{F_{3,5} < x\} = 0.9$, из таблице: $x = 3.62$
- ★ Таблице се праве за сваку вероватноћу посебно, приказана је за $P\{T_{\nu_1, \nu_2} \leq x\} = 0.90$



Случајна величине T има $\mathcal{F}_{\nu_1, \nu_2}$ расподелу

- ★ Вероватноћу догађаја $P\{F_{\nu_1, \nu_2} < x\}$ рачунамо функцијом $\text{pf}(x, \text{df1}=\nu_1, \text{df2}=\nu_2)$.

```
R > pf(3.62, df1=3, df2=5)
0.9000233
```

- ★ Вредност x за коју је $P\{F_{\nu_1, \nu_2} < x\} = p$ рачунамо функцијом $\text{qf}(p, \text{df1}=\nu_1, \text{df2}=\nu_2)$.

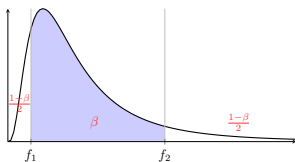
```
R > qf(0.90, df1=3, df2=5)
3.619477
```



Интервал поверења за $\frac{\sigma_1^2}{\sigma_2^2}$

Нека су дати независни узорци $X_1, \dots, X_{n_1}$ из популације чије обележје X има нормалну $\mathcal{N}(m_1, \sigma_1^2)$ расподелу и $Y_1, \dots, Y_{n_2}$ из популације чије обележје Y има нормалну $\mathcal{N}(m_2, \sigma_2^2)$ расподелу.

- ★ случајна величина: $T_n = \frac{\sigma_2^2 S_{n_1}^2(X)}{\sigma_1^2 S_{n_2}^2(Y)} \in \mathcal{F}_{n_1-1, n_2-1}$
- ★ расподела није симетрична па поред тога да је $P\{c_1 < T_n < c_2\}$ мора се додати још неки услов, а то је најчешће да је интервал „површински симетричан”



- ★ треба одредити константе f_1 и f_2 тако да $P\{T_n < f_1\} = \frac{1-\beta}{2}$ и $P\{T_n > f_2\} = \frac{1-\beta}{2}$

$$\star f_1 < T_n < f_2 \iff f_1 \leq \frac{\sigma_2^2 S_{n_1}^2}{\sigma_1^2 S_{n_2}^2} \leq f_2$$

$$\iff \frac{f_1 S_{n_2}^2}{S_{n_1}^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \frac{f_2 S_{n_2}^2}{S_{n_1}^2}$$

Двострани интервал поверења за $\frac{\sigma_1^2}{\sigma_2^2}$ је $(\frac{S_{n_1}^2}{f_2 S_{n_2}^2}, \frac{S_{n_1}^2}{f_1 S_{n_2}^2})$.

Пример Две куглашице, Марси и Пег, покушавају да се изборе за место у тиму. Обе имају просечну оцену од 160. Пошто је конзистентност важна у куглању, куглашица са најмањом варијабилношћу резултата је најкориснија за тим. Нека X означава резултат који постиже Марси, а Y резултат који постиже Пег. Желимо да упоредимо дисперзије њихових резултата.

На основу мерења, добијено је

X	163	160	155	160	157	159	166	159	166	169	157	160	165	159		
Y	165	177	140	166	170	168	170	163	146	155	146	167	171	163	153	140

Одредити 95% интервал поверења за $\frac{\sigma_1^2}{\sigma_2^2}$.

Двострани интервал поверења је $(\frac{S_{n_1}^2}{f_2 S_{n_2}^2}, \frac{S_{n_1}^2}{f_1 S_{n_2}^2})$.

$$\star f_1 = F_{\mathcal{F}_{13,15}}^{-1}(\frac{1-0.95}{2}) = 10.57, \quad c_2 = F_{\mathcal{F}_{13,15}}^{-1}(\frac{1+0.95}{2}) = 24.53$$

$$\text{qf}(0.025, 13, 15), \quad \text{qf}(0.975, 13, 15)$$

$$\Rightarrow \left(\frac{16.69}{24.53 \cdot 137.84}, \frac{16.69}{10.57 \cdot 137.87} \right) = (0.0049, 0.0114).$$

Ако интервал поверења за количник дисперзија не садржи вредност 1, онда се може рећи да је разлика између дисперзија статистички значајна.

Тестирање хипотеза у вези са количником дисперзија

хипотезе:

- ★ $H_0(\sigma_1^2 = \sigma_2^2)$ против $H_1(\sigma_1^2 \neq \sigma_2^2)$ (двострани тест)
- ★ $H_0(\sigma_1^2 = \sigma_2^2)$ против $H_1(\sigma_1^2 < \sigma_2^2)$ (једностран леви)
- ★ $H_0(\sigma_1^2 = \sigma_2^2)$ против $H_1(\sigma_1^2 > \sigma_2^2)$ (једностран десни)

тест статистика: $T_n = \frac{S_{n_1}^2}{S_{n_2}^2}$ при H_0 има $\mathcal{F}_{n_1-1, n_2-1}$

p -вредност теста:

- ★ код једностраног левог теста је површина лево од вредности t_0 коју статистика T_n узме у узорку;
- ★ код једностраног десног теста је површина десно од вредности t_0 коју статистика T_n узме у узорку;
- ★ код двостраног теста је **приближно** двострука површина лево или десно од вредности t_0 , у зависности од тога која од тих површина је мања.

У суштини може се тестирати и хипотеза $H_0(\frac{\sigma_1^2}{\sigma_2^2} = A)$, где је A нека константа.

Разлика у односу на претходно је једино што треба скалирати тест статистику константом A .

Пример Две куглашице, Марси и Пег, покушавају да се изборе за место у тиму. Обе имају просечну оцену од 160. Пошто је конзистентност важна у куглању, куглашица са најмањом варијабилношћу резултата је најкориснија за тим. Нека X означава резултат који постиже Марси, а Y резултат који постиже Пег. Желимо да упоредимо дисперзије њихових резултата. На основу мерења, добијено је

X	163	160	155	160	157	159	166	159	166	169	157	160	165	159		
Y	165	177	140	166	170	168	170	163	146	155	146	167	171	163	153	140

Да ли ови подаци указују на то да постоји разлика у варијабилности резултата ових куглашица?

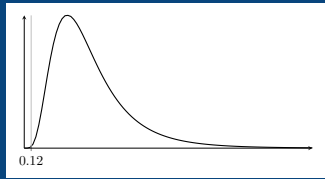
Поставимо хипотезе: $H_0(\sigma_1^2 = \sigma_2^2)$ $H_1(\sigma_1^2 \neq \sigma_2^2)$.

Тест статистика: $T_n = \frac{S_{n_1}^2}{S_{n_2}^2}$ при H_0 има $\mathcal{F}_{13,15}$

Реализована вредност тест статистике: $\frac{s_{n_1}^2}{s_{n_2}^2} = \frac{16.69}{137.87} = 0.12$.

p -вредност теста је приближно:

$$2P_{H_0}\{T_n \leq 0.12\} = 2F_{\mathcal{F}_{13,15}}(0.12) = 0.00044$$



Пошто је p -вредност мала, одбацујемо H_0 , па сматрамо да је разлика у дисперзијама статистички значајна.

У R-у за тестирање се може користити:

R `var.test()`

R

```
x<-c( 163, 160, 155, 160, 157, 159, 166, 159, 166, 169, 157, 160, 165,
      159)
y<-c(165, 177, 140, 166, 170, 168, 170, 163 , 146, 155, 146, 167, 171,
      163, 153, 140)
var.test(x, y)
```

F test to compare two variances

```
data: x and y
F = 0.12104, num df = 13, denom df = 15, p-value = 0.0004605
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.04138114 0.36948783
sample estimates:
ratio of variances
 0.1210359
```



С обзиром да сами одлучујемо која популација је прва а која друга можемо их означити како желимо. Пракса је да се као прва популација означи она која има већу узорачку дисперзију. У том случају, ако H_0 није тачно очекујемо да ће количник $\frac{S_{n_1}^2}{S_{n_2}^2}$ бити велики.

Урадити претходни пример са наведеном изменом!

