SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO-MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK
AK. GOD. 2020. / 2021.

SEMINARSKI RAD IZ PRIMJENJENE STATISTIKE

ZADATAK 2.

AUTOR: Josipa Radnić

ZAGREB, 2021. GOD

Sadržaj

1 Uvod

Zadatak

Zadatak 2.
Hotellingov T^2-test za jedan uzorak
Podaci: ALM, Exercise 1.8.3 (c), str. 68.
(*a*) Izvedite simultane pouzdane intervale za komponente očekivanja (TMS 8.3, str. 104) i procijenite ih na osnovi zadanih podataka.
(*b*) Ispitajte normalnost podataka.
(*c*) Sprovedite test iz navedenog zadatka.

Podaci

Podaci su dobiveni iz ALM, Exercise 1.8.3 (c), Table 1.4.

Dani su podaci o masama za tri skupine štakora u pet različitih vremena. Prva skupina je dobivala tiroksin, druga skupina tiouracil, a treća skupina je kontrolna skupina. Masa je mjerena u gramima i izvršeno je pet mjerenja, svako u razmaku od tjedan dana. Podaci su u sljedećim tablicama.

Tablica 1: Tiroksin

Vrijeme 0	Vrijeme 1	Vrijeme 2	Vrijeme 3	Vrijeme 4
59	85	121	156	191
54	71	90	110	138
56	75	108	151	189
59	85	116	148	177
57	72	97	120	144
52	73	97	116	140
52	70	105	138	171

Tablica 2: Tiouracil

Vrijeme 0	Vrijeme 1	Vrijeme 2	Vrijeme 3	Vrijeme 4
61	86	109	120	129
59	80	101	111	122
53	79	100	106	133
59	88	100	111	122
51	75	101	123	140
51	75	92	100	119
56	78	95	103	108
58	69	93	114	138
46	61	78	90	107
53	72	89	104	122

Tablica 3: Kontrolna

Vrijeme 0	Vrijeme 1	Vrijeme 2	Vrijeme 3	Vrijeme 4
57	86	114	139	172
60	93	123	146	177
52	77	111	144	185
49	67	100	129	164
56	81	104	121	151
46	70	102	131	153
51	71	94	110	141
63	91	112	130	154
49	67	90	112	140
57	82	110	139	169

2 Simultani pouzdani intervali za komponente očekivanja

Neka je $X_1, \dots, X_n$ slučajni uzorak iz *p*-dimenzionalnog normalnog modela $N_p(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$. Za svaki fikisirani $\beta = 1 - \alpha$ definiramo kvantil centralne F - distribucije *Fα(p1,p2)* jednadžbom

$$\mathbb{P}(F(p_1, p_2) \geq F_\alpha(p_1, p_2)) = \alpha.$$

Hotellingova testna statistika je $T^2 = n(\overline{\mathbb{X}} - \boldsymbol{\mu})^\tau S^{-1}(\overline{\mathbb{X}} - \boldsymbol{\mu}) \sim F(p, n-p)$, gdje je *S* uzoračka kovarijacijska matrica uzorka ***X*** i jednaka je $S = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbb{X}_{i\cdot} - \bar{X})(\mathbb{X}_{i\cdot} - \bar{X})^\tau$, što je ujedno i nepristrani procjenitelj za Σ. Sada vrijedi

$$\mathbb{P}(n(\overline{\mathbb{X}} - \boldsymbol{\mu})^\tau S^{-1}(\overline{\mathbb{X}} - \boldsymbol{\mu}) \leq c_\alpha) = \beta \tag{1}$$

gdje smo uzeli $c_\alpha = \frac{(n-1)p}{n-p} F_\alpha(p, n-p)$ kvantil korišten u testu, odnosno vrijedi

$$\mathbb{P}(T^2 \leq c_\alpha) = \beta.$$

Dakle, $\beta * 100\%$ „sigurno" će očekivanje uzorka $\boldsymbol{\mu}$ biti unutar pouzdanog područja – slučajnog elipsoida $\{\boldsymbol{\mu} \in \mathbb{R}^p : T^2 \leq c_\alpha\}$.

Linearne hipoteze

U mnogim eksperimentima želimo testirati određene linearne jednadžbe nad elementima vektora $\boldsymbol{\mu}$. Npr., je li druga komponenta jednaka trećoj komponenti, ili je li suma prve dvije komponente jednaka 2. Odgovor na prvo pitanje, odnosno drugo pitanje je zapravo testiranje nulhipoteze:

1. $H_0: \mu_2 = \mu_3$
2. $H_0: \mu_1 + \mu_2 = 2$

Svaku tu nulhipotezu možemo zapisati u obliku vektorske jednadžbe H_0: $a^\tau \boldsymbol{\mu} = c$, gdje je $c \in \mathbb{R}$ neka konstanta i $a \in \mathbb{R}^p$ neki vektor duljine p. Ako imamo više od jedne (specifično r njih) takve nulhipoteze koje želimo istovremeno testirati, tada zapravo imamo r vektora i konstanti a_i, c_i tako da je naša nulhipoteza:

$$H_0\colon\ {a_1}^\tau \boldsymbol{\mu} = c_1, {a_2}^\tau \boldsymbol{\mu} = c_2, \dots, {a_r}^\tau \boldsymbol{\mu} = c_r\ .$$

Ako definiramo $\boldsymbol{A} = (a_1, \dots, a_r)$ i $\boldsymbol{c} = (c_1, \dots, c_r)^\tau$ tada gornja nulhipoteza se može ekvivalentno zapisati

$$H_0\colon\ \boldsymbol{A}^\tau \boldsymbol{\mu} = \boldsymbol{c}\ .$$

Neka je $x_1, \dots, x_n$ realizacija slučajnog uzorka $X_1, \dots, X_n$ iz *p*-dimenzionalnog normalnog modela $N_p(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$. Stavimo $y_i = \boldsymbol{A}^\tau x_i, i = 1, \dots, n$. Tada su očito $y_1, \dots, y_n$ nezavisni jednakodistribuirani slučajni vektori s distribucijom $N_r(\boldsymbol{A}^\tau \boldsymbol{\mu}, \boldsymbol{A}^\tau \Sigma \boldsymbol{A})$, $\Sigma > 0$.

Uz pretpostavku da su vektori $a_1, \dots, a_r$ linearno nezavisni, možemo primijeniti multivarijantni rezultat direktno na ***y*** pa vrijedi

$$H_0\ istinita \Leftrightarrow n(\bar{y} - c)^\tau S_y^{-1}(\bar{y} - c) \stackrel{\mathrm{D}}{=} \frac{(n-1)r}{n-r} F(r, n-r).$$

Tada je $\beta * 100\%$ -pouzdano područje za $\boldsymbol{A}^\tau\boldsymbol{\mu}$, što je elipsoid, dano s

$$CR(\boldsymbol{A}^\tau\boldsymbol{\mu};\beta) = \left\{\boldsymbol{A}^\tau\boldsymbol{\mu} \in \mathbb{R}^r : n(\boldsymbol{A}^\tau\boldsymbol{\mu} - \bar{y})^\tau S_y^{-1}(\boldsymbol{A}^\tau\boldsymbol{\mu} - \bar{y}) \leqslant k_\alpha\right\} \tag{2}$$

gdje je $k_\alpha = \frac{(n-1)r}{n-r}F_\alpha(r, n-r)$. Primijetite da je $\bar{y} = \boldsymbol{A}^\tau\bar{x}$ i $S_y = \boldsymbol{A}^\tau S\boldsymbol{A}$.

Želimo izvesti *simultane pouzdane intervale* za komponente vektora $\boldsymbol{A}^\tau\boldsymbol{\mu}$, odnosno pouzdane intervale za $A_i^\tau\boldsymbol{\mu}$, $i = 1,\ldots,r$ koji su simultano pouzdani. Za to će nam trebati sljedeća lema:

Lema 1. Neka je $\boldsymbol{S} \in M_p$ i $\boldsymbol{A} = (a_1,\ldots,a_r) \in \mathbb{R}_r^p$ je ranga r. Tada je

$$\frac{(a_i^\tau x)^2}{a_i^\tau \boldsymbol{S} a_i} \leqslant x^\tau \boldsymbol{A}(\boldsymbol{A}^\tau\boldsymbol{S}\boldsymbol{A})^{-1}\boldsymbol{A}^\tau x \leqslant x^\tau \boldsymbol{S}^{-1}x, \forall x \in \mathbb{R}^p \tag{3}$$

Iz nejednadžbe (3) za svaku od komponenti $A_i^\tau\boldsymbol{\mu}$, $i = 1,\ldots,r$, dobivamo sljedeće nejednakosti

$$\frac{(\bar{y}_i - A_i^\tau\boldsymbol{\mu})^2}{a_i^\tau S a_i} \leqslant (\bar{y} - A^\tau\boldsymbol{\mu})^\tau S_y^{-1}(\bar{y} - A^\tau\boldsymbol{\mu}) \leqslant (\bar{x} - \boldsymbol{\mu})^\tau S^{-1}(\bar{x} - \boldsymbol{\mu}) \tag{4}$$

Sada iz (2) i (4) imamo:

$$\mathbb{P}\left(\sqrt{n}\frac{|\bar{y}_i - A_i^\tau\boldsymbol{\mu}|}{\sqrt{(a_i^\tau S a_i)}} \leqslant \sqrt{k_\alpha}, i = 1,\ldots,r\right) \geqslant \mathbb{P}\left(n(\bar{y} - A^\tau\boldsymbol{\mu})^\tau S_y^{-1}(\bar{y} - A^\tau\boldsymbol{\mu}) \leqslant k_\alpha\right) \overset{(1)}{=} \beta$$

Dakle, sljedeći simultani *Roy-Boseovi* intervali su $\beta * 100\%$ pouzdani

$$\bar{y}_i - \sqrt{\frac{k_\alpha}{n}(a_i^\tau S a_i)} \leqslant A_i^\tau\mu \leqslant \bar{y}_i + \sqrt{\frac{k_\alpha}{n}(a_i^\tau S a_i)}, i = 1,\ldots,r.$$

Uspoređujući c_α i k_α vidimo da vrijedi $k_\alpha \leq c_\alpha$, pa je bolje koristiti ocjenu k_α. Budući da vrijedi

$$sup_{a\neq 0}\frac{(a^\tau\bar{x} - a^\tau\boldsymbol{\mu})^2}{a^\tau S a} = (\bar{x} - \boldsymbol{\mu})^\tau S^{-1}(\bar{x} - \boldsymbol{\mu})$$

na analogan način kao u izvodu *Roy-Boseovih* intervala koristeći (1) imamo $\beta * 100\%$ simultane pouzdane intervale koje zovemo *Scheffeovim* intervalima:

$$a^T\bar{x} - \frac{c_\alpha^{1/2}}{n^{1/2}}(a^T S a)^{1/2} \leqslant a^T\boldsymbol{\mu} \leqslant a^T\bar{x} + \frac{c_\alpha^{1/2}}{n^{1/2}}(a^T S a)^{1/2}, \forall a \in \mathbb{R}^p$$

Iako su *Scheffeovi* intervali širi od *Roy-Boseovih* korisni su jer pomoću njih možemo raditi veliki broj neplaniranih usporedbi između komponenata vektora očekivanja.

Ako je r, veličina vektora $\boldsymbol{\mu}$, mali broj, možemo dodatno poboljšati ocjenu konstante k_α. Neka je

$$T_i = \sqrt{n}\frac{\bar{y}_i - A_i^\tau\boldsymbol{\mu}}{\sqrt{a_i^\tau S a_i}}, i = 1,\ldots,r.$$

Primijetimo da je tada $T_i \overset{D}{=} t_{n-1}, i = 1,\ldots,r$. Onda, ako označimo događaj $A_i = \{|T_i| \leqslant t_0\}$ vrijedi:

$$\mathbb{P}(|T_i| \leqslant t_0, i = 1,\ldots,r) = \mathbb{P}(\cap_{i=1}^r A_i) = 1 - \mathbb{P}(\cup_{i=1}^r A_i^c) \geqslant$$
$$\geqslant 1 - \sum_{i=1}^r \mathbb{P}(A_i^c) = 1 - r\mathbb{P}(|t_{n-1}| > t_0)$$

Gornja nejednakost se zove *Bonferronijeva* nejednakost. Ako tu nejednakost izjednačimo s $\beta = 1 - \alpha$ i riješimo po t_0, dobivamo $\mathbb{P}(t_{n-1} > t_0) = \frac{\alpha}{2r}$ što je ekvivalentno $t_0 = t_{\frac{\alpha}{2r},n-1}$. Ako stavimo $b_\alpha = {t_{\frac{\alpha}{2r},n-1}}^2$ i umjesto *Roy-Boseovih* intervala izračunamo *Bonferronijeve* intervale

$$\bar{y}_i - \sqrt{\frac{b_\alpha}{n}\left(a_i^\tau S a_i\right)} \leqslant A_i^\tau \boldsymbol{\mu} \leqslant \bar{y}_i + \sqrt{\frac{b_\alpha}{n}\left(a_i^\tau S a_i\right)}, i = 1, \dots, r$$

koji su barem $\beta * 100\%$ pouzdani. Vidimo da je odnos duljina *Bonferronijevih* i *Roy-Boseovih* intervala točno $\sqrt{b_\alpha / k_\alpha}$, stoga u praksi koristimo one koji su kraći. Za podatke koji nisu normalno distribuirani, ali imaju konačno očekivanje i kovarijacijsku matricu, se na sličan način mogu izvesti simultani pouzdani intervali korištenjem centralnog graničnog teorema.

3 Izvod simultanih pouzdanih intervala

Želimo odrediti simultane pouzdane intervale za komponente očekivanje za svaku od tri skupine štakora opisanih u uvodu (svaka skupina čini jedan slučajan uzorak čiji su vektori dimenzije 5). Pretpostavljamo da su uzorci iz višedimenzionalne normalne razdiobe. Zanimaju nas simultani pouzdani intervali za komponente očekivanja, a ne za linearne jednakosti vezane uz njih, pa imamo da je matrica *A* iz gore navedene teorije jednaka jediničnoj matrici $A = I_5$. Kako $r = 5$ vidimo da je $k_\alpha = c_\alpha$, što znači da su *Roy-Boseovi* i *Scheffeovi* intervali jednaki. Trebamo provjeriti koji je od brojeva k_α i b_α manji kako bi mogli izabrati hoćemo li uzeti *Roy-Boseove* ili *Bonferronijeve* intervale. Kako je prva skupina duljine 7, a druge dvije su duljine 10 morati ćemo napraviti dvije provjere. Uzeti ćemo $\alpha = 0.05$. Račune ćemo provesti u R-u:

```
> k_alpha=function(n,alpha)
 { return ((n-1)*5/(n-5)*qf(1-alpha,df1=5,df2=n-5)); }
> b_alpha=function(n,alpha)
{ return ((qt(1-alpha/10,df=n-1))^2); }
> k1=k_alpha(7,0.05)
> b1=b_alpha(7,0.05)
> k1
[1] 289.4461
> b1
[1] 13.74502
> k2=k_alpha(10,0.05)
> b2=b_alpha(10,0.05)
> k2
[1] 45.45296
> b2
[1] 10.56143
```

Vidimo da su *Bonferronijeve* konstante manje pa ćemo u nastavku računati *Bonferronijeve* simultane pouzdane intervale.

Prvo ćemo učitati podatke iz Tablica 1,2 i 3.

```
>tiroksin=matrix(c(59,85,121,156,191,54,71,90,110,138,56,75,108,151,18
9,59,85,116,148,177,57,72,97,120,144,52,73,97,116,140,52,70,105,138,1
71),nrow=7,byrow=T)

> tiroksin

     [,1] [,2] [,3] [,4]  [,5]
[1,]   59   85  121  156  191
[2,]   54   71   90  110  138
[3,]   56   75  108  151  189
[4,]   59   85  116  148  177
[5,]   57   72   97  120  144
[6,]   52   73   97  116  140
[7,]   52   70  105  138  171
```

Na analogan način unosimo podatke za druge dvije skupine.

```
> tiouracil

      [,1] [,2] [,3] [,4]  [,5]
[1,]    61   86  109  120  129
[2,]    59   80  101  111  122
[3,]    53   79  100  106  133
[4,]    59   88  100  111  122
[5,]    51   75  101  123  140
[6,]    51   75   92  100  119
[7,]    56   78   95  103  108
[8,]    58   69   93  114  138
[9,]    46   61   78   90  107
[10,]   53   72   89  104  122
```

```
> kontrolna

      [,1] [,2] [,3] [,4]  [,5]
[1,]    57   86  114  139  172
[2,]    60   93  123  146  177
[3,]    52   77  111  144  185
[4,]    49   67  100  129  164
[5,]    56   81  104  121  151
[6,]    46   70  102  131  153
[7,]    51   71   94  110  141
[8,]    63   91  112  130  154
[9,]    49   67   90  112  140
[10,]   57   82  110  139  169
```

Sada ćemo izračunati vektor očekivanja za svaku od skupina štakora po komponentama vremena. Na analogan način kako je izračunato za prvu skupinu, tako je i za preostale dvije skupine.

```
>ocek_tiroksin=c(mean(tiroksin[,1]),mean(tiroksin[,2]),mean(tiroksin[,3]),mean(tiroksin[,4]),
mean(tiroksin[,5]))

> ocek_tiroksin

[1]  55.57143  75.85714 104.85714 134.14286 164.28571

> ocek_tiouracil

[1]  54.7  76.3  95.8 108.2 124.0

> ocek_kontrolna

[1]  54.0  78.5 106.0 130.1 160.6
```

Sada računamo uzoračke kovarijacijske matrice za svaku skupinu štakora. Na analogan način kako je izračunato za prvu skupinu, tako je i za preostale dvije skupine.

```
> kovarijacijska_tiroksin=cov(tiroksin, y=tiroksin, use="all.obs")

> kovarijacijska_tiroksin

          [,1]      [,2]      [,3]      [,4]      [,5]

[1,]  8.952381 15.92857  22.42857  33.07143  36.80952

[2,] 15.928571 41.47619  61.80952  85.52381  95.88095

[3,] 22.428571 61.80952 123.14286 195.52381 232.04762

[4,] 33.071429 85.52381 195.52381 346.80952 427.45238

[5,] 36.809524 95.88095 232.04762 427.45238 537.23810

> kovarijacijska_tiouracil

         [,1]     [,2]     [,3]     [,4]      [,5]

[1,] 22.01111 27.87778 29.71111 28.40000  15.00000

[2,] 27.87778 62.67778 58.51111 43.15556  18.55556

[3,] 29.71111 58.51111 72.17778 69.04444  53.00000

[4,] 28.40000 43.15556 69.04444 95.06667  87.55556

[5,] 15.00000 18.55556 53.00000 87.55556 126.66667
```

```
> kovarijacijska_kontrolna

         [,1]     [,2]      [,3]      [,4]      [,5]

[1,] 29.55556 49.55556  39.55556  27.88889  26.22222

[2,] 49.55556 92.94444  84.44444  71.94444  68.77778

[3,] 39.55556 84.44444  98.44444 110.88889 120.00000

[4,] 27.88889 71.94444 110.88889 157.87778 180.04444

[5,] 26.22222 68.77778 120.00000 180.04444 230.93333
```

Sada računamo granice simultanih pouzdanih intervala za svaku skupinu na tri razine značajnosti ($\alpha = 0.1, 0.05, 0.01$) te ćemo ih prikazati u obliku matrice kojoj stupci predstavljaju vremena, prvi redak donju granicu intervala i drugi redak gornju granicu intervala. Na analogan način kako je izračunato za prvu skupinu, tako je i za preostale dvije skupine.

Granice za razinu značajnosti $\alpha = 0.05$:

```
> n_tiroksin=length(tiroksin[,1])
> donje_granice_tiroksin5=c(0,0,0,0,0)
> gornje_granice_tiroksin5=c(0,0,0,0,0)
> alpha=0.05
> for(i in 1:5)
 { donje_granice_tiroksin5[i]=ocek_tiroksin[i]-
sqrt(b_alpha(n_tiroksin,alpha)/n_tiroksin*t(A[,i])%*%kovarijacijska_tiroksin%*%A[,i])
gornje_granice_tiroksin5[i]=ocek_tiroksin[i]+sqrt(b_alpha(n_tiroksin,alpha)/n_tiroksin*t(A[,i])%*%
kovarijacijska_tiroksin%*%A[,i]) }
> granice_tiroksin5=matrix(c(donje_granice_tiroksin5,gornje_granice_tiroksin5),nrow=2,byrow=T)
> granice_tiroksin5
        [,1]     [,2]      [,3]     [,4]     [,5]
[1,] 51.37874 66.83264  89.30722 108.0471 131.8064
[2,] 59.76412 84.88164 120.40707 160.2386 196.7650
> granice_tiouracil5
        [,1]     [,2]      [,3]     [,4]     [,5]
[1,] 49.8785 68.16386  87.06902  98.17982 112.4338
[2,] 59.5215 84.43614 104.53098 118.22018 135.5662
> granice_kontrolna5
        [,1]     [,2]      [,3]     [,4]     [,5]
[1,] 48.41297 68.59229  95.80336 117.1872 144.9827
[2,] 59.58703 88.40771 116.19664 143.0128 176.2173
```

Također, na analogan način kako je izračunato za razinu značajnosti $\alpha = 0.05$, tako je i za preostale dvije razine značajnosti ($\alpha = 0.1,\ 0.01$).

Granice za razinu značajnosti $\alpha = 0.1$:

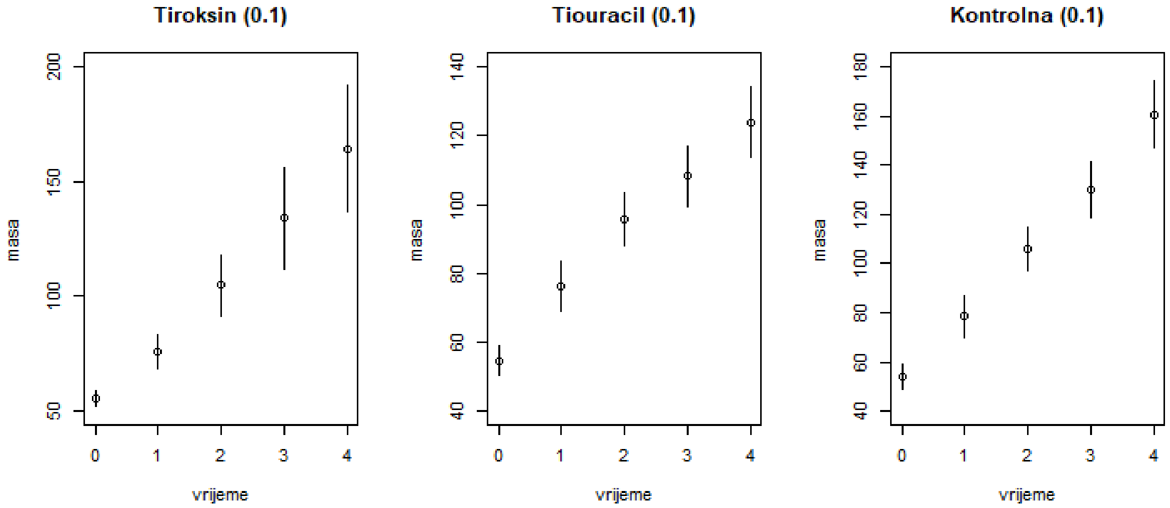
```
> granice_tiroksin10
         [,1]     [,2]      [,3]     [,4]     [,5]
[1,] 52.01742 68.20736  91.67597 112.0224 136.7540
[2,] 59.12544 83.50692 118.03832 156.2634 191.8174
> granice_tiouracil10
         [,1]     [,2]      [,3]     [,4]     [,5]
[1,] 50.51407 69.23638  88.21995  99.50069 113.9584
[2,] 58.88593 83.36362 103.38005 116.89931 134.0416
> granice_kontrolna10
         [,1]     [,2]      [,3]     [,4]     [,5]
[1,] 49.14946 69.89834  97.1475 118.8893 147.0414
[2,] 58.85054 87.10166 114.8525 141.3107 174.1586
```

Granice za razinu značajnosti $\alpha = 0.01$:

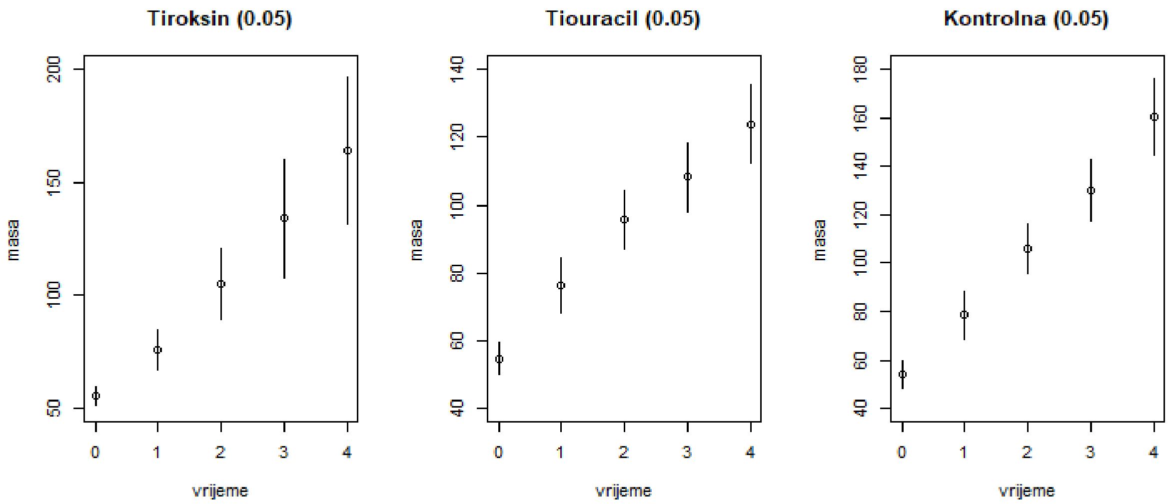
```
> granice_tiroksin1
         [,1]     [,2]      [,3]     [,4]     [,5]
[1,] 49.68218 63.18091  83.01499  97.4876 118.6637
[2,] 61.46068 88.53338 126.69929 170.7981 209.9077
> granice_tiouracil1
         [,1]     [,2]      [,3]     [,4]     [,5]
[1,] 48.3252 65.54272  84.25624  94.95171 108.7076
[2,] 61.0748 87.05728 107.34376 121.44829 139.2924
> granice_kontrolna1
         [,1]     [,2]      [,3]     [,4]     [,5]
[1,] 46.61305 65.40042  92.5184 113.0271 139.9515
[2,] 61.38695 91.59958 119.4816 147.1729 181.2485
```

Možemo primijetiti da postoje razlike u pouzdanim intervalima ovisno o razini značajnosti, no svi intervali sadrže pripadnu komponentu očekivanja skupine po vremenima, što na sljedećim grafovima (Slika 1,2 i 3) i vidimo. Na grafovima možemo još vidjeti kako se rasponi intervala mijenjaju promjenom razine značajnosti.

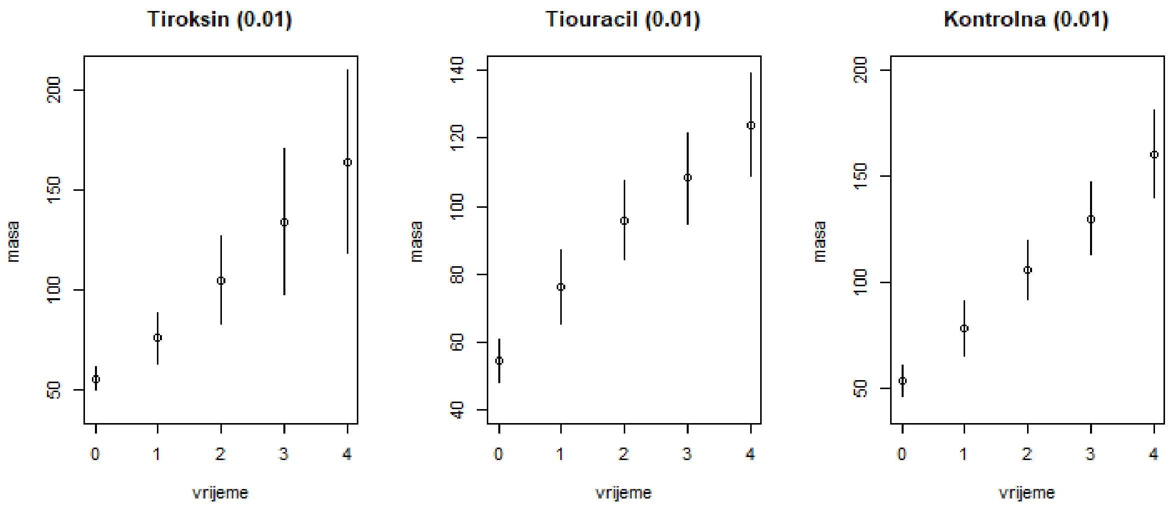
Slika 1: Prikaz simultanih 90% pouzdanih intervala s očekivanjem



Slika 2: Prikaz simultanih 95% pouzdanih intervala s očekivanjem



Slika 3: Prikaz simultanih 99% pouzdanih intervala s očekivanjem



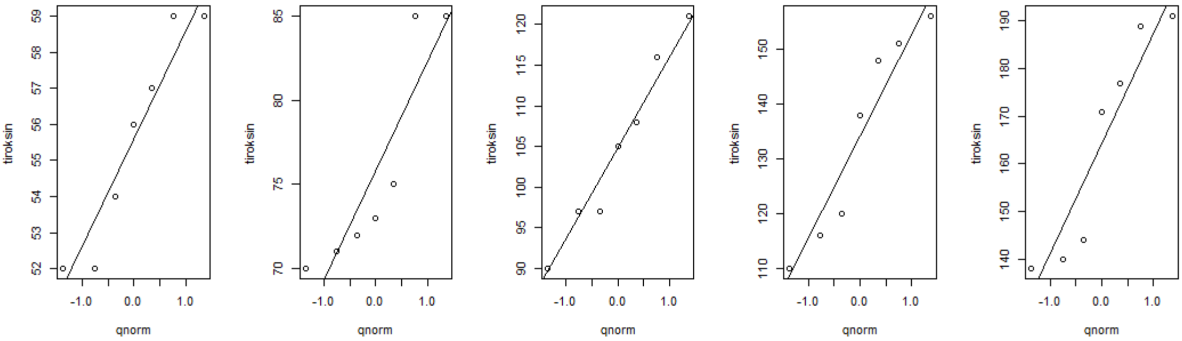
3 Ispitivanje normalnosti podataka

Podaci nad kojima provodimo test su podaci o tri skupine štakora u pet različitih vremena svako udaljeno tjedan dana. Svaka skupina je slučajni uzorak čiji su vektori dimenzije 5, te želimo testirati je li svaki od slučajnih uzoraka normalno distribuiran. Dakle, želimo testirati je li svaki stupac iz svake od te tri skupine dolazi iz normalne distribucije, odnosno nulhipoteza će biti da dolaze iz normalne distribucije, a alternativna da ne dolaze iz normalne distribucije. Kako ne postoji jedinstven način za testiranje takvih nulhipoteza, testirati ćemo pomoću normalnog vjerojatnosnog grafa i Lillieforsove inačice Kolmogorov-Smirnovljevog testa. Račune ćemo provesti u R-u.

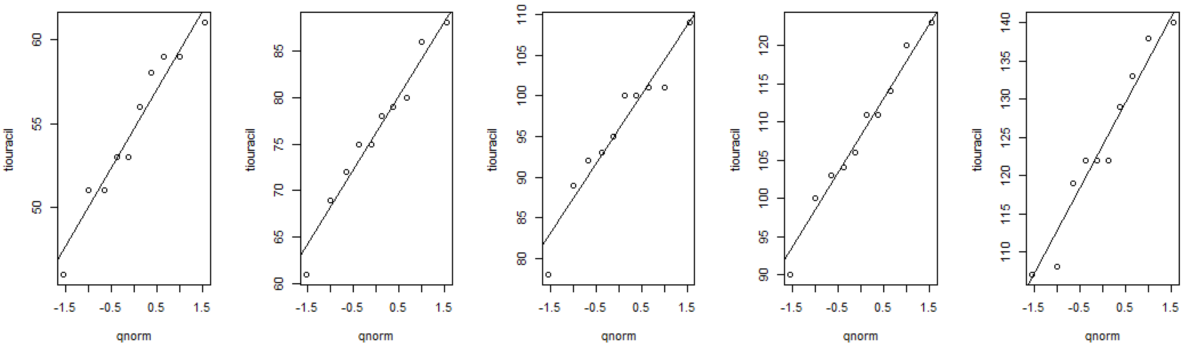
```
par(mfrow=c(1,5))
for(i in 1:5)
{
  plot(qnorm(((1:n_tiroksin)-3/8)/(n_tiroksin+1/4)),sort(tiroksin[,i]),xlab="qnorm",ylab="tiroksin")
  abline(mean(tiroksin[,i]),sd(tiroksin[,i]))
}
for(i in 1:5)
{
  plot(qnorm(((1:n_tiouracil)-3/8)/(n_tiouracil+1/4)),sort(tiouracil[,i]),xlab="qnorm",ylab="tiouracil")
  abline(mean(tiouracil[,i]),sd(tiouracil[,i]))
}
for(i in 1:5)
{
  plot(qnorm(((1:n_kontrolna)-3/8)/(n_kontrolna+1/4)),sort(kontrolna[,i]),xlab="qnorm",ylab="kontrolna")
  abline(mean(kontrolna[,i]),sd(kontrolna[,i]))
}
```

Pomoću gore navedenog koda dobili smo sljedeće normalne vjerojatnosne grafove:

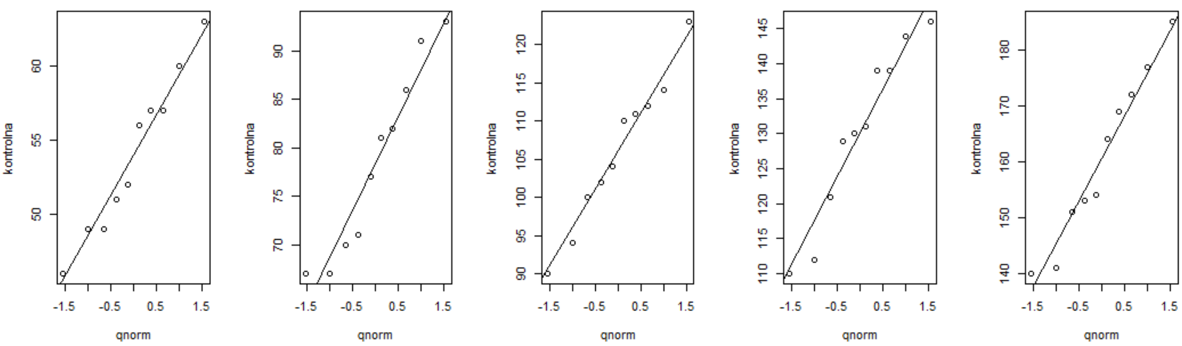
Slika 4: Normalni vjerojatnosni grafovi za skupinu tiroksin po vremenima



Slika 5: Normalni vjerojatnosni grafovi za skupinu tiouracil po vremenima



Slika 6: Normalni vjerojatnosni grafovi za kontrolnu skupinu po vremenima



Vidimo na Slici 4 da se podaci za skupinu tiroksin baš i ne grupiraju oko pravca pa možemo pretpostaviti da podaci ne dolaze iz normalne distribucije. Iz Slike 5 vidimo da se podaci za skupinu tiouracil malo bolje grupiraju oko pravca, ali i dalje ima odstupanja. Također, iz Slike 6 vidimo da se podaci za kontrolnu skupinu uglavnom grupiraju oko pravca, ali ima nekih odmaka od pravaca. Iz normalnih vjerojatnosnih grafova ne možemo zaključiti ništa konkretno, pa ćemo provesti Lillieforsovu inačicu Kolmogorov-Smirnovljevog testa kako bi došli do konkretnijeg zaključka.

```
> lillie.test(tiroksin[,1])

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  tiroksin[, 1]
D = 0.1694, p-value = 0.7868
```

Navedeni kod je primjer testiranja pripadnosti podataka skupine tiroksina za vrijeme 0 normalnoj distribuciji. Analogno radimo za sve tri skupine po svim vremenima i dobijemo sljedeće *p*-vrijednosti:

Tablica 4: *p*-vrijednosti Lillieforsove inačice KS-testa za svaku skupinu po vremenima

vrijeme	TIROKSIN	TIOURACIL	KONTROLNA
0	0.7868	0.6719	0.8096
1	0.1363	0.8732	0.4592
2	0.6303	0.3916	0.6955
3	0.5021	0.9735	0.6145
4	0.2696	0.5629	0.5872

Vidimo da nijedna *p*-vrijednost nije manja od razine značajnosti $\alpha = 0.05$, pa na razini značajnosti 5% **ne odbacujemo** hipotezu da svaki stupac u svakoj skupini dolazi iz normalne distribucije. Dakle, na razini značajnosti 5% zaključujemo da su podaci normalno distribuirani.

4 Hottelingov T^2-test za jedan uzorak

Želimo testirati je li istinita tvrdnja da je vektor očekivanja za kontrolnu skupinu jednak $\boldsymbol{\mu} = (60, 80, 100, 120, 140)^\tau$. Stavimo $\boldsymbol{\mu_0} = (60, 80, 100, 120, 140)^\tau$ i $\alpha = 0.05$. Iz dijela *3 Ispitivanje normalnosti podataka* zaključili smo da možemo pretpostaviti na razini značajnosti 5% da su podaci normalno distribuirani. Testiramo:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$$

$$H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu_0}$$

gdje je $\boldsymbol{\mu}$ vektor očekivanja za kontrolnu skupinu. Hipoteze ćemo testirati pomoću Hottelingovog T^2-testa, pa ćemo koristiti rezultat (1):

$$\mathbb{P}(n(\overline{\mathbb{X}} - \boldsymbol{\mu})^\tau S^{-1}(\overline{\mathbb{X}} - \boldsymbol{\mu}) \leq c_\alpha) = \beta.$$

Izračunat ćemo testnu statistiku $T^2 = n(\overline{\mathbb{X}} - \boldsymbol{\mu})^\tau S^{-1}(\overline{\mathbb{X}} - \boldsymbol{\mu}) \sim F(p, n-p)$ za kontrolnu skupinu, gdje je S uzoračka kovarijacijska matrica uzorka kontrolne skupine, i provjeriti je li se nalazi u kritičnom području $[c_\alpha, \infty\rangle$. Ako se nalazi unutar kritičnog područja, odbacujemo hipotezu H_0 u korist H_1 na razini značajnosti 5%. Također ćemo izračunati p-vrijednost testne statistike T^2 koristeći F-distribuciju, te ako je manja od 0.05 odbacujemo hipotezu H_0 u korist H_1 na razini značajnosti 5%. Račun ćemo provesti u R-u.

```
> n=n_kontrolna

> alpha=0.05

> c_alpha=k_alpha(n_kontrolna,alpha)

> c_alpha

[1] 45.45296

> mu0=c(60,80,100,120,140)

> ocek_kontrolna

[1]  54.0  78.5 106.0 130.1 160.6

> p=5

> n

[1] 10

> p

[1] 5

> T_2=n*t(ocek_kontrolna-mu0)%*%solve(kovarijacijska_kontrolna)%*%(ocek_kontrolna-mu0)

> T_2

      [,1]

[1,] 178.2226
```

Vidimo iz gornjih rezultata da je kritično područje $[45.45296\,,\infty\rangle$, a testna statistika jednaka je $T^2 = 178.2226$ što upada u kritično područje, pa na razini značajnosti 5% možemo **odbaciti** hipotezu H_0 u korist H_1. Sada ćemo izračunati *p*-vrijednost:

```
> pvrijednost=1-pf(T_2,df1=p,df2=n-p)

> pvrijednost

         [,1]

[1,] 1.255786e-05
```

Vidimo da je *p*-vrijednost = 1.255786e-05 što je manje od 0.05, dodatno manja je od bilo koje standardne razine značajnosti, pa na svakoj standardnoj razini značajnosti **odbacujemo** hipotezu H_0 u korist H_1, odnosno hipotezu da je vektor očekivanja kontrolne skupine jednak vektoru $(60, 80, 100, 120, 140)^\tau$.