

Primijenjena statistika

Seminarski rad - Analiza profila

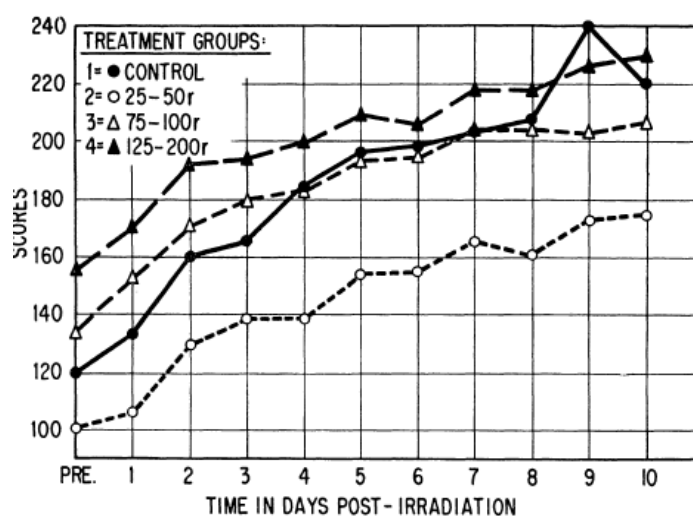
Hermina Petric Maretić

Podaci prikazuju testiranja 45 pacijenata koji pate od kancerogenih lezija na psihomotornom uređaju. Svi pacijenti su prošli početno treniranje na uređaju te je tada, od 45 pacijenata, 39 bilo izloženo različitim količinama radijacije, dok ih je 6 bilo kontrolno. Točan broj pacijenata u svakoj skupini, kao i količina radijacije kojoj su bili izloženi su prikazani u Tablici 1. Nakon početnog treniranja, performanse svakog pacijenta na uređaju mjerene su 4 puta dnevno, i to neposredno prije zračenja i svakodnevno kroz 10 dana nakon zračenja. Za svakog pacijenta i za svaki dan uzeto je usrednjenje 4 dnevnih mjerenja kao podatak za analizu. Ti su podaci prikazani u Tablici 2. Usrednjenja za svaku grupu za svaki dan su prikazana na Slici 1. Cilj eksperimenta bilo je određivanje utjecaja zračenja na psihomotorne performanse i veze između količine zračenja i performansi.

NUMBER OF SUBJECTS WITH AVERAGE AGE BY TREATMENT GROUP

Treatment group	No. of subjects	Average age
control	6	64.5
25-50 r	14	67.4
75-100 r	15	60.2
125-200 r	10	57.3
	45 (Total)	62.4 (average)

Tablica 1.



Slika 1.

AVERAGE DAILY SCORE (4 TRIALS/DAY) FOR EACH SUBJECT
ON PSYCHOMOTOR TESTING DEVICE

	Time (Days Post Irradiation)										
	Pre	1	2	3	4	5	6	7	8	9	10
<i>Controls</i>											
1	191	223	242	248	266	274	272	279	286	287	286
2	64	72	81	66	92	114	126	123	134	148	140
3	206	172	214	239	265	265	262	274	258	288	289
4	155	171	191	203	219	237	237	220	252	260	245
5	85	138	204	213	224	247	246	259	255	374	284
6	15	22	24	24	38	41	46	62	62	79	74
<i>25-50r</i>											
7	53	53	102	104	105	125	122	150	93	127	132
8	33	45	50	54	44	47	45	61	50	60	52
9	16	47	45	34	37	61	51	28	43	40	45
10	121	167	188	209	224	229	230	269	264	249	268
11	179	193	206	210	221	234	224	255	246	225	229
12	114	91	154	152	155	174	196	207	208	229	173
13	92	115	133	136	148	159	146	180	148	168	169
14	84	32	97	86	47	87	103	124	110	162	187
15	30	38	37	40	48	61	64	65	83	91	90
16	51	66	131	148	181	172	195	170	158	203	215
17	188	210	221	251	256	268	260	281	286	290	296
18	137	167	172	212	168	213	190	196	211	213	224
19	108	23	18	30	29	40	57	37	47	56	55
20	205	234	260	269	274	282	282	290	298	304	308
<i>75-100r</i>											
21	181	206	199	237	219	237	232	251	247	254	250
22	178	208	222	237	255	253	254	276	254	267	275
23	190	224	224	261	249	291	293	294	295	299	305
24	127	119	149	196	203	211	207	241	220	188	219
25	94	144	169	164	182	189	188	164	181	142	152
26	148	170	202	181	184	186	207	184	195	168	163
27	99	93	122	145	130	167	153	165	144	156	167
28	207	237	243	281	273	281	279	294	307	305	305
29	188	208	235	249	265	271	263	272	285	283	290
30	140	187	199	205	231	227	228	246	245	263	262
31	109	95	102	96	135	135	111	146	131	162	171
32	69	46	67	28	43	55	55	77	73	76	76
33	69	95	137	99	95	108	129	134	133	131	91
34	51	59	76	101	72	72	107	91	128	120	133
35	156	186	198	201	205	210	217	217	219	223	229
<i>125-250r</i>											
36	201	202	229	232	224	237	217	268	244	275	246
37	113	126	159	157	137	160	162	171	167	165	185
38	86	54	75	75	71	130	157	142	173	174	156
39	115	158	168	175	188	164	184	195	194	206	212
40	183	175	217	235	241	251	229	241	233	233	275
41	131	147	183	181	206	215	197	207	226	244	240
42	71	105	107	92	101	103	78	87	57	70	71
43	172	213	263	260	276	273	267	286	283	290	298
44	224	258	248	257	257	267	260	279	299	289	300
45	246	257	269	270	289	291	306	301	295	312	311

Tablica 2.

Analiza profila sastoji se od tri glavna dijela:

- Test paralelnosti krivulja
- Test prosječne razine krivulja
- Test vodoravnosti prosječne krivulje

Zadnja dva dijela su od posebne važnosti ako vrijedi prvi dio, jer očito vrijede sljedeći rezultati:

- Krivulje su paralelne i imaju jednaku prosječnu razinu ako i samo ako su krivulje identične
- Krivulje su paralelne i prosječna krivulja je vodoravna ako i samo ako su sve krivulje vodoravne

Definiramo matricu $\Lambda'_r = [J_{r-1}, -I_{r-1}]$, gdje J_{r-1} označava vektor jedinica dimenzije $r - 1$, a I_{r-1} jediničnu matricu dimenzija $(r - 1) \times (r - 1)$. Primijetimo da vrijedi:

$$\Lambda'_a B = \begin{bmatrix} \mu'_1 - \mu'_2 \\ \mu'_1 - \mu'_3 \\ \vdots \\ \mu'_1 - \mu'_a \end{bmatrix}$$

gdje je $\mu'_i = [\mu_{i1}, \dots, \mu_{iq}]$

Zato je hipoteza $H_0: \Lambda_a B = 0$ jednaka upravo hipotezi $H_0: \mu_1 = \mu_2 = \dots = \mu_a$.

Prvo unesemo podatke po grupama:

```
gr1=load('grupa1.txt');  
gr2=load('grupa2.txt');  
gr3=load('grupa3.txt');  
gr4=load('grupa4.txt');
```

Računamo matricu B koja je usrednjenje po krivuljama unutar grupa:

```
B(1,:)=mean(gr1,1);  
B(2,:)=mean(gr2,1);  
B(3,:)=mean(gr3,1);  
B(4,:)=mean(gr4,1);
```

```
>> B
```

```
B =
```

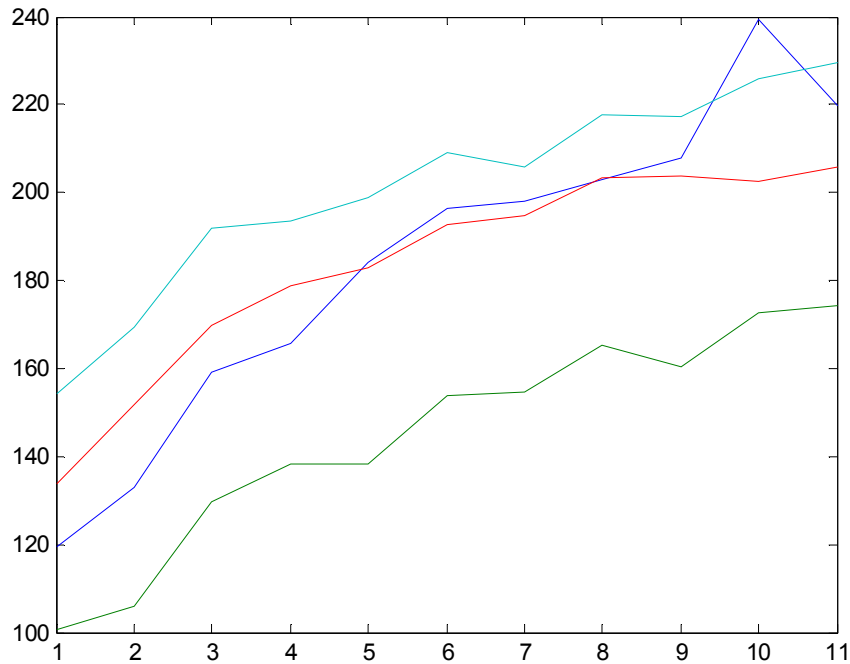
```
Columns 1 through 8
```

119.3333	133.0000	159.3333	165.5000	184.0000	196.3333	198.1667	202.8333
100.7857	105.7857	129.5714	138.2143	138.3571	153.7143	154.6429	165.2143
133.7333	151.8000	169.6000	178.7333	182.7333	192.8667	194.8667	203.4667
154.2000	169.5000	191.8000	193.4000	199.0000	209.1000	205.7000	217.7000

Columns 9 through 11

207.8333	239.3333	219.6667
160.3571	172.6429	174.5000
203.8000	202.4667	205.8667
217.1000	225.8000	229.4000

plot(B')



Računamo matrice transformacija koje ćemo koristiti:

```
lambda4(1,:) = ones(1,3);  
lambda4(2:4,:) = -eye(3);  
  
lambda11(1,:) = ones(1,10);  
lambda11(2:11,:) = -eye(10);  
  
J11 = ones(11,1);  
J4 = ones(4,1);
```

Računamo kovarijacijske matrice i matrice S i E, te ih onda transformiramo:

```
s1 = cov(gr1);  
s2 = cov(gr2);  
s3 = cov(gr3);  
s4 = cov(gr4);  
  
e = 5*s1 + 13*s2 + 14*s3 + 9*s4;  
s = e/41;
```

i matricu X koja zapravo predstavlja matricu $X'X$, gdje je X blok matrica s jediničnim matricama na dijagonali, pa kad ih pomnožimo dobivamo:

```
X = diag([6,14,15,10]);
```

Test paralelnosti krivulja

Hipoteza paralelnosti profila može se zapisati kao:

$$\mu_{i1} - \mu_{i'1} = \mu_{i2} - \mu_{i'2} = \dots = \mu_{iq} - \mu_{i'q} \quad \forall i, i', \text{ ili ekvivalentno}$$

$$\mu_{11} - \mu_{i1} = \dots = \mu_{1q} - \mu_{iq} \quad \forall i = 2, \dots, a, \text{ odnosno}$$

$$(\mu_{11} - \mu_{i1}) - (\mu_{12} - \mu_{i2}) = 0, \dots, (\mu_{11} - \mu_{i1}) - (\mu_{1q} - \mu_{iq}) = 0 \quad \forall i = 2, \dots, a$$

Sada vidimo da su profili paralelni ako i samo ako je $\Lambda'_a B \Lambda_q = 0$. Sada možemo postaviti test paralelnosti profila:

$$H_0: \Lambda'_a B \Lambda_q = 0 \text{ versus } H_1: \Lambda'_a B \Lambda_q \neq 0$$

Uz hipoteznu statistiku:

$$H_* = P = (\Lambda'_a \hat{B} \Lambda_q)' [\Lambda'_a (X'X)^{-1} \Lambda_a]^{-1} (\Lambda'_a \hat{B} \Lambda_q)$$

i pogreškovnu statistiku:

$$E_* = G = \Lambda'_q E \Lambda_q$$

Uz istinitost H_0 $\Lambda = \frac{\det(G)}{\det(P+G)}$ (tj. Wilkis' Lambda) ima distribuciju

$$\Lambda \sim U(q-1, a-1, n-a)$$

Računamo hipoteznu i pogreškovnu statistiku:

```
H=(lambda4'*B*lambda11) '*inv(lambda4'*inv(X)*lambda4) '*lambda4'*B*lambda11;  
E1=lambda11'*e*lambda11;
```

Ispišimo, kao primjer, hipoteznu i pogreškovnu statistiku:

H =

1.0e+004 *

0.1335	0.0838	0.0688	0.1224	0.0665	0.0646	0.0495	0.1063	-0.0003	0.0051
0.0838	0.0762	0.0480	0.1312	0.0957	0.0877	0.0691	0.1237	0.1493	0.0877
0.0688	0.0480	0.0599	0.1188	0.0952	0.1133	0.0832	0.1280	0.1187	0.0626
0.1224	0.1312	0.1188	0.3218	0.2797	0.3012	0.2275	0.3454	0.5111	0.2802
0.0665	0.0957	0.0952	0.2797	0.2623	0.2892	0.2174	0.3138	0.5289	0.2864
0.0646	0.0877	0.1133	0.3012	0.2892	0.3361	0.2492	0.3496	0.5627	0.2992
0.0495	0.0691	0.0832	0.2275	0.2174	0.2492	0.1855	0.2619	0.4280	0.2287

0.1063	0.1237	0.1280	0.3454	0.3138	0.3496	0.2619	0.3835	0.5918	0.3200
-0.0003	0.1493	0.1187	0.5111	0.5289	0.5627	0.4280	0.5918	1.2780	0.6957
0.0051	0.0877	0.0626	0.2802	0.2864	0.2992	0.2287	0.3200	0.6957	0.3805

E1 =

1.0e+004 *

3.2684	2.9417	3.1477	3.4296	2.9759	2.6052	2.6897	2.8459	2.6277	2.3969
2.9417	4.3338	4.2682	4.6122	4.1976	4.0900	4.2460	3.9621	4.5699	3.9433
3.1477	4.2682	5.6600	5.6784	5.3498	5.0082	5.3659	5.1430	5.6540	5.5019
3.4296	4.6122	5.6784	6.8902	5.9192	5.4687	5.8880	5.4488	5.9533	5.8396
2.9759	4.1976	5.3498	5.9192	5.9068	5.4040	5.7152	5.4652	6.0558	5.7597
2.6052	4.0900	5.0082	5.4687	5.4040	5.7851	5.5275	5.6598	6.2341	5.6868
2.6897	4.2460	5.3659	5.8880	5.7152	5.5275	6.6678	5.9591	6.8981	6.3590
2.8459	3.9621	5.1430	5.4488	5.4652	5.6598	5.9591	6.6359	6.9004	6.2848
2.6277	4.5699	5.6540	5.9533	6.0558	6.2341	6.8981	6.9004	9.3191	7.7850
2.3969	3.9433	5.5019	5.8396	5.7597	5.6868	6.3590	6.2848	7.7850	7.9652

Računamo U vrijednost:

```
hip=det(E1+H);
pog=det(E1);
rez=pog/hip;
```

Provodimo X2 test:

```
v = 10*(4-1);
X2 = (-1)*((44)-(5*(10+4)))*log(rez);
P = 1-chi2cdf(X2,v)
```

P =

0.8306

P vrijednost je vrlo velika pa ne odbacujemo nultu hipotezu H_0 u korist prve hipoteze, odnosno ne odbacujemo hipotezu da su krivulje paralelne na razini značajnosti od 5 %, što smo iz grafa mogli i očekivati.

Test prosječne razine krivulja

Prosječnu razinu krivulje definiramo kao prosjek po svim različitim zavisnim varijablama. Računamo prosječne vrijednosti profila:

$$\bar{\mu}_{i\cdot} = \frac{1}{q} \sum_{k=1}^q \mu_{ik}$$

Hipoteza da je $\bar{\mu}_{1\cdot} = \dots = \bar{\mu}_{a\cdot}$ može se zapisati u obliku

$$\bar{\mu}_{1\cdot} - \bar{\mu}_{2\cdot} = 0, \dots, \bar{\mu}_{1\cdot} - \bar{\mu}_{a\cdot} = 0$$

Ekvivalentnu hipotezu dobit ćemo ako promatramo sume bez usrednjavanja

$$\mu_{i\cdot} = \mu'_i J_q = \sum_{k=1}^q \mu_{ik}$$

Sada možemo postaviti test prosječne razine krivulja uz hipoteze oblika:

$$H_0: \Lambda_a^T B J_q = 0 \text{ versus } H_1: \Lambda_a^T B J_q \neq 0$$

Gdje je hipotezna statistika:

$$H_* = P = (\Lambda_a^T \hat{B} J_q)' [\Lambda_a^T (X'X)^{-1} \Lambda_a]^{-1} (\Lambda_a^T \hat{B} J_q)$$

i pogreškovna statistika:

$$E_* = G = J_q^T E J_q$$

Računamo hipoteznu i pogreškovnu statistiku:

```
H2=(lambda4'*B*J11) '*inv(lambda4'*inv(X)*lambda4) '*lambda4'*B*J11;  
E2=J11'*e*J11;
```

Testiranje se svodi na univarijantni test. Provodimo ANOVU na modificiranim podacima. Za ovaj dio testiranja korišten je gotov kod koji je javno dostupan na službenim MATLAB-ovim stranicama¹.

Rješenje:

Analysis of Variance Summary Table for Test of Levels Effect.

SOV	SS	df	MS	F	P
Between Groups	244142.864	3	81380.955	1.550	0.2141
Within Groups	2468204.975	47	52514.999		
Total	2712347.839	50			

¹ <http://www.mathworks.com/matlabcentral/fileexchange/13483>

P vrijednost je velika, pa ne odbacujemo hipotezu H_0 na razini značajnosti 5 %. Drugim riječima, ne odbacujemo hipotezu o jednakosti prosječnih razina krivulja.

Računamo U vrijednost:

```
hip2=det (E2+H2) ;  
pog2=det (E2) ;  
rez2=pog2/hip2;
```

Provodimo X2 test:

```
v = 3;  
X22 = (-1)*((44)-(5*(10+4)))*log(rez2);  
P2 = 1-chi2cdf(X22,v)
```

P2 =

0.3434

P vrijednost je malo drugačija, ali dolazimo do istog zaključka. Ne odbacujemo hipotezu H_0 na razini značajnosti 5%.

Test vodoravnosti prosječne krivulje

Prosječnom krivuljom smatramo:

$$\bar{\mu}^T = \frac{1}{a} J_a^T B = \frac{1}{a} \sum_{i=1}^a \mu_i^T. \quad \bar{\mu}^T = (\bar{\mu}_{\cdot 1}, \dots, \bar{\mu}_{\cdot q})$$

Testiranje vodoravnosti te krivulje svodi se na testiranje sljedećih jednakosti:

$$\bar{\mu}_{\cdot 1} - \bar{\mu}_{\cdot 2} = 0, \dots, \bar{\mu}_{\cdot 1} - \bar{\mu}_{\cdot q} = 0.$$

Kao i u testiranju prosječne razine krivulja, problem možemo promatrati i bez usrednjenja, uz

$$\mu^T = J_a^T B$$

Sada možemo postaviti test vodoravnosti prosječne krivulje:

$$H_0: J_a^T B \Lambda_q = 0 \quad \text{versus} \quad H_1: J_a^T B \Lambda_q \neq 0$$

Uz hipoteznu statistiku:

$$H_* = P = (J_a^T \hat{B} \Lambda_q)^T [J_a^T (X'X)^{-1} J_a]^{-1} (J_a^T \hat{B} \Lambda_q)$$

I pogreškovnu statistiku:

$$E_* = G = \Lambda_q^T E \Lambda_q$$

Računamo hipoteznu i pogreškovnu statistiku:

```
H3=(J4'*B*lambda11) '*inv(J4'*inv(X)*J4) '*J4'*B*lambda11;  
E3=lambda11'*e*lambda11;
```

Računamo U vrijednost:

```
hip3=det(E3+H3);  
pog3=det(E3);  
rez3=pog3/hip3;
```

Provodimo X2 test:

```
v = 10*(4-1);  
X23 = (-1)*((44)-(.5*(10+4)))*log(rez3);  
P3 = 1-chi2cdf(X23,v)
```

P3 =

1.1690e-004

P vrijednost je vrlo mala, pa odbacujemo hipotezu H_0 na razini značajnosti 5 %. Drugim riječima, odbacujemo hipotezu o vodoravnosti prosječne krivulje, što smo i očekivali.