

Линеарни статистички модели
септембар 2017.

1. Посматрајмо модел линеарни модел $Y = X\beta + \varepsilon$ где је X $n \times (p + 1)$ дизајн матрица максималног ранга, $\beta = (\beta_0, \dots, \beta_p)$ вектор непознатих параметара, и ε има вишедимензиону нормалну $\mathcal{N}_n(0, \sigma^2 I)$ расподелу, при чему је σ^2 такође непознат параметар.

а) Одредити $\hat{\beta}$ и $\hat{\sigma}^2$ методом максималне веродостојности и вредност функције веродостојности у тим оценама. (2 + 1)

б) Извести израз за статистику теста количника веродостојности којим се тестира нулта хипотеза $\beta_1 + \beta_2 = 0$. Коју расподелу има тест статистика? На ком месту се користи Кохранова теорема? (4 + 2 + 4)

в) Како бисте на основу оцењеног модела предвидели вредност независне променљиве у тачки x_0 Y ? Коју расподелу има та оцена? Извести израз за $\beta\%$ интервал предвиђања. (2 + 2 + 3)

У циљу проучавања загађења ваздуха сакупљени су следећи подаци.

Y -сумпор диоксид у микрограмима по кубном метру

X_1 -просечна годишња температура у Фаренхајтима

X_2 -број фабрика који запошљавају више од 20 људи

X_3 -величина популације у хиљадама

X_4 -просечна годишња брзина ветра у миљама на сат

X_5 -просечна количина падавина у кубном инчу

X_6 -просечан број кишних дана годисње

```
model <- lm(log(Y) ~ X1 + X2 + X3 + X4 + X5 + X6)
summary(model)
```

Call:

```
lm(formula = log(Y) ~ X1 + X2 + X3 + X4 + X5 + X6)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	7.2532456	1.4483686	5.008	1.68e-05	***
X1	-0.0599017	0.0190138	-3.150	0.00339	**
X2	0.0012639	0.0004820	2.622	0.01298	*
X3	-0.0007077	0.0004632	-1.528	0.13580	
X4	-0.1697171	0.0555563	-3.055	0.00436	**
X5	0.0173723	0.0111036	1.565	0.12695	
X6	0.0004347	0.0049591	0.088	0.93066	

Residual standard error: 0.448 on 34 degrees of freedom

Multiple R-Squared: 0.6541

F-statistic: 10.72 on 6 and 34 degrees of freedom, p-value: 1.126e-06

г) На који тест се односи вредност F статистике? Шта се на основу тога може закључити о моделу? (2 + 2)

д) Које бисте променљиве избацили из модела? Одговор образложити. Узети у обзир Бонферонијеву корекцију. (4)

ђ) Претпоставите да сте избацили неке предикторе. Када поновите оцењивање да ли очекујете да оцене коефицијената остану непромењене? Одговор образложити. (4)

е) Шта је модификовани коефицијент детерминације и за шта се он користи? (2)

з) Приметите да је извршена трансформација зависне променљиве. Који су могући разлози за то? (4)

2. а) Објаснити LASSO регресију. Како се оцењују непознати параметри модела? Да ли је добијена оцена непристрасна? (2+3+2)

б) Да ли се ова регресија може користити за селекцију предиктора? Одговор образложити. (2)

3. а) Навести два модела која припадају класи уопштених линеарних модела која се могу применити у случају да Y узима вредности 0,1,2,... (2)

б) Које су канонске линк функције у тим моделима? (2)

в) Реализовано је истраживање о утицају поена из математике на финалном испиту (math) и врсте студија које је студент похађао (prog), опште (1), академске (2) и стручне (3), на број освојених награда у току школовања. Део података изгледа овако: Предложено је да се за моделовање зависности искористи следећи

	id	num_awards	prog	math	
	195	61	1	2	60
	196	100	2	2	71
	197	143	2	3	75
	198	68	1	2	71
	199	57	0	2	72
	200	132	3	2	73

модел.

```
m1 <- glm(num_awards ~ prog + math, family="poisson")
```

 Добија се:

	Естимате	Стд. Ерроp	з валуе	Пр(> з)	
(Интерцепт)	-5.5781	0.6768	-8.24	<2e-16	***
prog	0.1233	0.1633	0.76	0.4502	
math	0.0861	0.0096	8.98	<2e-16	***

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 287.67 on 199 degrees of freedom

Residual deviance: 203.45 on 197 degrees of freedom

- г) Написати једначину добијеног модела. (2)
- д) Шта можете да закључите о значајности коефицијената модела? Одговор образложити. (3)
- ђ) Изведите израз за девијацију Пуасоновог модела. (4)
- е) Како бисте оценили параметар расејања? Да ли оцена упућује на одступање од Пуасоновог модела? (3+2)
- ж) Како бисте оценили број награда за студента који је имао 72 поена на финалном испиту из математике и похађао академске студије? (3)