

UNIVERZA V LJUBLJANI, EKONOMSKA FAKULTETA

AKTUARSTVO

AKTUARSKA STATISTIKA

PISNI IZPIT

10. OKTOBER 2008

IME IN PRIIMEK: _____

V PISNA ŠT:

--	--	--	--	--	--	--	--

NAVODILA

Pazljivo preberite besedilo naloge, preden se lotite reševanja. Naloge so 4. Dovoljena sredstva sta dva A4 format lista in matematični priročnik. Vaše odgovore prosim napišite na priložene liste. Na razpolago imate 2 uri.

Naloga	a.	b.	c.	d.	
1.				•	
2.				•	
3.			•	•	
4.			•	•	
Skupaj			•	•	

1. (25) Privzemite, da je populacija N posameznikov razdeljena na K skupinic velikosti M , tako da velja $N = KM$. Vzorčimo tako, da z enostavnim slučajnim vzorčenjem izberemo $k \leq K$ skupinic brez ponavljanja, iz vsake skupinice pa kar vse enote. Oceniti želimo populacijsko povprečje μ dane spremenljivke. Označite

$$\sigma_b^2 = \frac{1}{K} \sum_{i=1}^K (\mu_i - \mu)^2.$$

- a. (5) Cenilka za μ naj bo povprečje vrednosti spremenljivke za vse enote izbrane v vzorec. Pokažite, da je cenilka nepristranska.

Namig: Cenilko \bar{Y} lahko zapišete kot

$$\bar{Y} = \frac{1}{k} \sum_{i=1}^K \mu_i I_i,$$

kjer je μ_i povprečje znotraj i -te skupinice in

$$I_i = \begin{cases} 1 & \text{če } i\text{-to skupnico izberemo,} \\ 0 & \text{če } i\text{-te skupinice ne izberemo.} \end{cases}$$

- b. (10) Privzemite, da je standardni odklon spremenljivke znotraj vsake skupinice enak σ_w . Kako bi ocenili populacijsko varianco σ^2 ? Je vaša ocena nepristranska?

- c. (5) Privzemite spet, da je standardni odklon znotraj vsake skupinice enak σ_w . Izrazite standardno napako cenilke iz točke a. s σ_w^2 in σ_b^2 .
za enote iz vzorca?

2. (20) Naj bodo X_1, X_2, \dots, X_n med sabo neodvisne opazovane vrednosti z inverzno Gaussovo porazdelitvijo, dano z gostoto

$$f(x) = \sqrt{\frac{\tau}{2\pi x^3}} \exp \left\{ -\frac{\tau}{2x\mu^2} (x - \mu)^2 \right\}, \quad x > 0, \tau > 0, \mu > 0.$$

Kot znano upoštevajte, da je $E(X_1) = \mu$. Privzemite tudi, da so izpolnjene vse predpostavke za uporabo asimptotskih rezultatov za cenilke po metodi največjega verjetja.

- a. (5) Poiščite cenilki za parametra μ in τ po metodi največjega verjetja.

- b. (10) Izračunajte Fisherjevo matriko informacije $I(\mu, \tau)$.

- c. (5) Zapišite interval zaupanja pri stopnji zaupanja $1 - \alpha = 0,95$ za parameter μ na osnovi podatkov x_1, x_2, \dots, x_n .

3. (25) Na razpolago imamo opazovane vrednosti X_1, X_2, \dots, X_n iz normalne porazdelitve $N(\mu, \sigma^2)$. Predpostavljamo, da so opazovane vrednosti med sabo neodvisne, enako porazdeljene slučajne spremenljivke. Preizkusiti želimo domnevo $H_0 : \mu = 0$ proti $H_1 : \mu \neq 0$. Privzemite, da je parameter σ^2 znan.

a. (15) Domnevo H_0 lahko preizkusimo pri dani stopnji tveganja α na dva načina:

- H_0 zavržemo, če je $|\bar{X}| > c$ za primerno izbran c , tako da bo tveganje ravno α .
- Iz opazovanih vrednosti ocenimo μ in izračunamo zgornjo in spodnjo mejo intervala zaupanja pri stopnji tveganja α . Če interval ne pokriva 0, domnevo H_0 zavržemo.

Ali sta zgornja testa popolnoma enaka? Komentirajte.

b. (10) Izračunajte še testno statistiko za zgornjo situacijo po metodi kvocienta verjetij.

4. (25) Dane imamo cene stanovanj v različnih četrtih v Bostonu in vrednosti spremenljivk, ki vplivajo na ceno stanovanj v tisti četrti. Imena spremenljivk so navedena spodaj.

Vir: Harrison, Rubinfeld, HEDONIC PRICES AND THE DEMAND FOR CLEAN AIR, Economics & Management, VOL.5, 81-102, 1978

VKLJUČENE SPREMENLJIVKE:

CRIM	Stopnja kriminala v mestu.
ZN	Delež stanovanjskih parcel preko 25000 kvadratnih čevljev.
INDUS	delež površine parcel podjetij, ki se ne ukvarjajo s trgovino.
CHAS	indikator bližine reke Charles.
NOX	Koncentracija dušikovega oksida.
RM	povprečno število sob na bivališče.
AGE	delež stanovanj, zgrajenih pred letom 1940 v privatni lasti.
DIS	ponderirana razdalja do petih bostonskih con zaposlovanja.
RAD	indeks dostopnosti do obvoznic
TAX	stopnja davka na lastnino na \$10.000.
PTRATIO	razmerje št. učencev–št. učiteljev v mestu.
B	1000*(bk-0.63), bk= delež črnega prebivalstva.
LSTAT	odstotek prebivalstva nižjega sloja.
MEDV	mediana vrednosti lastniških stanovanjskih enot v \$1000.

Program za regresijo po vrsti izbira spremenljivke, ki najbolj vplivajo na mediano vrednosti stanovanjskih enot. Izpis je naslednji:

Forward selection. Alpha-to-Enter: 0,01

Response is MEDV on 13 predictors, with N = 452
N(cases with missing observations) = 54 N(all cases) = 506

Step	1	2	3	4	5	6
Constant	-38,276	-10,916	7,023	13,659	6,916	16,229
RM	9,78	6,51	5,89	5,51	5,65	5,54
T-Value	23,35	14,08	13,13	12,41	12,93	12,79
P-Value	0,000	0,000	0,000	0,000	0,000	0,000
LSTAT		-0,582	-0,537	-0,644	-0,596	-0,552
T-Value		-11,61	-11,17	-12,44	-11,48	-10,41
P-Value		0,000	0,000	0,000	0,000	0,000
PTRATIO			-0,79	-0,83	-0,82	-0,87
T-Value			-7,01	-7,48	-7,57	-8,04
P-Value			0,000	0,000	0,000	0,000

DIS	-0,60	-0,65	-1,02
T-Value	-4,88	-5,44	-6,28
P-Value	0,000	0,000	0,000
B		0,0148	0,0121
T-Value		4,31	3,48
P-Value		0,000	0,001
NOX			-10,7
T-Value			-3,30
P-Value			0,001

S	5,93	5,21	4,95	4,83	4,73	4,68	4,63
R-Sq	54,79	65,23	68,67	70,25	71,44	72,12	72,80
R-Sq(adj)	54,69	65,08	68,46	69,99	71,12	71,75	72,37
C-p	315,2	140,9	84,9	60,1	42,1	32,6	23,2

a. (5) S katerimi količinami na izpisu bi utemeljili, da sta neodvisna spremenljivka LSTAT in mediana cen stanovanj MEDV med sabo povezani?

b. (5) Bi na podlagi izpisa sklepali, da vseh 6 izbranih neodvisnih spremenljivk vpliva na odvisno spremenljivko MEDV? Utemeljite zakaj.

- c. (5) V drugem poskusu izberemo le 4 neodvisne spremenljivke in izvedemo regresijsko analizo. Izpis je nasledni:

Regression Analysis: MEDV versus RM; LSTAT; PTRATIO; DIS

The regression equation is

$$\text{MEDV} = 13,7 + 5,51 \text{ RM} - 0,644 \text{ LSTAT} - 0,828 \text{ PTRATIO} - 0,595 \text{ DIS}$$

452 cases used 54 cases contain missing values

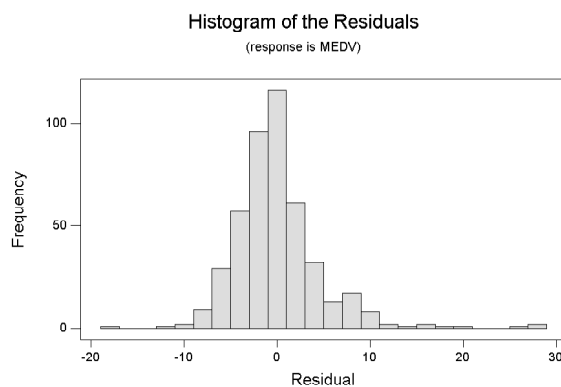
Predictor	Coef	SE Coef	T	P
Constant	13,659	4,192	3,26	0,001
RM	5,5141	0,4444	12,41	0,000
LSTAT	-0,64376	0,05173	-12,44	0,000
PTRATIO	-0,8283	0,1108	-7,48	0,000
DIS	-0,5953	0,1219	-4,88	0,000

S = 4,826 R-Sq = 70,3% R-Sq(adj) = 70,0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	24584,4	6146,1	263,93	0,000
Residual Error	447	10409,3	23,3		
Total	451	34993,7			

Histogram napak pri napovedih (residuals) je na spodnji sliki 1.



Slika 1 Histogram napak pri napovedih za dane podatke.

Ali znate z izpisa razbrati, kolikšen je standardni odklon histograma na Sliki 1?

- d. (10) Ali lahko trdimo, da vsaka dodatna soba v enoti v povprečju poveča ceno za približno \$5514? Utemeljite vaš odgovor!

