

DODATAK 1. METODE PROCJENJIVANJA

Metoda pomoćnih momenata

Definirali smo pomoćne momente populacije:

$$k\text{-ti pomoćni moment populacije: } m_k = \sum_{x \in D} x^k p(x) = E(X^k)$$

Možemo definirati k -ti pomoćni moment uzorka: $m_{ku} = \frac{1}{n} \sum_{i=1}^n x_i^k$

Metoda momenata sastoji se u izjednačavanju k -tog momenta uzorka s k -tim momentom populacije.

To smo malo prije i napravili kad smo izjednačili $\mu = \bar{x}$.

Primjer:

Ako imamo uzorak iz populacije za koju pretpostavljamo da je raspodijeljena prema općenitoj gama raspodjeli $f(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}$, ($x \geq 0$) (8. predavanje), želimo procijeniti parametre α i β . Znademo da je $E(X) = \alpha \cdot \beta$ i $E(X^2) = \beta^2(\alpha + 1)\alpha$.

Procjenjitelji parametara α i β su

$$\hat{\alpha} = \frac{\bar{X}^2}{(1/n) \sum X_i^2 - \bar{X}^2}$$

$$\hat{\beta} = \frac{(1/n) \sum X_i^2 - \bar{X}^2}{\bar{X}}$$

Procjenjitelj najvjerojatnije vrijednosti

Neka slučajne varijable X_1, X_2, \dots, X_n imaju združenu raspodjelu vjerojatnosti ili funkciju gustoće vjerojatnosti danu s

$$f(x_1, x_2, \dots, x_n; \vartheta_1, \vartheta_2, \dots, \vartheta_m),$$

gdje su parametri $\vartheta_1, \vartheta_2, \dots, \vartheta_m$ nepoznati. Ako su x_i -ovi opažene vrijednosti uzorka, onda je f funkcija parametara $\vartheta_1, \vartheta_2, \dots, \vartheta_m$. Procjena najvjerojatnijih vrijednosti tih parametara $\hat{\vartheta}_1, \hat{\vartheta}_2, \dots, \hat{\vartheta}_m$ je ona za koju vrijedi

$$f(x_1, x_2, \dots, x_n; \hat{\vartheta}_1, \hat{\vartheta}_2, \dots, \hat{\vartheta}_m) \geq f(x_1, x_2, \dots, x_n; \vartheta_1, \vartheta_2, \dots, \vartheta_m) \text{ za sve } \vartheta_1, \vartheta_2, \dots, \vartheta_m.$$

Ako x_i -ove zamijenimo X_i -ovima dobivamo **procjenjitelje najvjerojatnijih vrijednosti**.

Primjer:

Na 7. predavanju tražili smo najvjerojatniju vrijednost mjerene fizikalne veličine. Tada je združena funkcija vjerojatnosti za n mjerena bila:

$$f(x_1, x_2, \dots, x_n; \mu, \sigma^2) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{(x_1-\mu)^2 + (x_2-\mu)^2 + \dots + (x_n-\mu)^2}{2\sigma^2}}.$$

Tražili smo maksimum te funkcije tako da smo derivirali njezin logaritam i dobili smo princip najmanjih kvadrata. Ovaj procjenitelj za μ je nepristran. Međutim, procjenitelj najvjerojatnije vrijednosti varijance σ^2 je $\hat{\sigma}^2 = \frac{\sum(X_i - \bar{X})^2}{n}$, što je pristrani procjenitelj.

Svojstva procjenitelja najvjerojatnije vrijednosti

Procjenitelj najvjerojatnije vrijednosti ima dva vrlo poželjna svojstva:

1. Invarijantnost na matematičke operacije.

Procjenitelj najvjerojatnije vrijednosti funkcije parametara jednak je funkciji procjenitelja najvjerojatnijih vrijednosti tih parametara, tj. $\hat{h}(X) = h(\hat{X})$.

Sjetimo se da to svojstvo nema očekivanje.

2. Za uzorak s velikim n , procjenitelj najvjerojatnije vrijednosti je gotovo nepristran.

Želimo pokazati da je $E[h(X)] \approx h(E(X))$:

$$E(X) = \mu$$

Funkciju $h(X)$ razvijimo u Taylorov red:

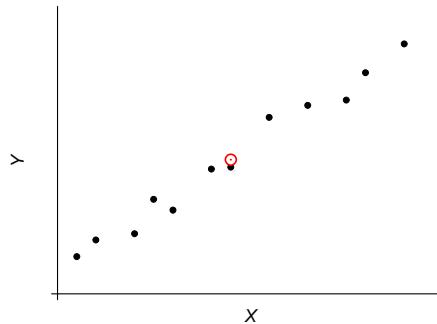
$$h(X) = h(\mu) + (X - \mu) \frac{dh}{dX} \Big|_{\mu} + \frac{1}{2} (X - \mu)^2 \frac{d^2 h}{dX^2} \Big|_{\mu} + \dots$$

$$E[h(X)] = E[h(\mu)] + E(X - \mu) \cdot \frac{dh}{dX} \Big|_{\mu} + \frac{1}{2} E[(X - \mu)^2] \frac{d^2 h}{dX^2} \Big|_{\mu} + \dots$$

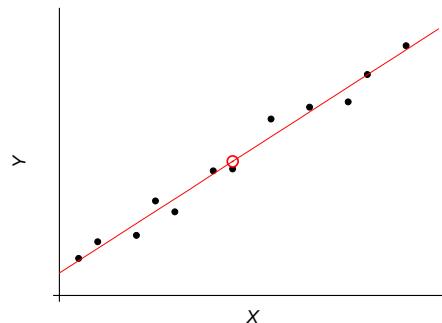
$$= h(\mu) + 0 + \frac{1}{2} \sigma^2 \frac{d^2 h}{dX^2} \Big|_{\mu} + \dots \approx h(\mu) \quad (\text{za male } \sigma)$$

DODATAK 2. Određivanje pravca regresije “od oka”

Najprije pronađemo vrijednosti $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ i $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ i ucrtamo točku (\bar{x}, \bar{y}) :



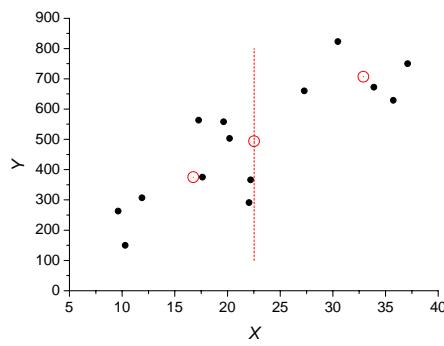
Kroz tu točku povučemo pravac za koji nam se čini da najbolje prolazi kroz podatke te očitamo nagib a i odsječak na ordinati b .



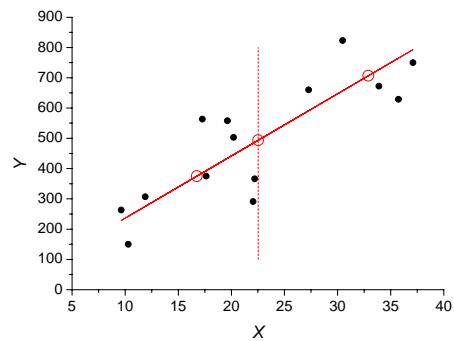
Ako je rasipanje veliko, najprije nacrtamo točku (\bar{x}, \bar{y}) . Zatim podijelimo podatke na dvije skupine – lijevo i desno od (\bar{x}, \bar{y}) . Izračunamo

$$\bar{x}_L = \frac{1}{n_L} \sum_{x_i < \bar{x}} x_i \quad \bar{y}_L = \frac{1}{n_L} \sum_{x_i < \bar{x}} y_i \quad \bar{x}_D = \frac{1}{n_D} \sum_{x_i > \bar{x}} x_i \quad \bar{y}_D = \frac{1}{n_D} \sum_{x_i > \bar{x}} y_i$$

i ucrtamo točke (\bar{x}_L, \bar{y}_L) i (\bar{x}_D, \bar{y}_D) :



Zatim povučemo pravac koji najbolje prolazi kroz te tri točke:



DODATAK 3.

TESTIRANJE DOBROTE PRILAGODBE

Zamislimo da imamo pokus koji nam kao rezultat daje niz opaženih frekvencija. Možemo pretpostaviti da neka teorijska raspodjela dobro opisuje opaženu raspodjelu frekvencija. Da bismo tu pretpostavku (hipotezu) provjerili, primjenjujemo jedan od statističkih testova prilagodbe. Najčešće se upotrebljava χ^2 -test. Najprije ćemo se podsetiti što znamo o χ^2 -raspodjeli:

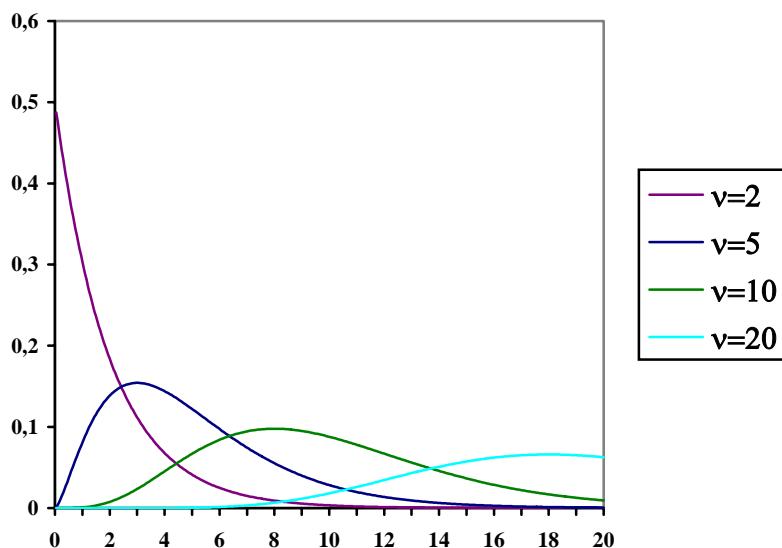
χ^2 raspodjela i tablice

Def:

Neka je $v \in N$. Kontinuirana slučajna varijabla X ima **χ^2 raspodjelu** s parametrom v ako je njezina funkcija gustoće vjerojatnosti dana s

$$f(x; v) = \chi_v^2(x) = \begin{cases} \frac{x^{v/2-1} e^{-x/2}}{2^{v/2} \Gamma(v/2)} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$$

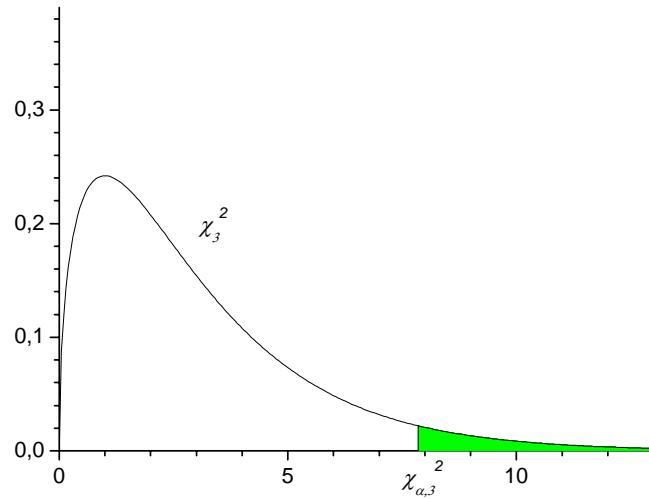
Parametar v zove se **“broj stupnjeva slobode”** varijable X .



Vjerojatnost da varijabla X poprimi vrijednost veću od neke određene vrijednosti $\chi_{\alpha, v}^2$ dana je površinom ispod repa krivulje:

$$P(X > \chi_{\alpha, v}^2) = \int_{\chi_{\alpha, v}^2}^{\infty} \chi_v^2(x) dx$$

Indeks α označava iznos te vjerojatnosti. Ako X poprimi vrijednost veću od $\chi_{\alpha,\nu}^2$, kažemo da je u **kritičnom području**.



Obično unaprijed odredimo koliku vjerojatnost α želimo pa na osnovu toga određujemo kritičnu vrijednost $\chi_{\alpha,\nu}^2$.

Te vjerojatnosti je teško izračunati pa se služimo tablicama.
U testiranju hipoteza najčešće nas zanimaju vjerojatnosti:

$$P(X > \chi_{0,05,\nu}^2) = 0,05$$

$$P(X > \chi_{0,01,\nu}^2) = 0,01$$

Za $\nu=3$, vrijednosti su:

$$\chi_{0.05,\nu}^2 = 7,82 \quad \chi_{0.01,\nu}^2 = 11,34$$

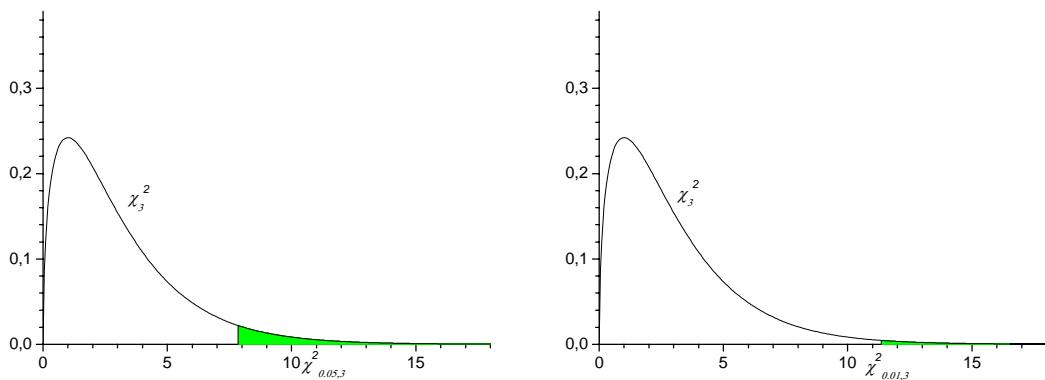
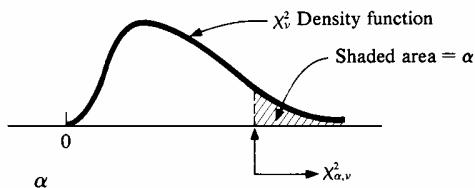


Table A.6 Critical Values $X^2_{\alpha,\nu}$ for the Chi-Squared Distribution



ν	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.843	5.025	6.637	7.882
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.344	12.837
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.085	16.748
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.440	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.012	18.474	20.276
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.534	20.090	21.954
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.022	21.665	23.587
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.724	26.755
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.735	27.687	29.817
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.600	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.577	32.799
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.407	7.564	8.682	10.085	24.769	27.587	30.190	33.408	35.716
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.843	7.632	8.906	10.117	11.651	27.203	30.143	32.852	36.190	38.580
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.033	8.897	10.283	11.591	13.240	29.615	32.670	35.478	38.930	41.399
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.195	11.688	13.090	14.848	32.007	35.172	38.075	41.637	44.179
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.519	11.523	13.120	14.611	16.473	34.381	37.652	40.646	44.313	46.925
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.807	12.878	14.573	16.151	18.114	36.741	40.113	43.194	46.962	49.642
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.120	14.256	16.147	17.708	19.768	39.087	42.557	45.772	49.586	52.333
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.457	15.655	17.538	19.280	21.433	41.422	44.985	48.231	52.190	55.000
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.814	17.073	19.046	20.866	23.110	43.745	47.400	50.724	54.774	57.646
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.191	18.508	20.569	22.465	24.796	46.059	49.802	53.203	57.340	60.272
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.584	19.960	22.105	24.075	26.492	48.363	52.192	55.667	59.891	62.880
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.994	21.425	23.654	25.695	28.196	50.660	54.572	58.119	62.426	65.473
40	20.706	22.164	24.433	26.509	29.050	51.805	55.758	59.342	63.691	66.766

$$\text{For } \nu > 40, x_{\alpha,\nu}^2 \doteq \nu \left(1 - \frac{2}{9\nu} + z_\alpha \sqrt{\frac{2}{9\nu}} \right)^3$$

χ^2 -test

Razmotrimo pokus koji daje n opaženih frekvencija f_i . Želimo li provjeriti hipotezu da ta opažanja slijede neku teorijsku raspodjelu, najprije izračunamo očekivane teorijske vrijednosti f_{ti} . Opažene frekvencije, naravno, odstupaju od teorijskih, a mi želimo donijeti odluku možemo li ta odstupanja pripisati slučaju. Nul-hipoteza je: "Opažanja slijede teorijsku raspodjelu".

Definiramo statistiku:

$$\chi_{\text{op}}^2 = \sum_{i=1}^n \frac{(f_i - f_{ti})^2}{f_{ti}}$$

Postoji teorem (K. Pearson) koji kaže:

Statistika χ_{op}^2 približno je raspodijeljena prema χ^2 raspodjeli sa stupnjem slobode ν koji ovisi o broju opažanja (ili razreda) i o broju postavljenih ograničenja.

Broj stupnjeva slobode ν

Parametar ν definiran je kao:

$$\nu = \text{broj nezavisnih varijabli uključenih u izračun } \chi_{\text{op}}^2$$

Nalazimo ga na sljedeći način:

$$\nu = \text{broj razreda} - \text{broj ograničenja}$$

Napomene:

- Uzimanjem χ^2 raspodjele, diskretnu raspodjelu aproksimiramo kontinuiranom. Ta aproksimacija ne vrijedi ako je očekivana frekvencija manja od 5. Taj problem prevladavamo grupiranjem nekoliko razreda male frekvencije u jedan razred dovoljno velike frekvencije.
- Ako je vrijednost χ_{op}^2 vrlo mala, mudro je upitati se nisu li opaženi rezultati namješteni.

Dobrota prilagodbe

Binomna raspodjela, p je poznat

Primjer:

Bacamo četiri novčića 160 puta. Ako su novčići pošteni, očekujemo bimomnu raspodjelu $X \sim \text{Bin}(4, 1/2)$. Opažene su frekvencije:

x	0	1	2	3	4	Total
f(x)	5	35	67	41	12	160

Provjeri s 5% signifikantnosti jesu li novčići pošteni!

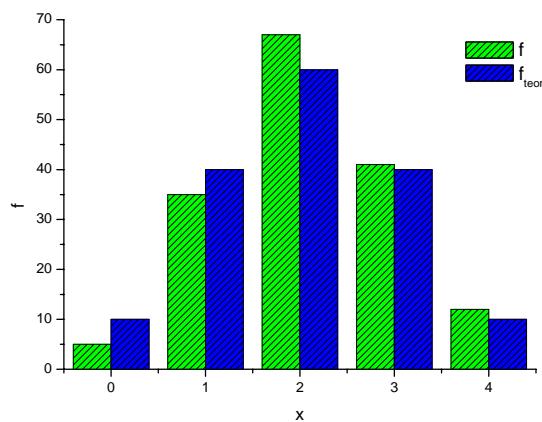
Rješenje: H_0 : Novčići su pošteni, tj. $p=1/2$

Teorijska raspodjela je binomna.

Imamo jedno ograničenje: $\sum f_i = \sum f_{ti} = 160$

Dakle, $v=5-1=4$

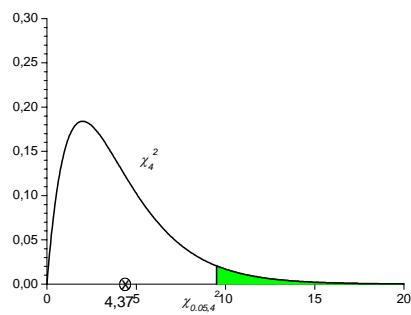
x	f_i	f_{ti}	$(f_i - f_{ti})^2 / f_{ti}$
0	5	10	2,5
1	35	40	0,625
2	67	60	0,817
3	41	40	0,025
4	12	10	0,4
zbroj	160	160	4,367



Za 4 stupnja slobode i 5% signifikantnosti, kritična vrijednost (iz tablica) je

$$\chi^2_{0.05,4} = 9,49$$

Izračunata vrijednost $\chi^2_{\text{op}} = 4,37$ ne pada u kritično područje (vidi sliku). Stoga hipotezu H_0 zadržavamo.



Primjeri testova dobrote prilagodbe.

Pravokutna raspodjela

Kockica se baca 120 puta i dobivaju se rezultati:

x	1	2	3	4	5	6	Total
f(x)	14	16	24	22	24	20	120

Provjeri s 5% signifikantnosti je li kocka poštena!

Rješenje: H_0 : Kocka je poštena

Teorijska raspodjela je pravokutna: $f_{ti}=20$

Imamo jedno ograničenje: $\sum f_i = \sum f_{ti} = 120$

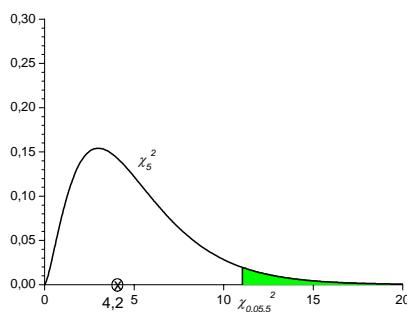
Dakle, $v=6-1=5$

x	f_i	f_{ti}	$(f_i - f_{ti})^2 / f_{ti}$
1	14	20	1,8
2	16	20	0,8
3	24	20	0,8
4	22	20	0,2
5	24	20	0,8
6	20	20	0
zbroj	120	120	4,2

Za 5 stupnjeva slobode i 5% signifikantnosti, kritična vrijednost (iz tablica) je

$$\chi^2_{0,05,5} = 11,07$$

Izračunata vrijednost $\chi^2_{\text{op}} = 4,2$ ne pada u kritično područje (vidi sliku). Stoga hipotezu H_0 zadržavamo.



Binomna raspodjela, p je nepoznat

Primjer s četvrtog predavanja:

Imamo velik broj proizvoda i uzimamo slučajni uzorak od $n = 20$ komada. Pokus ponavljamo $N = 100$ puta. Zanima nas kolika je vjerojatnost škarta (neispravnog proizvoda) i da li je raspodjela binomna. Opažene su frekvencije:

\underline{x}	0	1	2	3	4	5	6	Total
$f(x)$	14	25	27	23	7	3	1	100

Provjeri s 5% signifikantnosti je li raspodjela binomna!

Rješenje: H_0 : Raspodjela je binomna; $p = \frac{\bar{x}}{n}$.

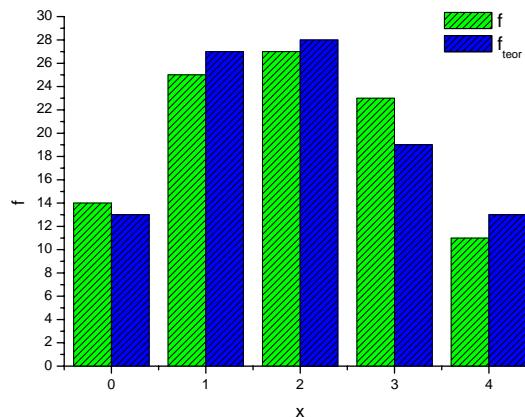
Izračunali smo na četvrtom predavanju $\bar{x} = 1,97$; $p = 0,0985$

Imamo dva ograničenja: $\sum f_i = \sum f_{ti} = N = 100$ i $np = \bar{x}$

Razredi 5 i 6 su premali pa grupiramo u jedan razred $x \geq 4$.

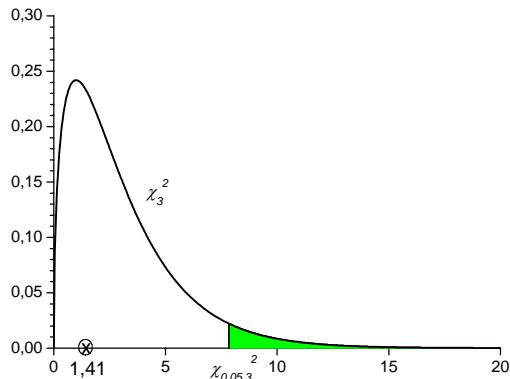
Dakle, imamo pet razreda i dva ograničenja pa je $v=5-2=3$

x	f_i	f_{ti}	$(f_i - f_{ti})^2 / f_{ti}$
0	14	13	1/13
1	25	27	4/27
2	27	28	1/28
3	23	19	16/19
≥ 4	11	13	4/13
zbroj	100	100	1,41



Za 3 stupnja slobode i 5% signifikantnosti, kritična vrijednost (iz tablica) je $\chi^2_{0.05,3} = 7.82$

Izračunata vrijednost $\chi^2_{\text{op}} = 1.41$ ne pada u kritično područje (vidi sliku). Stoga hipotezu H_0 zadržavamo.



Poissonova raspodjela

Bilježen je broj poziva u privatni stan u jednom danu u periodu od $N = 150$ dana.
Opažene su frekvencije:

X	0	1	2	3	4	Total
f(x)	51	54	36	6	3	150

Provjeri s 5% signifikantnosti je li raspodjela Poissonova!

Rješenje: H_0 : Raspodjela je Poissonova; $\lambda = \bar{x}$.

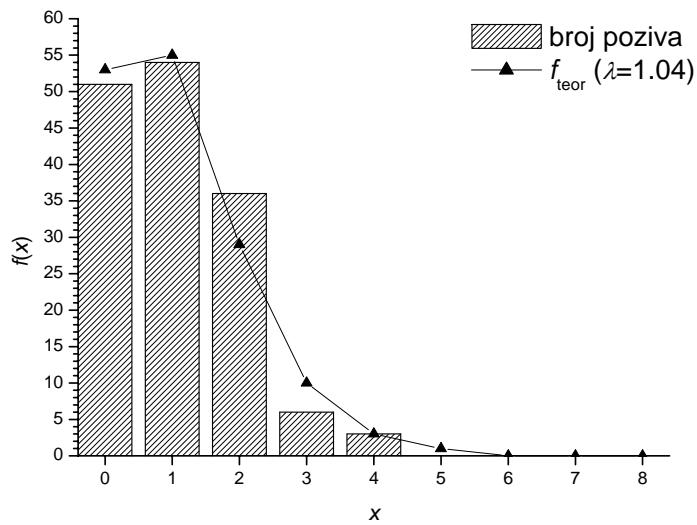
Izračunali smo na petom predavanju $\bar{x} = 1,04$; $\lambda = 1,04$

Imamo dva ograničenja: $\sum f_i = \sum f_{ti} = N = 150$ i $\lambda = \bar{x}$

Razred 4 je preveliki pa grupiramo u jedan razred $x \geq 3$.

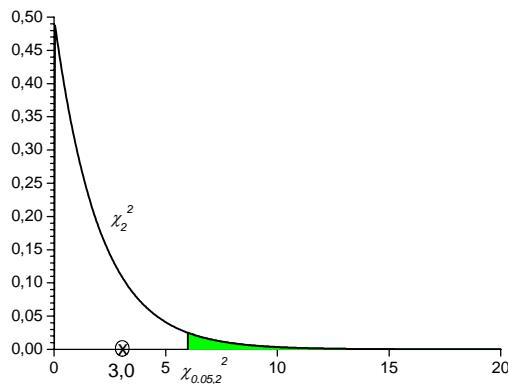
Dakle, imamo četiri razreda i dva ograničenja pa je $v=4-2=2$

x	f_i	f_{ti}	$(f_i - f_{ti})^2 / f_{ti}$
0	51	53	4/53
1	54	55	1/55
2	36	29	49/29
≥ 3	9	13	16/13
zbroj	150	150	3,01



Za 2 stupnja slobode i 5% signifikantnosti, kritična vrijednost (iz tablica) je $\chi^2_{0.05,2} = 5.99$

Izračunata vrijednost $\chi^2_{\text{op}} = 3.01$ ne pada u kritično područje (vidi sliku). Stoga hipotezu H_0 zadržavamo.



Normalna raspodjela, μ i σ poznati

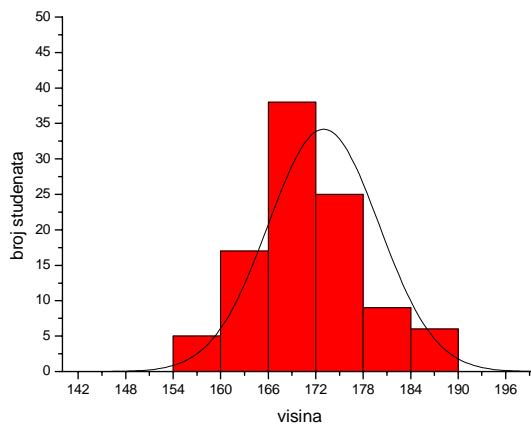
Primjer:

Dugogodišnje statistike pokazuju da je visina studenata na nekom sveučilištu normalno raspodijeljena s očekivanjem $\mu = 173$ cm i standardnom devijacijom $\sigma = 7$ cm. Iz jedne generacije studenata na tom sveučilištu izdvojeno je 100 studenata, mjerene su njihove visine i svrstane u razrede. Opažene su frekvencije:

<u>x</u>	154-160	160-166	166-172	172-178	178-184	184-190	Total
<u>f(x)</u>	5	17	38	25	9	6	100

Provjeri s 5% signifikantnosti je li raspodjela normalna s s očekivanjem $\mu = 173$ cm i standardnom devijacijom $\sigma = 7$ cm !

Rješenje: H_0 : *Raspodjela je Gaussova s očekivanjem $\mu = 173$ cm i standardnom devijacijom $\sigma = 7$ cm.*



Imamo jedno ograničenje: $\sum f_i = \sum f_{ti} = N = 100$

Uvodimo standardnu normalnu slučajnu varijablu $z_i = \frac{x - \mu}{\sigma}$. Za te

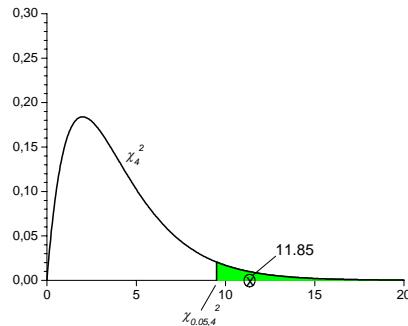
vrijednosti z_i određujemo iz tablica funkciju raspodjele $F(z)$ i vjerojatnosti razreda $f(z)$:

razred i	f_i	z_i	$F(z)$	$f(z)$	f_{ti}	$(f_i f_{ti})^2 / f_{ti}$
154-160	5	$z < -1,86$	0,0314	0,0314	3	
160-166	17	$-1,86 < z < -1$	0,1587	0,1273	13	2,25
166-172	38	$-1 < z < -0,14$	0,4443	0,2856	28	3,57
172-178	25	$-0,14 < z < 0,71$	0,7611	0,3168	32	1,53
178-184	9	$0,71 < z < 1,57$	0,9418	0,1807	18	4,5
184-190	6	$z > 1,57$	1	0,0582	6	0
zbroj	100				100	11,85

Dakle, imamo pet razreda i jedno ograničenje pa je $v=5-1=4$

Za 4 stupnja slobode i 5% signifikantnosti, kritična vrijednost (iz tablica) je $\chi^2_{0,05,4} = 9.49$.

Izračunata vrijednost $\chi^2_{\text{op}} = 11.85$ pada u kritično područje (vidi sliku). Stoga hipotezu H_0 odbacujemo.



Normalna raspodjela, μ i σ nepoznati

Isti primjer:

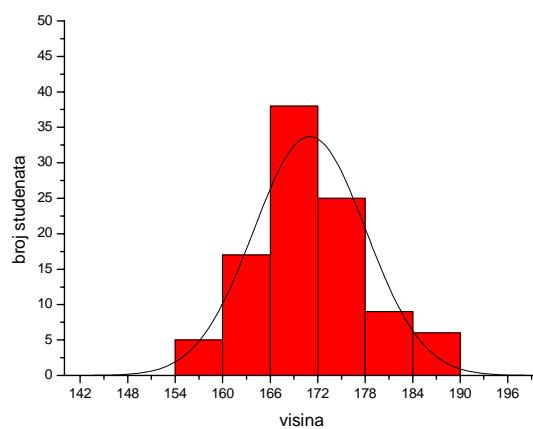
Odbacili smo hipotezu o očekivanju i standardnoj devijaciji, ali i dalje vjerujemo da je raspodjela normalna.

Provjeri s 5% signifikantnosti je li raspodjela normalna!

Rješenje: H_0 : *Raspodjela je Gaussova s očekivanjem $\mu = \bar{x}_{\text{uzorka}}$ i standardnom devijacijom $\sigma = \sigma_{\text{uzorka}}$.*

Izračunamo:

$$\bar{x}_{\text{uzorka}} = \frac{1}{N} \sum f_i x_i = 171 \text{ cm} \quad \sigma_{\text{uzorka}} = \sqrt{\frac{1}{N} \sum f_i (x_i - \bar{x}_{\text{uzorka}})^2} = 7,1 \text{ cm}$$



razred i	f_i	z_i	$F(z)$	$f(z)$	f_{ti}	$(f_i - f_{ti})^2 / f_{ti}$
154-160	5	$z < -1,55$	0,0606	0,060	6	0,167
160-166	17	$-1,55 < z < -0,70$	0,2420	0,181	18	0,056

166-172	38	-0,70<z<0,14	0,5557	0,314	32	1,125
172-178	25	0,14<z<0,99	0,8389	0,283	28	0,321
178-184	9	0,99<z<1,83	0,9664	0,128	13	0,063
184-190	6	z>1,83	1	0,033	3	
zbroj	100			1	100	1,732

Dakle, imamo pet razreda i tri ograničenja: $N = 100$

$$\mu = \bar{x}_{\text{uzorka}}$$

$$\sigma = \sigma_{\text{uzorka}}.$$

pa je $v=5-3=2$

Za 2 stupnja slobode i 5% signifikantnosti, kritična vrijednost (iz tablica) je $\chi^2_{0.05,2} = 5,99$.

Izračunata vrijednost $\chi^2_{\text{op}} = 1,73$ ne pada u kritično područje (vidi sliku). Stoga hipotezu H_0 zadržavamo.

