Poissonova distribucija se primjenjuje na događaje za koje vrijedi:

- Događaji se mogu brojati nenegativnim brojem
- Događaji su međusobno nezavisni
- Prosječan broj nastupa događaja u danom vremenskom periodu je poznat i konstantan
- Moguće je odrediti broj nastupa, ali je besmisleno pitati koliko se puta događaj *nije* dogodio (npr. broj kupaca u nekom periodu).

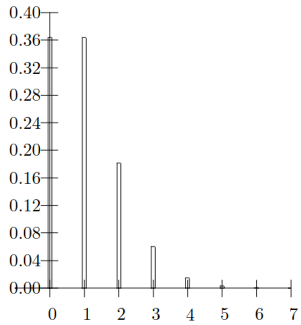
Poissonova distribucija određena je parametrom $\lambda \geq 0$, čije značenje je očekivani broj događanja tijekom zadanog vremenskog intervala (npr. 10 događanja u 4 minute $\rightarrow \lambda = \frac{10}{4} = 2.5$)
Slučajna varijabla koja ima Poissonovu distribuciju se označava sa $X \sim \text{Poi}(\lambda)$:

$$\text{Poi}(\lambda) = \begin{pmatrix} 0 & 1 & \dots & k & \dots \\ p_0 & p_1 & \dots & p_k & \dots \end{pmatrix}$$

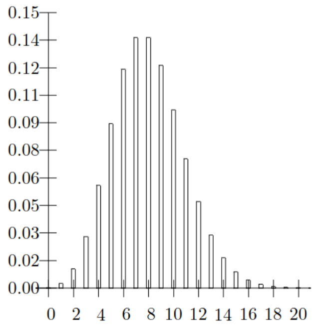
Vjerojatnost p_k da slučajna varijabla X poprimi vrijednost k:

$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

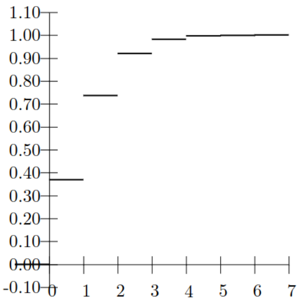
$$E(X) = \lambda, \quad \text{Var}(X) = \lambda, \quad \sigma(X) = \sqrt{\lambda}$$



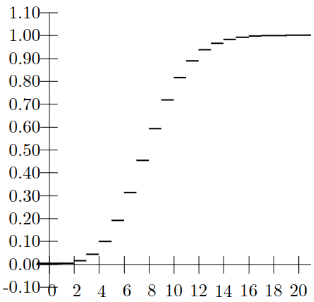
funkcija vjerojatnosti
Poissonove distribucije za $\lambda = 1$



funkcija vjerojatnosti
Poissonove distribucije za $\lambda = 8$



funkcija distribucije vjerojatnosti
Poissonove distribucije za $\lambda = 1$



funkcija distribucije vjerojatnosti
Poissonove distribucije za $\lambda = 8$

Kontinuirana slučajna varijabla se koristi kada ishodi slučaja nisu izolirani brojevi (npr. koliko sati žarulja izdrži prije nego što pregori). U ovim slučajevima se koristi kontinuirana slučajna varijabla koja može poprimiti vrijednosti $< -\infty, +\infty >$ (tj. u slučaju sa žaruljom, $[0, +\infty >$)

Vjerojatnost svakog elementarnog događaja kod kontinuirane slučajne varijable je 0. Ima smisla tražiti jedino vjerojatnost da varijabla poprimi vrijednost u nekom intervalu.

Funkcija gustoće vjerojatnosti kontinuriane slučajne varijable je funkcija $f : \mathbb{R} \rightarrow \mathbb{R}$ sa svojstvima:

- f je nenegativna
- $\int_a^b f(t)dt = P(a \leq X \leq b)$
- $\int_{-\infty}^{+\infty} = 1$

Zbog toga što je vjerojatnost da X poprimi neku konkretnu vrijednost, sve sljedeće vjerojatnosti su jednake:

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

Funkcija distribucije vjerojatnosti F kontinuirane slučajne varijable X se definira formulom:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

, vrijede:

$$P(a < X \leq b) = F(b) - F(a)$$
$$F'(x) = f(x)$$

Normalna/Gaussova distribucija je distribucija koja prima parametre μ (očekivanje) i σ^2 (varijanca). Da neprekidna slučajna varijabla X ima Gaussovu distribuciju označavamo sa:

$$X \sim N(\mu, \sigma^2)$$

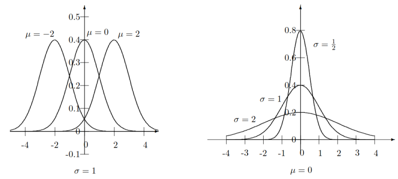
Za standardnu normalnu distribuciju vrijede $\mu = 0$ i $\sigma^2 = 1$. Jednadžba Gaussove distribucije je:

$$f(x) = \frac{1}{\sigma\sqrt{2\varpi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

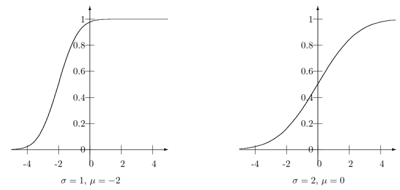
Funkcija distribucije vjerojatnosti se definira sa:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Funkcije gustoće vjerojatnosti za različite σ^2 i μ



Funkcije distribucije vjerojatnosti:



Vjerojatnost da slučajna varijabla $X \sim N(\mu, \sigma^2)$ poprimi vrijednosti manju od ili jednaku b:

$$P(X \leq b) = \int_{-\inf}^{b} f(x)dx = F(b)$$
$$= F^*\underbrace{\left(\frac{b-\mu}{\sigma}\right)}_{b^*} = F^*(b^*)$$

$$P(a < X) = 1 - P(X \leq a) = 1 - F(a) = 1 - F^*(a^*)$$
$$P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$
$$= F(b) - F(a) = F^*(b^*) - F^*(a^*)$$

Približnu vrijednost za $F^*(b^*)$ uzimamo direktno iz tabele. Za $F(x) = F^*(x^*)$, za $x^* > 4.09$ uzimamo da je $F^*(x^*) \approx 1$. Za $x^* < 0$ uzimamo $F^*(x^*) = 1 - F^*(-x^*)$, tj. za $x^* = -0.5 \rightarrow F^*(-0.5) = 1 - F(0.5)$

Broj sunčanih dana u jednoj godini na otoku Hvaru normalno je distribuirana slučajna varijabla X s parametrima $\mu = 251$ i $\sigma^2 = 49$. Izračunajte vjerojatnosti da je broj sunčanih dana u godini veći od 256 i između 249 i 259

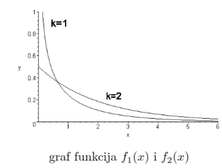
$$P(X > 256) = 1 - P(X \leq 256) = 1 - F(256) =$$
$$= 1 - F^*\left(\frac{256-\mu}{\sigma}\right) = 1 - F\left(\frac{256-251}{7}\right) \approx$$
$$\approx 1 - F^*(0.71) \approx 1 - 0.76115 = 0.23885$$

$$P(249 < X \leq 259) = P(X \leq 259) - P(X \leq 249) =$$
$$= F(259) - F(249) \approx F^*(1.14) - F^*(-0.29) =$$
$$= F^*(1.14) - (1 - F^*(0.29)) \approx$$
$$\approx 0.87286 - (1 - 0.61409) = 0.48695$$

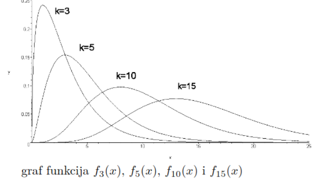
χ^2 **distribucija** ovisi o parametru $k \in \mathbb{N}$ kojeg zovemo stupanj slobode. Funkcija gustoće vjerojatnosti:

$$f_k(x) = c \cdot x^{\frac{k}{2}-1} \cdot e^{-\frac{x}{2}}, \quad x > 0$$

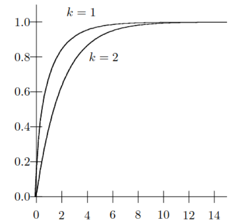
, gdje je $cx > 0$ kosntanta koja ovisi o k tako da vrijedi $\int_0^{\infty} f_k(t)dt = 1$ (pogledaj χ^2 test na idućoj str.)



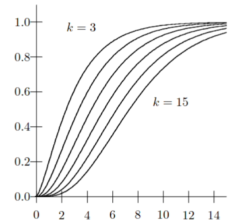
graf funkcija $f_1(x)$ i $f_2(x)$



graf funkcija $f_3(x)$, $f_5(x)$, $f_{10}(x)$ i $f_{15}(x)$



graf funkcija $F_1(x)$ i $F_2(x)$



graf funkcija $F_3(x)$, $F_5(x)$, $F_{10}(x)$ i $F_{15}(x)$

Statistički testovi se dijele na dva tipa:

1. Testiranje hipoteze o očekivanju normalno distribuirane varijable uz poznatu varijancu
2. χ^2 test

Testiranje hipoteze. Pri testiranju hipoteze uvijek imamo nul-hipotezu (H_0). Rezultat testa može biti njeno prihvaćanje ili odbacivanje. Ako odbacimo nul-hipotezu, znači da prihvaćamo alternativnu hipotezu (H_1) koja može i ne mora biti suprotnost nul-hipoteze.

Testovi hipoteze imaju jedan parametar, razina značajnosti (α), koja ima značenje vjerojatnosti da ćemo napraviti grešku prve vrste. Taj parametar treba biti što manji, ali ne može biti 0 (inače se uzima $\alpha = 0.05$ tj. 5% ili $\alpha = 0.01$ tj. 1%).

Razinu značajnosti možete shvatiti kao nekakav threshold, ili granica, ispod koje kažete da neki događaj *nije* mogao biti slučajan (nul-hipoteza se odbacuje), a iznad kažete da je mogao biti slučajan (nul-hipoteza idalje vrijedi).

Razlikuju se 3 vrste testa:

$$\textbf{lijevi}\begin{cases} H_0 : \mu = \mu_0 & \text{nul-hipoteza} \\ H_1 : \mu < \mu_0 & \text{alternativna hipoteza} \end{cases}$$

$$\textbf{desni}\begin{cases} H_0 : \mu = \mu_0 & \text{nul-hipoteza} \\ H_1 : \mu > \mu_0 & \text{alternativna hipoteza} \end{cases}$$

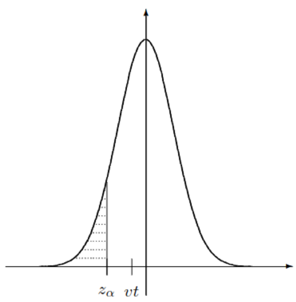
$$\textbf{obostrani}\begin{cases} H_0 : \mu = \mu_0 & \text{nul-hipoteza} \\ H_1 : \mu \neq \mu_0 & \text{alternativna hipoteza} \end{cases}$$

U svakom od njih se računa vrijednost testa $\dfrac{\overline{x} - \mu}{\sigma}\sqrt{n}$. Uz to, mora nam biti poznata razina značajnosti α Testovi se odabiru po pravilu:

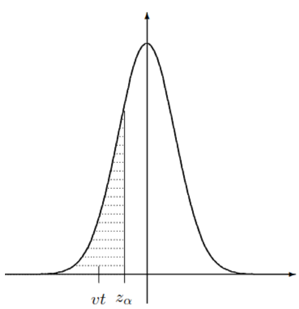
1. **lijevi** ako su sve vrijednosti u uzorku manje od μ_0
2. **desni** ako su sve vrijednosti u uzorku veće od μ_0
3. **obostrani** ako nijedno iznad ne vrijedi

lijevi testis: za lijevi test vrijedi $z_\alpha = -z_{(1-\alpha)}$.

Ako je $\frac{\overline{x}-\mu}{\sigma}\sqrt{n} > z_\alpha = -z_{(1-\alpha)}$, onda nul-hipotezu prihvaćamo s razinom značajnosti α. U suprotnom ju odbacujemo s razinom značajnosti $\alpha = 0.05$.



nul hipoteza se prihvaća
(vt=vrijednost testa)



nul hipoteza se odbacuje

desni testis: Ako je $\frac{\overline{x}-\mu}{\sigma}\sqrt{n} < z_{1-\alpha}$, onda nul-hipotezu prihvaćamo s razinom značajnosti α. U suprotnom ju odbacujemo s razinom značajnosti $\alpha = 0.05$.

obostrani testis: Ako:

- $\dfrac{\overline{x} - \mu}{\sigma}\sqrt{n} > z_{\frac{\alpha}{2}} = -z_{(1-\frac{\alpha}{2})}$
- $\dfrac{\overline{x} - \mu}{\sigma}\sqrt{n} < z_{(1-\frac{\alpha}{2})}$

oboje vrijede, tada nul-hipotezu prihvaćamo s razinom značajnosti α. U suprotnom ju odbacujemo s razinom značajnosti $\alpha = 0.05$

Očitavanje vrijednosti z_α: Vrijednost z_α se očitava u tablici vrijednosti standardne normalne razdiobe.

1. Pronađite α ili najbliži broj u samoj tablici (npr. 0.925, tj. u tablici je najbliži 0.92507)
2. Zbrojite "naziv" redka i stupca u kojem se nalazi za vrijednost z_α (za $\alpha = 0.925 \approx 0.92507$ to su redak "1.4" i stupac "0.04", tj. $z_{(0.925)} = 1.4 + 0.04 = 1.44$)

U tablici su popisana izmjerena odstupanja promjera matica od propisane vrijednosti (mm). Uz pretpostavku da je vrijednost odstupanja normalno distribuirana slučajna varijabla s parametrima μ i $\sigma^2 = 0.0036$, odgovarajućim testom testirajte hipotezu $H_0 : \mu = 0.29$ uz razinu značajnosti $\alpha = 0.001$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
0.27	0.35	0.21	0.28	0.27	0.27	0.31	0.24

Izračunajmo aritmetičku sredinu:

$$\overline{x} = \frac{0.27 + 0.35 + \cdots + 0.24}{8} = 0.275$$

Vrijednost testa je:

$$\frac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \frac{0.275 - 0.29}{0.06}\sqrt{8} \approx -0.7071$$

U uzorku ima vrijednosti ispod i iznad $\mu_0 = 0.29$ pa koristimo obostrani test:

$$\begin{cases} H_0 : \mu = 0.29\text{mm} & nul - hipoteza \\ H_1 : \mu \neq 0.27\text{mm} & alternativnahipoteza \end{cases}$$

Treba izračunati kritične vrijednosti $z_{\frac{\alpha}{2}}$ i $z_{(1-\frac{\alpha}{2})}$. Za obostrani test znamo da vrijedi $z_{\frac{\alpha}{2}} = -z_{(1-\frac{\alpha}{2})}$, pa je dovoljno naći samo $z_{(1-\frac{\alpha}{2})} = z_{0.9995} = 3.29$

Imamo $z_{0.9995} = 3.29$ pa prema tome imamo i $z_{0.0005} = -3.29$. Pošto vrijednost testa ($\frac{\overline{x}-\mu_0}{\sigma}\sqrt{n}$, dobiveno -0.7071) nije u kritičnom području (tj. vrijednost testa *je između* $z_{0.0005}$ i $z_{0.9995}$, -3.29 i 3.29 ("kritično područje" se nalazi *izvan* tog raspona)), nul-hipotezu prihvaćamo s razinom značajnosti $\alpha = 0.001$

Isti zad, drugaciji brojevi: $\sigma^2 = 0.0016$, testirajte hipotezu H_0 : $\mu = 1.15$ uz razinu značajnosti $\alpha = 0.05$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1.19	1.17	1.19	1.16	1.18	1.17	1.16	1.19

$$\overline{x} = 1.17625$$

$$\text{v.t.}\ \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \frac{1.17625 - 1.15}{0.04}\sqrt{8} \approx 1.8562$$

Pošto su u uzorku svi elementi veći od μ_0, koristimo desni test:

$$\begin{cases} H_0 : \mu = 1.15 & \text{nul-hipoteza} \\ H_1 : \mu > 1.15 & \text{alternativna hipoteza} \end{cases}$$

Treba izračunati vrijednost $z_{(1-\alpha)} = z_{(0.95)} = 1.64$. Pošto je vrijednost testa (1.8562) veća od $z_{(1-\alpha)}$, vrijednost testa se nalazi u kritičnom području te nul-hipotezu odbacujemo s razinom značajnosti $\alpha = 0.05$

χ^2 **test** se koristi kada želimo provjeriti da li dobivena očitanja nekog slučajnog događaja odgovaraju očekivanim frekvencijama. Recimo da imamo tabelu:

X	x_1	x_2	x_3	$\ldots$	x_{m-1}	x_n
f	f_1	f_2	f_3	$\ldots$	f_{m-1}	f_m

I da šta god mjerimo očekujemo da će imati (zadanu) distrubiciju:

$$\begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_{m-1} & x_m \\ p_1 & p_2 & p_3 & \ldots & p_{f-1} & p_m \end{pmatrix}$$

Tada, ako je $n \cdot p_j \geq 5$ **za svaki** j, gdje je n broj uzoraka, onda možemo primjeniti χ^2 test.

χ^2 testom provjeravamo nultu hipotezu (H_0) koja glasi "*distribucija mjerene veličine X jednaka je teoretskoj distribuciji*" (tj. vjerojatnosti događaja su one koje očekujemo). Tj.:

$$H_0\ :\ P(x_1) = p_1,\ P(x_2) = p_2,\ \ldots,\ P(x_m) = p_m$$

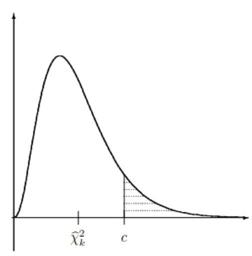
Da provjerimo to sa nekom razinom značajnosti α (obično $\alpha = 0.05, \alpha = 0.01$ ili $\alpha = 0.1$) računamo vrijednost χ^2 testa $\widehat{\chi}^2_k$, gdje je $k = m - 1$ stupanj slobode

$$\widehat{\chi}^2_k = \frac{(f_1 - np_1)^2}{np_1} + \frac{(f_2 - np_2)^2}{np_2} + \cdots + \frac{(f_m - np_m)^2}{np_m}$$

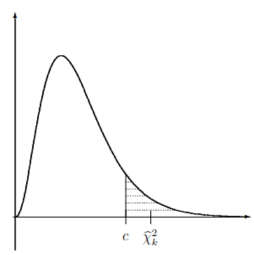
Dobivenu vrijednost onda uspoređujete sa vrijednosti očitane iz tabele vrijednosti funkcije distribucije χ^2 razdiobe.

Očitavanje iz χ^2 tabele: Uz poznat stupanj slobode $k = m - 1$ i razinu značajnosti α:

1. Pronađite redak koji odgovara stupnju slobode k (u tabeli su retci dani labelu m, ali tražite vrijednost za k)
2. Pronađite stupac sa vrijednosti $1 - \alpha$
3. Očitajte vrijednost na sjecištu, c
4. Ako je dobivena vrijednost za $\widehat{\chi}^2_k < c$, vrijednost se *ne nalazi* u kritičnom području i nul-hipoteza je potvrđena sa razinom značajnosti α.
5. U suportnom, ako je $\widehat{\chi}^2_k \geq c$, vrijednost *se nalazi* u kritičnom području, i nul-hipoteza se tada odbacuje sa razinom značajnosti α



nul hipoteza se prihvaća



nul hipoteza se odbacuje

1. Pretpostavimo da je masa čipsa kojeg stroj pakira u jednu vrećicu normalno distribuirana slučajna varijabla s aparametrima $\mu = 161g$ i $\sigma^2 = 64g^2$ $(X \sim N(161, 64), \sigma = \sqrt{\sigma^2} = 8)$. Izračunajte vjerojatnost da je masa čipsa u slučajno odabranoj vrećici:

   (a) Veća od 180g ($X > 180$ ili ti $180 < X$)

   Za ovaj zadatak se koristi formula $P(X \leq b) = F^*\left(\frac{b-\mu}{\sigma}\right) = F^*(b^*)$ i spoznaja da $P(a < X) = 1 - P(X \leq a)$. Računajući a^* se dobije:

   $$a^* = \frac{a - \mu}{\sigma} = \frac{180 - 161}{8} = 2.375$$

   $F^*(2.375)$ se onda iščita iz tablice: $F^*(2.375 \approx 2.38) = 0.99134$, pa je $P(a < X) = 1 - P(X \leq a) = 1 - 0.99134 = 0.00866$

   (b) Između 154g i 165g ($154 < X \leq 165$)

   Tu se koristi spoznaja da je $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$ i, ako je $a^* < 0 \rightarrow F^*(a^*) = 1 - F^*(-a^*)$.

   $$b^* = \frac{b - \mu}{\sigma} = 0.5, \quad a^* = -0.875 \approx 0.88$$

   $$F^*(b^*) \approx 0.69146, \quad F^*(a^*) = 1 - F*(-a^*) = 1 - F^*(0.88) \approx 1 - 0.81057 = 0.18943$$

   $$P(154 < X \leq 165) \approx 0.69146 - 0.18943 \approx 0.50203$$

---

2. Provjeren je rad stroja za prozivodnju matrica. U tabeli su popisana izmjerena odstupanja promjera matica od propisane vrijednosti (u mm). Uz pretpostavku da je vrijednost odstupanja normalno distribuirana slučajna varijabla s parametrima μ i $\sigma^2 = 0.0036$, odgovarajućim testom testirajte hipotezu $H_0 : \mu = 0.24$ uz razinu značajnosti $\alpha = 0.005$. Navedite alternativnu hipotezu.

x_1	x_2	x_3	x_4	x_5	x_6	x_7
0.33	0.3	0.33	0.33	0.3	0.33	0.27

Pošto u uzorcima postoje vrijednosti striktno veće od $\mu = 0.24$, koristimo **desni** test: $\begin{cases} H_0 : \mu = 0.24 & \text{nul-hipoteza} \\ H_1 : \mu > 0.24 & \text{alternativna hipoteza} \end{cases}$

$$\overline{x} \approx 0.313 \quad \text{v.t.: } \frac{\overline{x} - \mu}{\sigma}\sqrt{n} = \frac{0.313 - 0.24}{0.06} \approx 3.219$$

Za desni test provjerava se $\frac{\overline{x}-\mu}{\sigma}\sqrt{n} < z_{(1-\alpha)}$. Ako vrijedi, nul-hipoteza se prihvaća

$$z_{(1-\alpha)} = z_{(0.995)} \approx z_{(0.99506)} = 2.58 \qquad 3.219 \not< 2.58 \rightarrow \text{ nul-hipoteza se odbacuje}$$

---

3. Proizvođač loptica za stolni tenis pakira po pet loptica u jednu kutiju. Loptice mogu biti ili ispravne ili neispravne. Radi kontrole kvalitete proizvoda i načina pakiranja loptica potrebno je odrediti teoretsku razdiobu pakiranja neispravnih loptica po kutijama. Na slučajan način je odabrano 77 kutija za koje su dobiveni sljedeći podaci:

broj neispravnih loptica x_i	0	1	2	3	4	5
broj kutija f_i	28	12	16	5	7	9

Na ovom uzorku testirajte nulhipotezu, uz razinu značajnosti $\alpha = 0.025$, da broj neispravnih loptica po kutiji ima **teoretsku razdiobu**:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ \frac{5}{12} & \frac{1}{6} & \frac{5}{24} & \frac{1}{12} & \frac{1}{24} & \frac{1}{12} \end{pmatrix}$$

Računamo vrijednost testa sa stupnjem slobode $k = m - 1 = 6 - 1 = 5$ i brojem uzorka $n = 77$ (*btw, kalkulator ima limit na broj razlomaka u jednom izrazu. Možda će te morat na više navrata računat sve ovo*):

$$\widehat{\chi}_k^2 = \frac{(f_1 - np_1)^2}{np_1} + \frac{(f_2 - np_2)^2}{np_2} + \cdots + \frac{(f_m - np_m)^2}{np_m}$$

$$= \frac{(\overset{x_1}{28} - \overset{n}{77} \cdot \overset{p_1}{\frac{5}{12}})^2}{77 \cdot \frac{5}{12}} + \frac{(12 - 77 \cdot \frac{1}{6})^2}{77 \cdot \frac{1}{6}} + \frac{(16 - 77 \cdot \frac{5}{24})^2}{77 \cdot \frac{5}{24}} + \cdots + \frac{(9 - 77 \cdot \frac{1}{12})^2}{77 \cdot \frac{1}{12}}$$

$$\approx 6.4078$$

Sada u tablici za χ^2 distribuciju nalazimo redak $k = 5$ (u tablici su retci označeni labelom m, no tražite redak sa vrijednosti k) i stupac $1 - \alpha = 0.975$ i očitavamo vrijednost c na sjecištu: $c = 12.832$. Ako je dobivena vrijednost $\widehat{\chi}_k^2 < c$, nul-hipoteza se prihvaća sa razinom značajnosti α. U suprotnom se odbacuje sa razinom značajnosti α. U ovom zadatku, $6.4078 < 12.832$ ✓, pa nul-hipotezu prihvaćamo.

$$x^* = \frac{x - \mu}{\sigma}, \qquad P(X \leq b) = F^*(b^*), \qquad P(a < X) = 1 - P(X \leq a)$$

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a), \qquad x^* < 0 \rightarrow F^*(x^*) = 1 - F(-x^*)$$

l.t.: $\frac{\overline{x} - \mu}{\sigma}\sqrt{n} > -z_{(1-\alpha)}?$ **d.t.:** $\frac{\overline{x} - \mu}{\sigma}\sqrt{n} < z_{(1-\alpha)}?$ **o.t.:** $\left(\frac{\overline{x} - \mu}{\sigma}\sqrt{n} > -z_{(1-\frac{\alpha}{2})}?\right)$ && $\left(\frac{\overline{x} - \mu}{\sigma}\sqrt{n} < z_{(1-\frac{\alpha}{2})}?\right)$ true $\rightarrow H_0$✓ false $\rightarrow \cancel{H_0}$

$$\widehat{\chi}_k^2 = \frac{(f_1 - np_1)^2}{np_1} + \frac{(f_2 - np_2)^2}{np_2} + \cdots + \frac{(f_m - np_m)^2}{np_m}, \qquad c = \text{tabela } \chi^2 \begin{Bmatrix} \text{redak } k \\ \text{stupac } 1 - \alpha \end{Bmatrix} \rightarrow \text{sjecište}, \quad \widehat{\chi}_k^2 < c \;?$$

true $\rightarrow H_0$✓ false $\rightarrow \cancel{H_0}$