(06) Testovanie hypotéz, testovanie hypotéz o pravdepodobnostných rozdeleniach

1. Princíp testovania hypotéz: príklad o cukríkoch

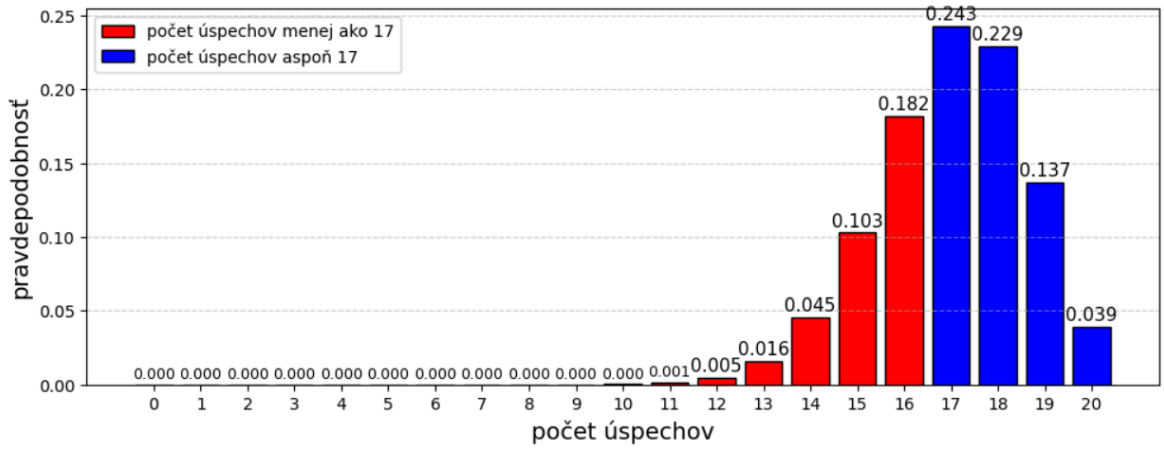
<u>Príklad 1:</u> rozoznávanie Lentiliek a M&M's (Katarína Kocsisová, bakalárska práca *Experimenty a vlastné dáta pri vyučovaní pravdepodobnosti a štatistiky*, 2015)

Človek tvrdí, že dokáže na základe chute rozlíšiť Lentilky a M&M's. Nie stopercentne, ale tvrdí, že s pravdepodobnosťou 0,85 to dokáže. My pochybujeme, že to dokáže až tak presne. Chceme preto zrealizovať experiment, na základe ktorého budeme vedieť (s určitou presnosťou) rozhodnúť, či jeho tvrdenie o schopnosti rozlišovať cukríky s uvedenou pravdepodobnosťou 0,85 zamietnuť alebo nie.

<u>Intuícia:</u> Necháme ho určovať druh cukríkov (so zaviazanými očami, cukríky podávame na lyžičke, aby nemohol hmatom zistiť ich veľkosť, možnosť napiť sa medzi pokusmi kvôli neutralizácii chute) viackrát a ak bude mať správnych odpovedí príliš málo, jeho tvrdenie o presnosti 0,85 zamietneme. Povedzme, že spravíme 20 pokusov. Kam dať hranicu - koľko správnych odpovedí bude "príliš málo"?

<u>Prvý nápad:</u> Očakávaný počet správnych odpovedí je 20*0,85 = 17. Čo sa stane, ak náš test bude vyzerať tak, že jeho tvrdenie zamietneme, ak správne určí menej ako 17 cukríkov (teda "menej ako by v priemere mal")?

Spočítajme, s akou pravdepodobnosťou jeho tvrdenie zamietneme v prípade, že by bolo v skutočnosti pravdivé. Počet správnych odpovedí má binomické rozdelenie, zaujíma nás pravdepodobnosť toho, že ich bude menej ako 17.




```
from scipy.stats import binom

print(f"Pravdepodobnosť zamietnutia pravdivého tvrdenia: {binom.cdf(16, 20, 0.85):.4f}")
```

```
Pravdepodobnosť zamietnutia pravdivej hypotézy: 0.3523
```

Ak by sme tento postup opakovali s viacerými ochutnávačmi a každý z nich by mal pravdepodobnosť správnej odpovede 0,85, tak s viac ako 35 percentnou pravdepodobnosťou by sme ich *pravdivé* tvrdenie o ich schopnostiach zamietli. To je príliš veľa.

Úplne sa takejto chybe nemôžeme vyhnúť. Ak by sme sa jej chceli zaručene vyhnúť, nemohli by sme uvažované tvrdenie zamietnuť nikdy. Takýto test by však nemal žiadny zmysel.

Čo sa teda robí: **Stanoví sa hranica, s akou maximálnou pravdepodobnosťou sme ochotní zamietnuť hypotézu, ktorá je v skutočnosti pravdivá. Štandardnou voľbou je 0,05.** V našom prípade:

- Ak je naše kritérium "tvrdenie o 85 percentnej úspešnosti zamietame, ak je počet správnych odpovedí menej ako **k**", tak pravdepodobnosť zamietnutia pravdivého tvrdenia je **binom.cdf(k - 1, 20, 0.85).**
- Na základe grafu vieme odhadnúť, aké maximálne **k** uvažovať:

```
Kritérium zamietnutia: menej ako 14 správnych odpovedí
Pravdepodobnosť zamietnutia pravdivého tvrdenia: 0.02194

Kritérium zamietnutia: menej ako 15 správnych odpovedí
Pravdepodobnosť zamietnutia pravdivého tvrdenia: 0.06731
```

- Prečo chceme maximálne možné **k**: lebo chceme čo najlepšie identifikovať nesprávne tvrdenie a ak je skutočná pravdepodobnosť menšia ako 0,85, chceme s čo najväčšou pravdepodobnosťou tvrdenie o hodnote 0,85 zamietnuť. To by sa nestalo, ak by sme zvolili napríklad **k=12**. Stále by platilo, že pravdivé tvrdenie zamietneme s pravdepodobnosťou nie väčšou ako 0,05. Ale často by sme robili opačnú chybu: testované tvrdenie by neplatilo, ale my by sme ho nezamietli.
- Záver: **k=14,** a teda
    - Ak je počet správnych odpovedí 0-13, tak tvrdenie o presnosti 0,85 zamietneme.
    - Ak je počet správnych odpovedí aspoň 14, tak tvrdenie o presnosti 0,85 nezamietneme.

Reálny výsledok: V experimente z bakalárskej práce sa dosiahlo 18 správnych odpovedí, takže hypotéza sa nezamietla.

Otázka: Predpokladajme, že skutočná pravdepodobnosť je menšia ako 0,85. S akou pravdepodobnosťou to odhalíme v prípade, že zvolíme náš test s hranicou 14 a v prípade, že zvolíme test s hranicou 12, ak

- skutočná pravdepodobnosť je 0,5 (človek iba háda),
- skutočná pravdepodobnosť je 0,65 (človek nehádа, ale precenil svoje schopnosti, keď povedal, že s 85 percentnou pravdepodobnosťou určí druh cukríka).

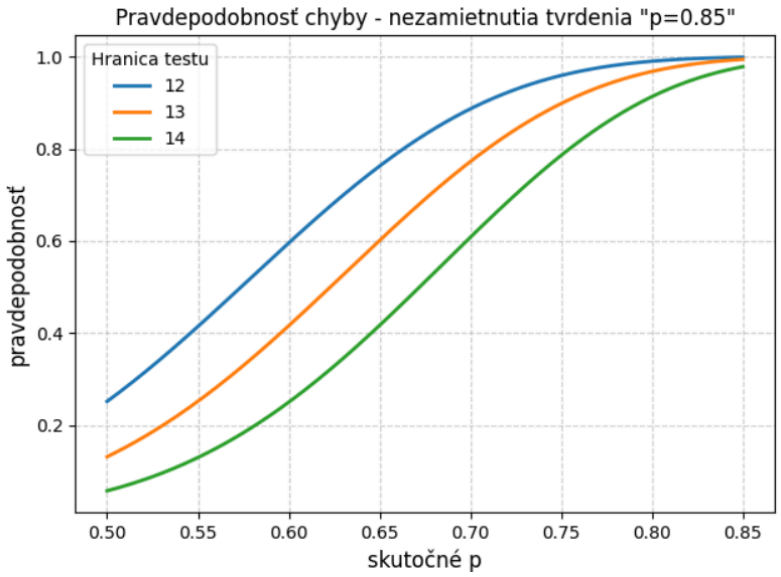
Odpoveď:

- Odhaliť nesprávne tvrdenie znamená, že hypotézu 0,85 zamietneme.
- Ak je hranica testu **k**, tak zamietame pre počet úspechov od 0 do k -1 (vrátane).
- Počet úspechov má binomické rozdelenie, pričom **n** je 20 (počet pokusov) a parameter **p** je skutočná pravdepodobnosť.

```
def pravdepodobnost_nezamietnutia(hranica_testu, skutocne_p, pocet_pokusov=20):
  return(1 - binom.cdf(hranica_testu - 1, pocet_pokusov, skutocne_p))
```

Pre interval pravdepodobnosti medzi 0,5 (menšiu neuvažujeme, ak by si niekto odsledoval, že má presnosť menšiu ako ½, stačilo by mu povedať opak toho, čo sa mu zdá ako správna odpoveď a pravdepodobnosť si zvýšiť nad ½) a 0,85 (to, čo testujeme):

```
hranica testu: 12, skutočné p: 0.5
pravdepodobnosť chyby, že nezamietneme tvrdenie "p=0.85": 0.25172

hranica testu: 14, skutočné p: 0.5
pravdepodobnosť chyby, že nezamietneme tvrdenie "p=0.85": 0.05766

hranica testu: 12, skutočné p: 0.65
pravdepodobnosť chyby, že nezamietneme tvrdenie "p=0.85": 0.76238

hranica testu: 14, skutočné p: 0.65
pravdepodobnosť chyby, že nezamietneme tvrdenie "p=0.85": 0.41663
```



2. Terminológia pri testovaní hypotéz

pojem	**vysvetlenie**	**v príklade o cukríkoch**
nulová hypotéza, označujeme ju H_0	hypotéza, ktorú testujeme	p=0,85
alternatívna hypotéza, označueme ju H_1	čo platí, ak neplatí nulová hypotéza	môže to byť napríklad • p<0,85 (nevie to tak presnne, ako to tvrdí) • p=0,5 (iba háda)
testovacia štatistika	hodnota určená z dát, na základe ktorej sa rozhodne, či nulovú hypotézu zamietnuť alebo nezamietnuť	počet správnych odpovedí z 20 pokusov
obor zamietnutia	pre aké hodnoty testovacej štatistiky nulovú hypotézu zamietneme	{0, 1, 2, …., 13}
hladina významnosti	maximálna povolená pravdepodobnosť, s akou sa zamieta pravdivá nulová hypotéza	0,05 (štandardná hladina významnosti)
sila testu	pravdepodobnosť, že zamietneme nulovú hypotézu, ktorá neplatí	počíta sa funkciou pravdepodobnost_nezamietnutia

3. Príklad o eurovej minci

<u>Príklad 2:</u> Euro coin accused of unfair flipping (NewScientist, 2002)
https://www.newscientist.com/article/dn1748-euro-coin-accused-of-unfair-flipping/

- Hádzanie eurovou mincou, 250 pokusov, 140-krát padla hlava
- Pri pravidelnej minci očakávame 125 hláv – je tento rozdiel "príliš veľký" alebo ho môžeme pripísať náhode?

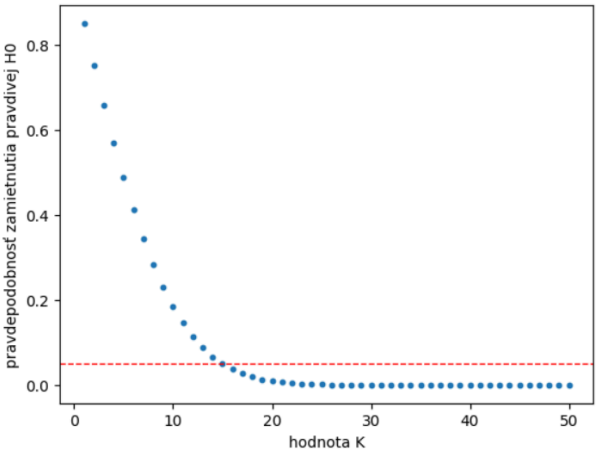
Spravme to isté ako v príklade o cukríkoch:

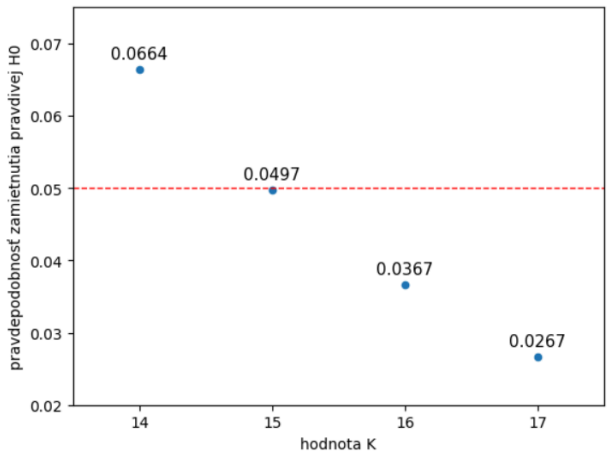
- H_0: minca je pravidelná, teda ***p=½***
- H_1: minca nie je pravidelná, teda **p nie je 1/2**
- Testovacia štatistika: **počet hláv**
- Hypotézu **zamietneme, ak sa bude skutočný počet hláv príliš líšiť od očakávaného** – bude príliš veľký alebo príliš malý (hranice nastavíme tak, aby pravdepodobnosť zamietnutia pravdivej H_0 nebola viac ako 5 percent)

<u>Postup 1:</u>
Kritérium pre zamietnutie bude mať tvar *"počet hláv je viac ako* **125+K** *alebo menej ako* **125-K***"*. Predpokladajme, že H_0 platí, teda minca je pravidelná. Aká je pravdepodobnosť, že ju zamietneme:

```
from scipy.stats import binom

def pravd_zamietnutia_pravdivej_H0(K):
    viac_ako_125plusK = 1 - binom.cdf(125 + K, 250, 0.5)
    menej_ako_125minusK = binom.cdf(125 - K, 250, 0.5) - binom.pmf(125 - K,
250, 0.5)
    return viac_ako_125plusK + menej_ako_125minusK
```

Graf + zoom:





Znovu:

- Pravdepodobnosť zamietnutia pravdivej H_0 nemôže byť viac ako 0,05.
- Na druhej strane nechceme "zbytočne malú" množinu výsledkov, pri ktorých H_0 zamietame, lebo to súčasne znamená, že klesá pravdepodobnosť odhalenia nepravidelnej mince.

Zoberieme teda **K=15,** kedy prvýkrát klesla pravdepodobnosť zamietnutia pravdivej H_0 pod 0,05 → **hypotézu o pravidelnej minci zamietneme, ak je hláv viac ako 125+15 = 140 alebo menej ako 125-15=110.**

V našom prípade je hlác 140, teda hypotézu nezamietame. (Ale vidíme, že je to tesné.)

V článku:
Howard Grubb, an applied statistician at the University of Reading, notes that, "with a sample of only 250, anything between 43.8 per cent and 56.2 per cent on one side or the other cannot be said to be biased". The range of 6.2 per cent on either side of 50 per cent is expected to cover the results, even with a fair coin, in 95 of every 100 experiments.

43,8% z 250 je 109,5 a 56,2% z 250 je 140,5.

Postup 2:
Nebudeme pracovať s binomickým rozdelením, ale na základe centrálnej limitnej vety ho aproximujeme normálnym rozdelením.

Počet hláv **X** má za platnosti H_0 binomické rozdelenie **Bin(250, ½),** aproximujeme ho normálnym rozdelením **N(125, 250 * ½ * (1 - ½)).** Náhodnú premennú **X** normalizujeme a vytvoríme náhodnú premennú s **N(0, 1)** rozdelením:

Z = (X - 125)/sqrt(250 * ½ * (1 - ½))

Nájdeme teraz množinu extrémne veľkých a extrémne malých hodnôt (zodpovedá to extrémne veľkým a extrémne malým hodnotám **X**), ktoré majú spolu pravdepodobnosť 0,05.

Kvantily:

```
q_0025 = norm.ppf(0.025)
q_0975 = norm.ppf(0.975)
print(f"2.5% kvantil: {q_0025:.5f}")
print(f"97.5% kvantil: {q_0975:.5f}")
```

```
2.5% kvantil: -1.95996
97.5% kvantil: 1.95996
```

```
normalizovana_statistika = (140 - 250*0.5)/np.sqrt(250*0.5*0.5)
print(f"Normalizovaná_štatistika: {normalizovana_statistika:.5f}")
```

```
Normalizovaná_štatistika: 1.89737
```

Hypotézu H_0 zamietame, ak je hodnota štatistiky menšia ako -1,95996 alebo väčšia ako 1,95996. To v našom prípade neplatí, takže hypotézu o pravidelnej minci nezamietame.

Všimnime si, že tieto hodnoty kvantilov sú univerzálne, nezávisia od testovaných dát, počtu pozorovaní, ani testovanej pravdepodobnosti. Využívajú len to, že testovacia štatistika má pri platnosti nulovej hypotézy (v tomto prípade približne) N(0, 1) rozdelenie.

4. P-hodnota testu

Motivácia. V článku sa píše:
Nonetheless, Grubb cautions, the Polish result is at the outside of this range, and would be expected in only about 7 of every 100 experiments with a fair coin, leaving a glimmer of hope for their hypothesis. Clearly, more research is needed.

Počítali teda pravdepodobnosť dosiahnutého výsledku (“the Polish result”) v prípade, že minca je pravidelná. Nejde o konkrétne 140 hláv ale o to, že pri pravidelnej minci nastane “takýto extrémny výsledok”, ako sme pozorovali. Teda 140 a viac hláv, alebo tiež 140 a viac znakov. Prečo nás to zaujíma:

- Ak je táto pravdepodobnosť pri platnosti nulovej hypotézy “príliš malá”, znamená to, že pozorovaný výsledok hovorí v neprospech jej platnosti. Hypotézu v takom prípade zamietneme.
- Ak je pravdepodobnosť “dostatočne veľká”, znamená to, že pozorovaný výsledok nie je “veľmi výnimočný” a pri platnosti nulovej hypotézy “bežne nastáva”. Takže hypotézu nezamietneme.

Čo znamená “príliš malá pravdepodobnosť”? Aby bol tento postup ekvivalentný s predchádzajúcim postupom, kde sme porovnávali štatistiku s kvantilom, hranicou, s ktorou sa vypočítaná pravdepodobnosť porovnáva, je kritická hodnota testu, štandardne 0,05. Ak je teda pravdepodobnosť menšia ako 0,05 (resp. iná hladina významnosti, ktorú chceme použiť), tak sa nulová hypotéza zamieta. Ak je väčšia ako 0,05, tak sa nulová hypotéza nezamieta.

Vypočítaná pravdepodobnosť sa nazýva **p-hodnota** (***p-value***). Predchádzajúce pravidlo teda môžeme sformulovať:

- Ak je p-hodnota menšia ako hladina významnosti, nulovú hypotézu zamietame.
- Ak je p-hodnota väčšia ako hladina významnosti, nulovú hypotézu nezamietame.

Postup 1 (binomické rozdelenie):
"Takýto extrémny výsledok" znamená **0-105** hláv alebo **140-250** hláv. Pravdepodobnosť takéhoto výsledku je:

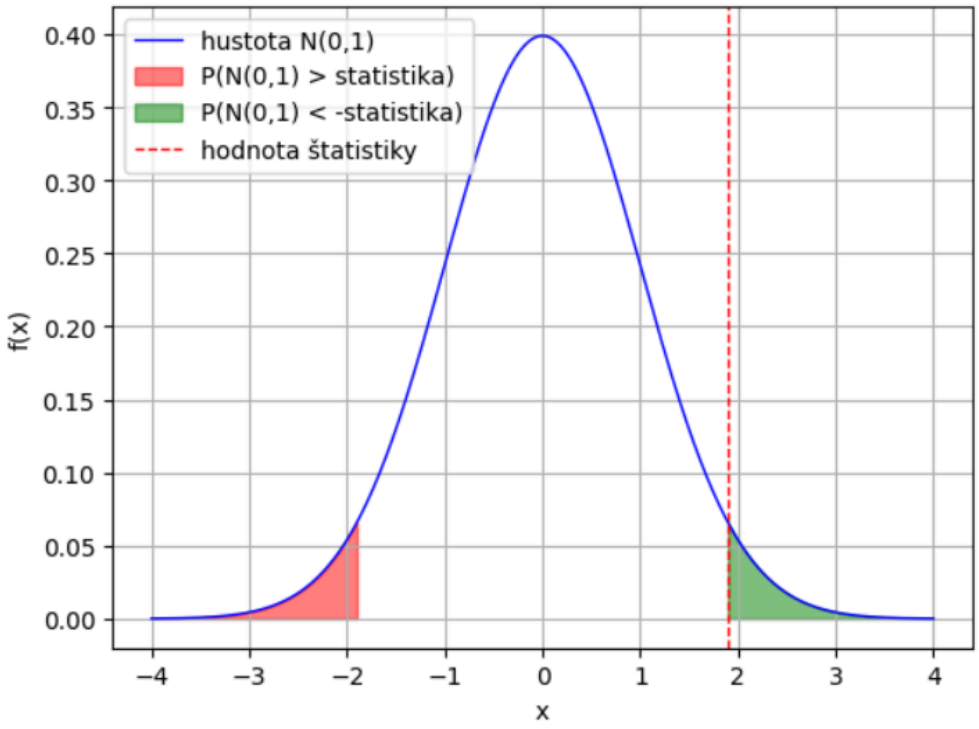
```
p_odchylka_15aviac = 1 - binom.cdf(139, 250, 0.5) + binom.cdf(110, 250, 0.5)
print(f"Pravdepodobnosť odchýlky 15 a viac: {p_odchylka_15aviac:.5f}")
```

```
Pravdepodobnosť odchýlky 15 a viac: 0.06642
```

Je to viac ako 0,05, takže hypotézu H_0 o pravidelnej minci nezamietame.

Všimnime si: 0,06642 je po zaokrúhlení na celé percentá 7 percent (hodnota spomínaná v článku)

Postup 2 (aproximácia normálnym rozdelením):
Hodnota Z štatistiky vyšla 1,89737. Ak platí nulová hypotéza, táto štatistika má N(0, 1) rozdelenie. Aká je pri tomto rozdelení pravdepodobnosť, že štatistika bude väčšia ako 1,89737 alebo menšia ako -1,89737:



```
plocha_vpravo = 1 - norm.cdf(normalizovana_statistika)
plocha_vlavo = norm.cdf(-normalizovana_statistika)
p_hodnota = plocha_vlavo + plocha_vpravo
# ekvivaletne: 2*plocha_vpravo alebo 2*plocha_vlavo

print(f"Dosiahnutá p-hodnota: {p_hodnota:.5f}")
```

```
Dosiahnutá p-hodnota: 0.05778
```

P-hodnota je väčšia ako 0,05, takže hypotézu nezamietame.

5. Testovanie hypotézy o parametre *p* binomického rozdelenia

Výpočet v Pythone:
https://www.statsmodels.org/stable/generated/statsmodels.stats.proportion.proportions_ztest.html

```
from statsmodels.stats.proportion import proportions_ztest

pocet_pokusov = 250
pocet_uspechov = 140
testovane_p = 1/2

statistika, p_hodnota = proportions_ztest(pocet_uspechov,
                                          pocet_pokusov,
                                          testovane_p,
                                          prop_var=testovane_p)
print(f"Z štatistika: {statistika:.5f}")
print(f"p-hodnota: {p_hodnota:.5f}")

Z štatistika: 1.89737
p-hodnota: 0.05778
```

Parameter **prop_var** s hodnotou rovnou pravdepodobnosti z nulovej hypotézy je potrebný kvôli tomu, aby sa pri výpočte menovateľa štatistiky použila hodnota **p** z nulovej hypotézy tak, ako sme to odvodili.

Záver: Vidíme tu hodnotu štatistiky aj p-hodnotu (rovnaké, ako vyšli pri “ručnom” výpočte). P-hodnota je väčšia ako 0,05, takže na 5% hladine významnosti nezamietame hypotézu, že minca je pravidelná.

<u>Príklad 3:</u> Padnutie šestky na kocke

1024 hodov hracej kocky | Matfyz livestream

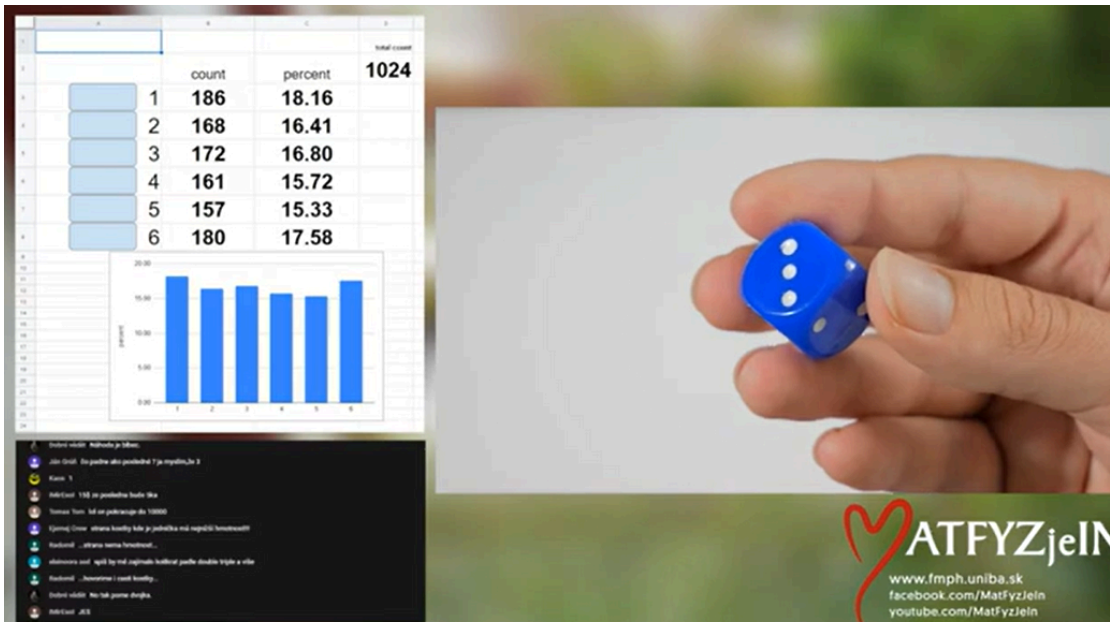
3.7K views • Streamed 5 years ago

https://www.youtube.com/watch?v=rjYT3WUWq0s

H_0: p = ⅙
H_1: p ≠ ⅙

Vypočítame štatistiku a p-hodnotu.

Aby sme si zopakovali princíp, spravíme výpočet ručne a potom porovnáme s výstupom z Pythonu.



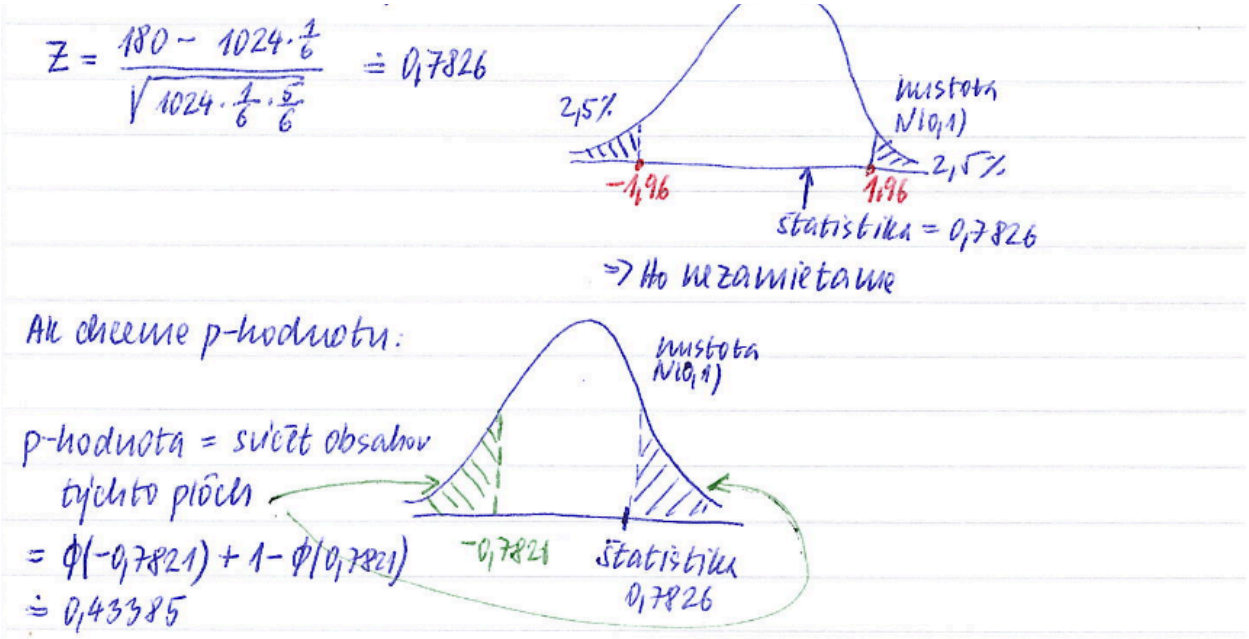
X = počet šestiek, realizovaná hodnota: 180

Ak platí H_0, X má rozdelenie $Bin(1024, \frac{1}{6})$

$E(X) = 1024 \cdot \frac{1}{6}$, $D(X) = 1024 \cdot \frac{1}{6}(1-\frac{1}{6})$

X aproximujeme $N(\mu, \sigma^2)$, $\mu = E(X), \sigma^2 = D(X)$

Z štatistika $= \frac{X-\mu}{\sigma} \overset{H_0, aprox.}{\sim} N(0,1)$




```python
from statsmodels.stats.proportion import proportions_ztest

pocet_pokusov = 1024
pocet_uspechov = 180
testovane_p = 1/6

statistika, p_hodnota = proportions_ztest(pocet_uspechov,
                                          pocet_pokusov,
                                          testovane_p,
                                          prop_var=testovane_p)
print(f"Z štatistika: {statistika:.5f}")
print(f"p-hodnota: {p_hodnota:.5f}")
```

```
Z štatistika: 0.78262
p-hodnota: 0.43385
```

P-hodnota je väčšia ako 0,05, takže hypotézu o pravdepodobnosti ⅙ nezamietame.

6. Chí-kvadrát test dobrej zhody

Motivácia 1.

Namiesto testovania šestky by sme chceli testovať naraz pravdepodobnosti všetkých čísel, teda či je kocka pravidelná.

Motivácia 2.

Pri Poissonovom rozdelení sme si ukazovali dáta o počte gólov vo futbalových zápasoch na majstrovstvách sveta, ktoré sa veľmi podobali na Poissonovo rozdelenie. Môžeme toto otestovať pomocou štatistického testu?

Základná myšlienka chí-kvadrát testu dobrej zhody: **porovnajú sa pozorované a očakávané počty hodnôt v zadaných množinách.** Množiny môžu byť hodnoty, ktoré nadobúda diskrétna náhodná premenná alebo intervaly hodnôt pre spojitú náhodnú premennú.

Predpoklad v základnej verzii testu: **parametre rozdelenia sú dané, nie odhadované z dát**

Najskôr si takýmto spôsobom prepíšeme predchádzajúci test, potom to zovšeobecníme. Pri testovaní parametra **p** binomického rozdelenia sa robilo **n** pokusov, v nich nastalo **A** úspechov a **B** neúspechov. Z toho sa vypočítala Z-štatistika, ktorá má za platnosti nulovej hypotézy približne normálne N(0, 1) rozdelenie. To znamená, že jej druhá mocnina má približne chí-kvadrát rozdelenie s jedným stupňom voľnosti

$$Z^2 = \frac{(A-np)^2}{np(1-p)} = \frac{(A-np)^2}{n} \cdot \frac{1}{p(1-p)} = \frac{(A-np)^2}{n}\left(\frac{1}{p} + \frac{1}{1-p}\right) =$$

$$= \frac{(A-np)^2}{np} + \frac{(A-np)^2}{n(1-p)} = \left| \begin{array}{l} A-np = (n-B)-np = (1-p)n - B = \\ = -(B-(1-p)n) \end{array} \right| =$$

$$= \frac{(A-np)^2}{np} + \frac{(B-(1-p)n)^2}{n(1-p)} = \frac{(\text{počet Ú} - \text{očakávaný \#Ú})^2}{\text{očakávaný \#Ú}} + \frac{(\#N - \text{očakávaný \#N})^2}{\text{očakávaný \#N}}$$

Zovšeobecnenie pre **k** množín, do ktorých zaraďujeme dáta:

- Oi = pozorované počty *(observed)*
- Ei = očakávané počty *(expected)*

Testovacia štatistika:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Hodnota štatistiky je vždy nezáporná, nulová hodnota by znamenala dokonalú zhodu. Veľká hodnota znamená, že sa očakávané a pozorované hodnoty výrazne líšia. Nulovú hypotézu teda zamietneme pre “veľmi veľké” hodnoty testovacej štatistiky.

Iný pohľad na testovaciu štatistiku:

hodnota	pozorovaná početnosť	očakávaná početnosť	rozdiel	druhá mocnina rozdielu
1	186	170.67	15.33	235.11
2	168	170.67	-2.67	7.11
3	172	170.67	1.33	1.78
4	161	170.67	-9.67	93.44
5	157	170.67	-13.67	186.78
6	180	170.67	9.33	87.11

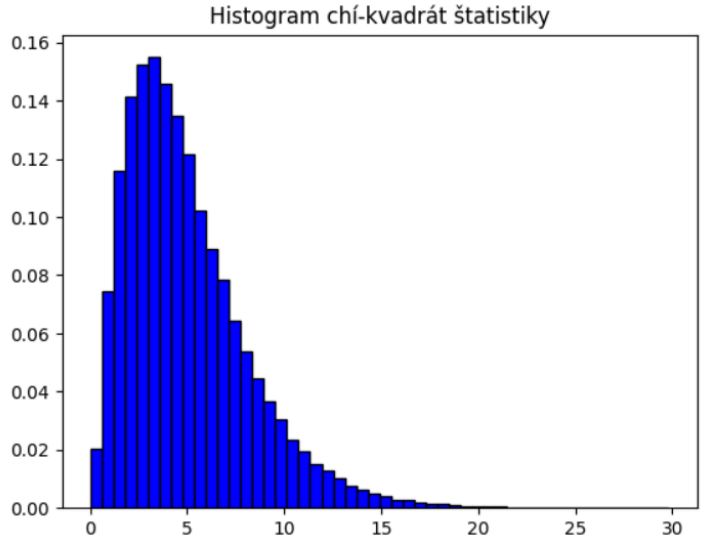
- Počítame rozdiely medzi pozorovanými a očakávanými početnosťami.
- Aby sa kladné a záporné odchýlky navzájom nezrušili, umocníme ich na druhú.
- Rozdiel napríklad 10 (resp. druhá mocnina 100) neznamená to isté v prípade, ak je očakávaný počet 20 (vtedy pozorujeme len polovicu očakávaného počtu) a 200 00 (vtedy je to zanedbateľný rozdiel). Preto druhé mocniny preškálujeme očakávanými početnosťami a tieto sčítame.

očakávaná početnosť	rozdiel	druhá mocnina rozdielu	(druhá mocnina rozdielu) / (očakávaná početnosť)
170.67	15.33	235.11	1.38
170.67	-2.67	7.11	0.04
170.67	1.33	1.78	0.01
170.67	-9.67	93.44	0.55
170.67	-13.67	186.78	1.09
170.67	9.33	87.11	0.51
		SPOLU	3.58

Čo znamená "príliš veľká hodnota" tejto štatistiky?

```
import numpy as np

def sim_kocka(pocet_pokusov=1024):
    kocky = np.random.randint(1, 7, size=pocet_pokusov)
    pozorovane = np.bincount(kocky, minlength=7)[1:7]  # pocet nul nechceme
    ocakavane = pocet_pokusov/6
```

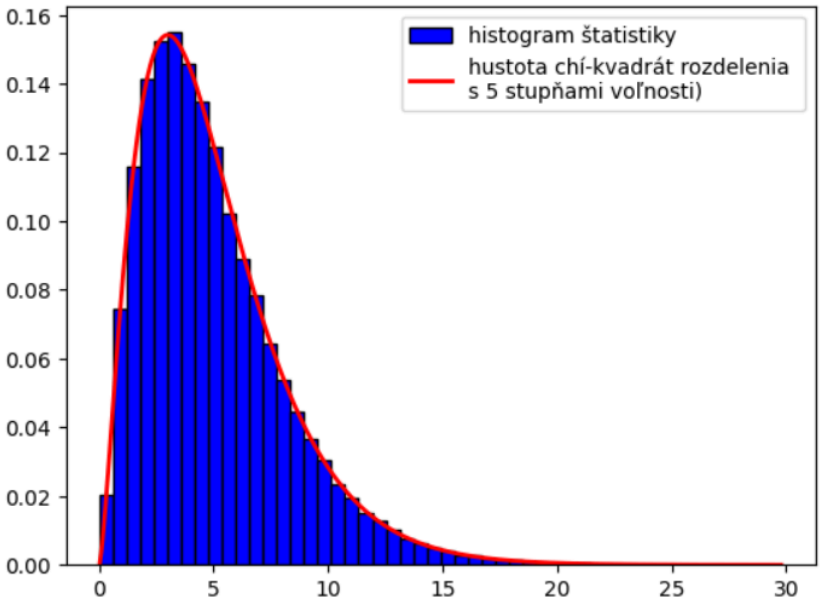
```
    chi2_stat = np.sum((pozorovane - ocakavane)**2/ocakavane)
    return chi2_stat

simulacie_statistiky = [sim_kocka() for _ in range(10**5)]
```




```
print(np.mean(simulacie_statistiky))
```

```
4.9996691796875
```

Histogram porovnáme s hustotou chí-kvadrát rozdelenia s 5 stupňami voľnosti:



Vo všeobecnosti:

- **Testovacia štatistika má za platnosti nulovej hypotézy približne k-1 stupňov voľnosti, kde k je počet množín, do ktorých zaraďujeme dáta.**
- Aby to bola dobrá aproximácia, požaduje sa, aby **očakávaný počet hodnôt v každej množine bol aspoň 5**

Pre naše kocky:

- Hypotéza o pravidelnej kocke sa zamietne, ak testovacia štatistika prekročí **kritickú hodnotu**, ktorou je 95-percentný kvantil chí-kvadrát rozdelenia s 5 stupňami voľnosti.

```
kriticka_hodnota = chi2.ppf(0.95, 5)
print(f"kritická hodnota: {kriticka_hodnota:.5f}")
```

```
kritická hodnota: 11.07050
```

- V našom prípade je štatistika menšia, takže hypotézu nezamietame.
- Výpočet p-hodnoty: Hľadáme pravdepodobnsť, s akou nastane hodnota za platnosti nulovej hypotézy štatistika "taká extrémna", ako vyšla nám. Vtedy má štatistika chí-kvadrát rozdelenie s 5 stupňami voľnosti a nás zaujíma, že bude mať hodnotu aspoň takú, ako v našom výpočte (lebo teraz "extrémna hodnota", pri ktorej H_0 zrejme neplatí, je veľká hodnota štatistiky).

```
pocet_pokusov = 1024
pozorovane = np.array([186, 168, 172, 161, 157, 180])
ocakavane = pocet_pokusov/6
chi2_stat = np.sum((pozorovane - ocakavane)**2/ocakavane)
print(chi2_stat)
```

```
3.58203125
```

```
p_hodnota = 1 - chi2.cdf(chi2_stat, 5)
print(p_hodnota)
```

```
0.6110132241946511
```

Funkcia v Pythone:
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chisquare.html

```
from scipy.stats import chisquare
statistika, p_hodnota = chisquare(pozorovane, ocakavane)
```

Pre náš príklad s kockami:

```
print(f"štatistika: {statistika:.5f}")
print(f"p-hodnota: {p_hodnota:.5f}")
```

```
štatistika: 3.58203
p-hodnota: 0.61101
```

Čo v prípade, ak sa parametre rozdelenia odhadujú:

- **Počet stupňov voľnosti sa zmenší o počet odhadovaných parametrov**
- Aké odhady parametrov zobrať:
  - Najčastejšie sa stretneme s tým, že sa vezmú "štandardné odhady", napríklad sa parameter určí tak, aby sa stredná hodnota zhodovala s priemerom, alebo sa použije tzv. metóda maximálnej vierohodnosti (budeme robiť neskôr počas semestra).
  - Nie je to celkom presné, teoretický výsledok o pravdepodobnostnom rozdelení štatistiky je pre tzv. modifikovanú metódu minimálneho chí-kvadrátu (postup, ktorý vznikne z myšlienky minimalizovať chí-kvadrát štatistiku, ale nie je to presne jej minimalizácia).
  - Zostaneme pri prvom postupe.

Príklad 3:
góly vo futbalových zápasoch

Chceme testovať Poissonovo rozdelenie.

Z kapitoly o diskrétnych náhodných premenných:

26. Poissonovo rozdelenie: porovnanie dát s Poissonovým rozdelením

- **How random was the 1998 World Cup?**
  Getting the Best from Teaching Statistics

Príklad 10

The average number of goals per game scored by the 32 teams in the group matches was 1.3125 (i.e. 48 matches yielded 126 goals: 126/96 = 1.3125).

The actual distribution was as follows:

Goals	Frequency
0	26
1	34
2	24
3	8
4	1
5	2
6	1

A Poisson distribution with $m = 1.3125$ gives expected frequencies:

Goals	Frequency
0	25.8
1	33.9
2	22.3
3	9.7
>3	4.3

NOT A BAD FIT!!!

Parameter **lambda** je strednou hodnotou Poissonovho rozdelenia, preto má zmysel odhadnúť ho ako aritmetický priemer dát.

```
hodnoty = np.array(range(0, 7))
pozorovane = np.array([26, 34, 24, 8, 1, 2, 1])

lambda0 = np.sum(hodnoty*pozorovane)/np.sum(pozorovane)
print(lambda0)
```

```
1.3125
```

Čo budeme teraz robiť:

- Najskôr spravíme chí-kvadarát test pre možnosti **0, 1, 2, 3, 4+** tak, ako je to v tabuľke. Uvidíme, že aj takýto odhad parametra nám bude stačiť na to, aby sme hypotézu o Poissonovom rozdelení nezamietli.
- Potom to zopakujeme pre **0, 1, 2, 3+.** V predchádzajúcom bode sme tesne pod hranicou 5 pre očakávané početnosti. Rovnako sa Poissonovo rozdelenie nezamietne

(1) *Chí-kvadrát test pre triedy* ***0, 1, 2, 3, 4+*** **:**

```
from scipy.stats import poisson

pozorovane_upravene = np.r_[pozorovane[:4], sum(pozorovane[4:])]
print(f"pozorované početnosti: {pozorovane_upravene}")

n = np.sum(pozorovane_upravene)
ocakavane = np.r_[poisson.pmf(range(0,4), lambda0), 1 - poisson.cdf(3, lambda0)]*n
print(f"očakávané početnosti: {ocakavane}")
```

```
pozorované početnosti: [26 34 24  8  4]
očakávané početnosti: [25.83804948 33.91243994 22.25503871  9.73657944  4.25789244]
```

Na výpočet testovacej štatistiky môžeme použiť funkciu **chisquare**. V tomto prípade však je **počet stupňov voľnosti rovný 3, lebo odhadujeme jeden parameter.** Bez odhadovaného parametra by pri 5 triedach pre dáta boli 4 stupne voľnosti (o 1 menej ako počet tried). Takto sa znižuje ešte o počet odhadovaných parametrov, teda v našom prípade o 1. V Pythone to zadáme ako parameter **ddof** *(delta degrees of freedom*), vyjadrujúci, o koľko sa musí znížiť počet stupňov voľnosti kvôli odhadovaným parametrom.

```
from scipy.stats import chisquare
from scipy.stats import chi2

statistika, p_hodnota = chisquare(pozorovane_upravene, ocakavane, ddof=1)
print(f"štatistika: {statistika:.5f}")
print(f"p hodnota: {p_hodnota:.5f}")
```

```
štatistika: 0.46341
p hodnota: 0.92685
```

Porovnáme s ručným výpočtom p-hodnoty pomocou distribučnej funkcie:

```
print(f"p-hodnota pomocou distribučnej funkcie: {(1 - chi2.cdf(statistika, df=3)):.5f}")
```

```
p-hodnota pomocou distribučnej funkcie: 0.92685
```

Hypotéza o Poissonovom rozdelení sa znovu nezamieta.

Ak nás zaujíma kritická hodnota (hypotéza sa zamieta pre hodnoty štatistiky väčšie ako je táto kritická hodnota):

```
print(f"kritická hodnota: {chi2.ppf(0.95, df=3):.5f}")
```

```
kritická hodnota: 7.81473
```

(2) *Chí-kvadrát test pre triedy* ***0, 1, 2, 3+*** **:**

```
pozorovane_upravene = np.r_[pozorovane[:3], sum(pozorovane[3:])]
print(f"pozorované početnosti: {pozorovane_upravene}")

n = np.sum(pozorovane_upravene)
ocakavane = np.r_[poisson.pmf(range(0,3), lambda0), 1 - poisson.cdf(2, lambda0)]*n
print(f"očakávané početnosti: {ocakavane}")
```

```
pozorované početnosti: [26 34 24 12]
očakávané početnosti: [25.83804948 33.91243994 22.25503871 13.99447187]
```

Testovanie:
Zmenil sa počet stupňov voľnosti. Máme 4 triedy pre dáta, 1 odhadovaný parameter, teda počet stupňov voľnosti je 4-1-1=2.

```
statistika, p_hodnota = chisquare(pozorovane_upravene, ocakavane, ddof=1)
print(f"štatistika: {statistika:.5f}")
print(f"p-hodnota: {p_hodnota:.5f}")
```

```
štatistika: 0.42231
p-hodnota: 0.80965
```

Hypotéza o Poissonovom rozdelení sa znovu nezamieta.

7. Kolmogorovov-Smirnovov test - jednovýberový

Predpoklady:

- **testujeme zhodu dát s daným spojitým rozdelením**
- **parametre sa neodhadujú.**

Príklad 4

Máme nasledovný algoritmus, v ktorom sa generujú nezávislé náhodné premenné **Y** z rovnomerného rozdelenia na (0, 1) a potom sa postupuje nasledovne:

Select a random number Y_1, then another one Y_2, etc., continuing as long as the numbers so far chosen form a decreasing sequence, i.e.,

$$Y_1 > Y_2 > \cdots > Y_n < Y_{n+1}$$

As soon as a random number is drawn that exceeds its immediate predecessor, the process halts.

If n is even, forget the entire sequence of random numbers, call this experiment a failure, and repeat from the beginning.

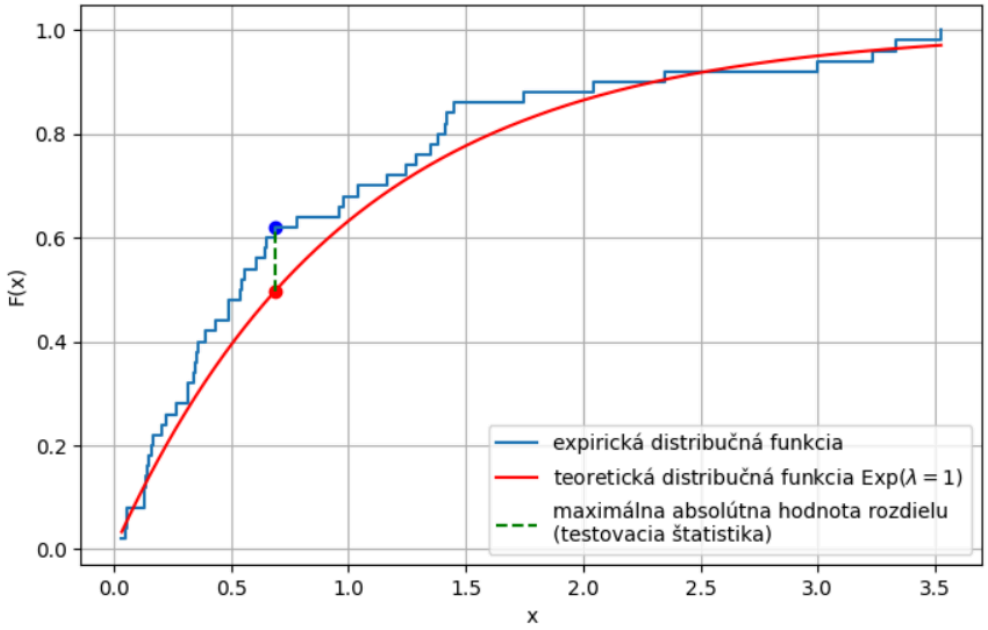
If, however, n is odd then we're ready for output. The number X that will be output is just this: the integer part of X is the number of failures so far seen, and the fractional part of X is Y_1, the largest (and first) member of the last decreasing sequence chosen.

For example, if we choose random numbers .63, .19, .37 we would record a failure and start again. If the second sequence is .44, .91 then we are ready for output, and the number that is selected by these events is $X = 1.44$.

Naprogramovali sme algoritmus, nechali ho dostatočne veľakrát zbehnúť a analyzujeme výsledky. Histogram pripomína exponenciálne rozdelenie, priemer aj výberová disperzia je "podozrivo" blízko k jednotke. To nasvedčuje tomu, že by mohlo ísť o exponenciálne rozdelenie so strednou hodnotou 1.

Mohli by sme to otestovať chí-kvadrát testom dobrej zhody, dajú sa vytvoriť intervaly a ich očakávané početnosti. Inou možnosťou je použiť **Kolmogorovov-Smirnovov test**.

Jeho základnou myšlienkou je **zostrojenie empirickej distribučnej funkcie z dát a jej porovnanie s distribučnou funkciou testovaného rozdelenia.** Testovacou štatistikou je maximum z absolútnej hodnoty ich rozdielu, teda ich maximálna vzdialenosť.



Pravdepodobnostné rozdelenie testovacej štatistiky (pre veľké **n**) nie je úplne jednoduché, ale podstatné je, že nezávisí od testovaného rozdelenia. Takže sa dajú počítať p-hodnoty.

V Pythone:
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ks_1samp.html

```
from scipy.stats import ks_1samp, expon

statistika, p_hodnota = ks_1samp(data, cdf=expon.cdf)
print(f"štatistika: {statistika:.5f}")
print(f"p-hodnota: {p_hodnota:.5f}")
```

```
štatistika: 0.12317
p-hodnota: 0.40155
```

Vidíme, že v tomto prípade hypotézu o exponenciálnom rozdelení so strednou hodnotou 1 nezamietame.

Viac o výstupe z testu:

```
vystup = ks_1samp(data, cdf=expon.cdf)
print(f"štatistika: {vystup.statistic:.5f}")
print(f"p-hodnota: {vystup.pvalue:.5f}")
print(f"poloha maximálnej odchýlky: {vystup.statistic_location:.5f}")
print(f"znamienko maximálnej odchýlky: {vystup.statistic_sign}")
```

```
štatistika: 0.12317
p-hodnota: 0.40155
poloha maximálnej odchýlky: 0.68683
znamienko maximálnej odchýlky: 1
```

8. Lillieforsov test

Kolmogorovov-Smirnovov test predpokladá, že sa parametre rozdelenia neodhadujú z dát, ale sú presne dané. Napríklad v našom príklade sme testovali exponenciálne rozdelenie s parametrom rovným jednej, nie s parametrom **lambda** rovným prevrátenej hodnote priemeru.

Lillieforsov test je modifikáciou Kolmogorovovho-Smirnovovho test, ktorý **testuje, či dáta pochádzajú z normálneho rozdelenia.** Využíva podobnú myšlienku ako pôvodný KS test, ale **nevyžaduje dopredu znalosť parametrov.**

Príklad 5 Máme vygenerované dve sady dát. Zobrazíme:

- pomocou **Q-Q plotu** ich porovnanie s normálnym rozdelením

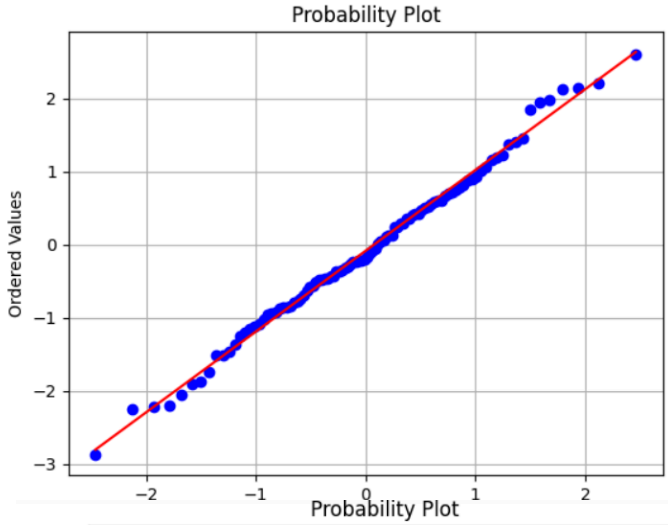
```
from scipy.stats import probplot

probplot(data1, dist="norm", plot=plt)
plt.grid(True)
plt.show()
```

- výsledky Lillieforsovho testu

```
from statsmodels.stats.diagnostic import lilliefors
```

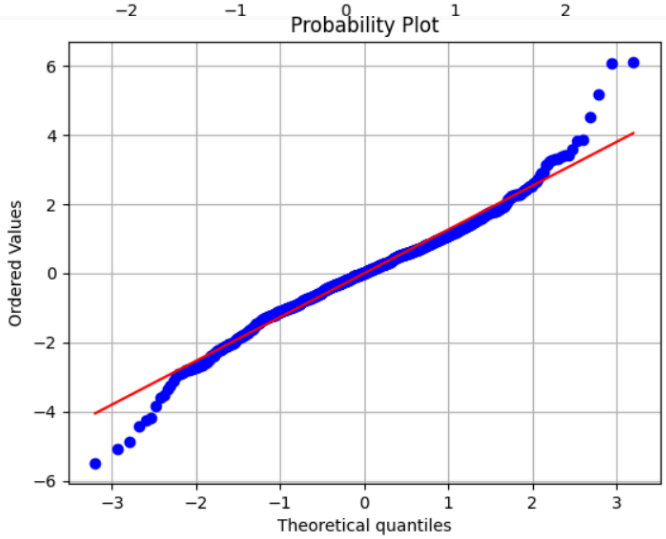
```
statistika, p_hodnota = lilliefors(data1)
print(f"štatistika: {statistika:.5f}")
print(f"p-hodnota: {p_hodnota:.5f}")
```

Kvôli výstižnosti necháme teraz defaultné popisy osí v Q-Q grafe.



Lillieforsov test:

```
štatistika: 0.04333
p-hodnota: 0.92201
```



Lillieforsov test:

```
štatistika: 0.03950
p-hodnota: 0.00168
```

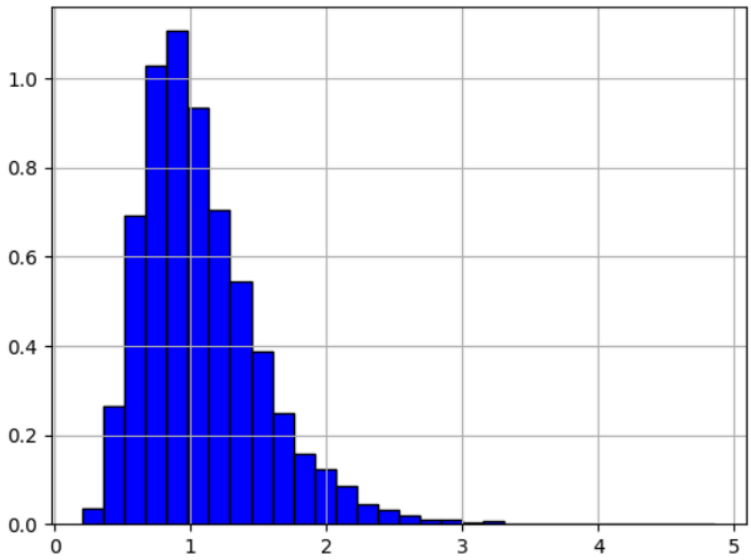
9. Iné testy normálneho rozdelenia

Uveďme aspoň niektoré:

- **Shapiro-Wilkov test**
  - Na rozdiel od predchádzajúcich testov sa **dá použiť aj v prípade malého rozsahu dát.**
  - https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html
- **D'Agostinov-Pearsonov test**
  - Založený je na šikmosti a špicatosti (tretí a štvrtý moment), ktoré porovnáva s hodnotami pre normálne rozdelenie.
  - https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html

Pojem **šikmosti** sa často používa na popis rozdelenia – hovorí sa napríklad, že dáta sú **zošikmené doprava**.

Na obrázku sú dáta vygenerované z **lognormálneho rozdelenia**. Náhodná premenná má lognormálne rozdelenie, ak jej logaritmus má normálne rozdelenie.



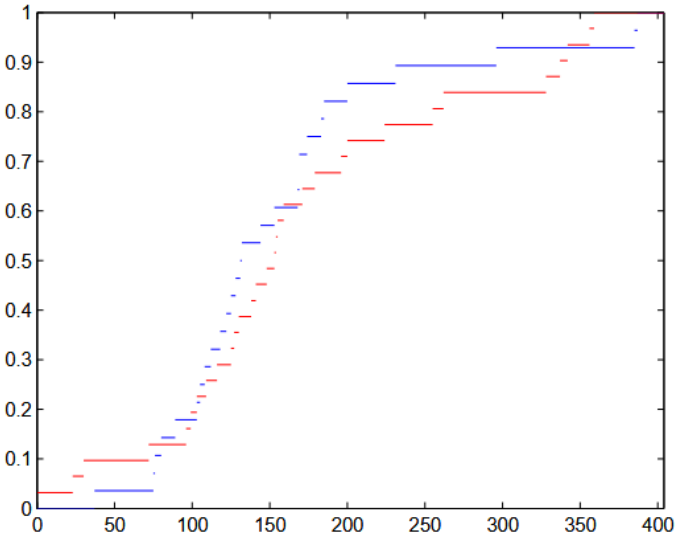
10. Kolmogorovov-Smirnovov test - dvojvýberový

Testujeme, či dve sady dát pochádzajú z toho istého rozdelenia.
Pritom sa predpokladá, že ide o spojité pravdepodobnostné rozdelenie.

Základná myšlienka testu je analogická ako v prípade jednovýberového testu – zostroja sa empirické distribučné funkcie a testovacou štatistikou je maximálna absolútna hodnota ich rozdielu.

Príklady:

- V príklade s programovaním algoritmu a zisťovaní pravdepodobnostného rozdelenia si chcú dvaja študenti skontrolovať výpočet. Nechcú pritom zdieľať kód, a tak sa rozhodnú, že si pošlú len vygenerovaný výstup. Ak budú oba výstupy z toho istého rozdelenia, zrejme to majú naprogramované správne – pravdepodobnosť, že nezávisle spravia chybu, ktorá vedie k rovnakému nesprávnemu rozdeleniu, je malá.
- Katarína Kocsisová, bakalárska práca *Experimenty a vlastné dáta pri vyučovaní pravdepodobnosti a štatistiky*, 2015: porovnávanie času čakania na obed v dvoch jedálňach na internátoch.



Obr. 6: Distribučné funkcie 1. jedálne (modrou farbou) a 2. jedálne (červenou farbou)

Hypotéza o zhode pravdepodobnostných rozdelení sa v tomto prípade nezamietla.

V Pythone: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ks_2samp.html