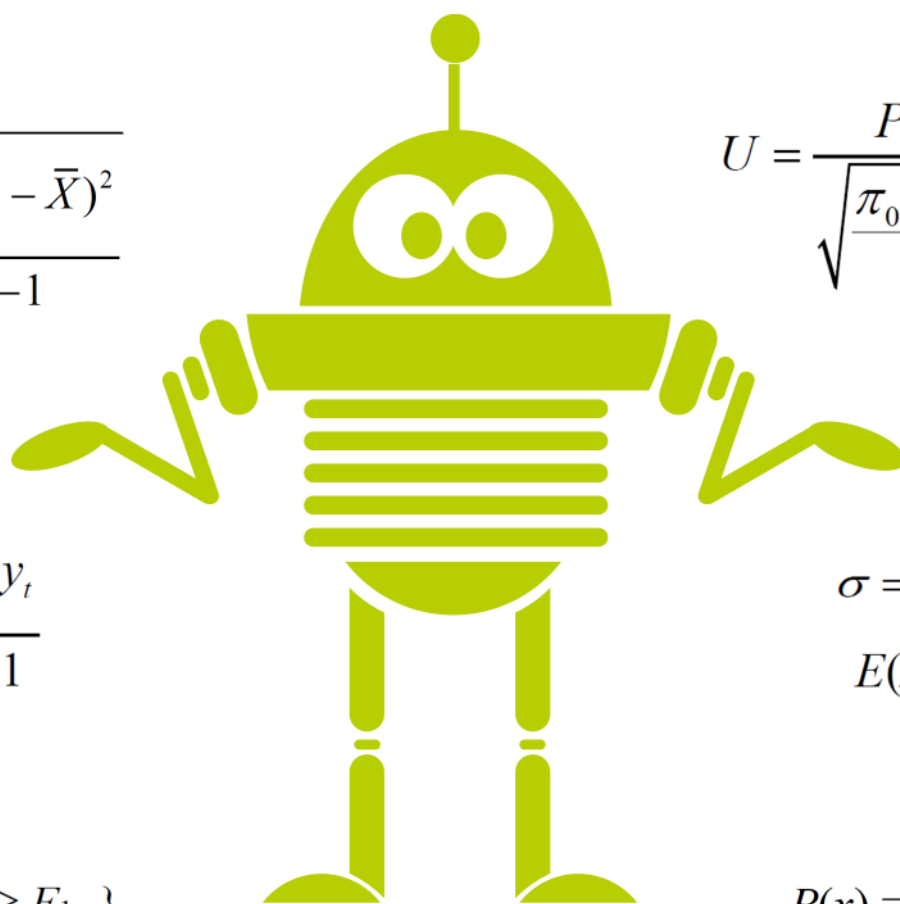


STATISTIKA

VŠTE V ČESKÝCH BUDĚJOVICÍCH PRVNÍ ČÁST STUDUJNÍHO PLÁNU

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

$$U = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$$



$$\bar{\Delta}y = \frac{\sum_{t=2}^T \Delta y_t}{T-1}$$

$$\sigma = \sqrt{D(X)}$$

$$E(X) = \pi$$

$$W_\alpha = \{F; F \geq F_{1-\alpha}\}$$

$$P(x) = P(X = x)$$



EDU FOR LIFE

VZDĚLÁNÍ, KTERÉ SE TI BUDE HODIT

Adriana Řeháčková
www.statistickyneklasicky.cz

STATISTIKA

VŠTE V ČESKÝCH BUDĚJOVICÍCH

Ukážeme si, že statistika není žádná nuda a že má opravdu smysl! Cílem kurzu je, abys pochopil základní až lehce pokročilé statistické metody. Nemine Tě ani osvojení Excelu a orientace ve vzorcích. Po kurzu bys měl vědět, co, kdy a jak použít a dokázat tyto znalosti využít na reálných datech.

OBSAH KURZU:

Úvod do statistiky

Četnosti

Charakteristiky polohy

Charakteristiky variability

Indexy - jednoduché a složené

Pravděpodobnost - základní

Náhodná veličina - nespojitá a spojitá

Binomické rozdělení, Poissonovo rozdělení, Hypergeometrické rozdělení

Normální rozdělení

Bodové a intervalové odhady

Testování statistických hypotéz

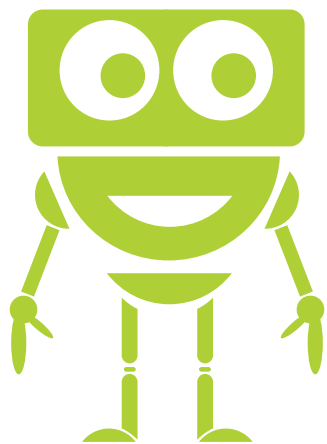
O TOMTO MATERIÁLU:

Žádná část tohoto materiálu nesmí být nijak použita či reprodukována bez písemného svolení autora.

Copyright ©Statistickyneklasicky 2020

Autor materiálu: Adriana Řeháčková

Sazba a grafické úpravy: Adriana Řeháčková



STATISTICKY NEKLASICKY

Statistika - analýza velkého množství dat - třídění, hledání souvislostí a odhady pravděpodobností. Zkoumá, zpracovává a vyhodnocuje data.

Statistický soubor - skupina prvků (počet lidí v ČR, studentů VŠE, maturantů, skleniček v baru,...)

Statistická jednotka - konkrétní prvek (jeden občan, student, sklenička,...)

Statistický znak - To s čím pracujeme, co nás zajímá, co chceme měřit.

1) Identifikační znak - společná vlastnost pro všechny jednotky (občanství, ČVUT, ...)

2) Statistická proměnná -

2a) Slovní (kvalitativní): vyjádřená slovně

i) Nominální (nelze seřadit podle pořadí) - barva, zaměstnání, muž/žena, ano/ne

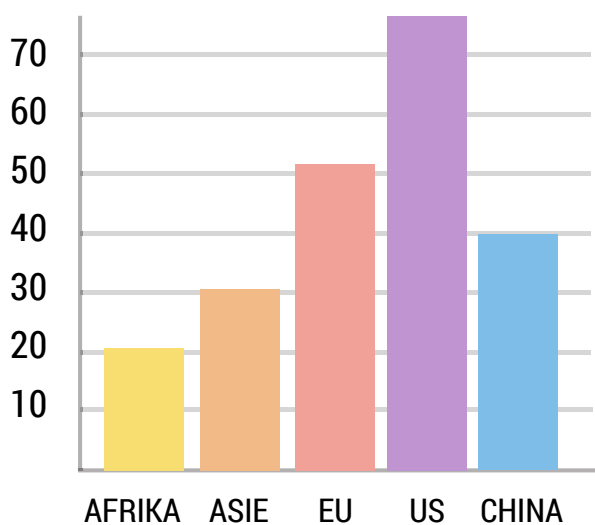
ii) Ordinální (lze seřadit, ale nelze říct o kolik) - rok narození, známka ve škole, ukončené vzdělání

2b) Číselná (kvantitativní): vyjádřená čísly

i) Spojité - hodnoty z konečného nebo nekonečného intervalu (výška, váha, měs. příjem)

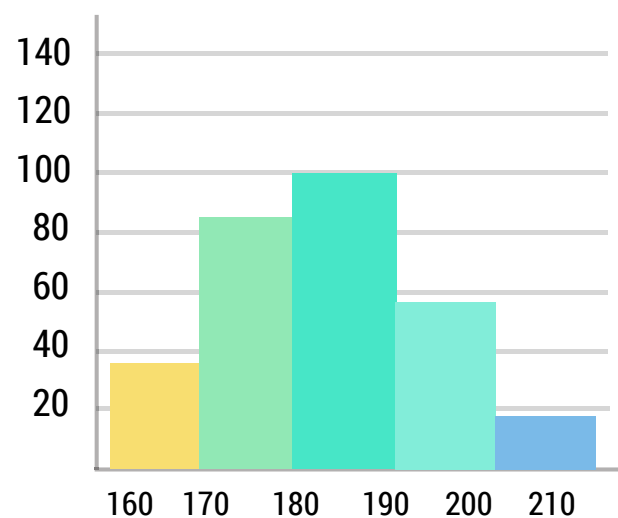
ii) Nespojité (diskrétní) - hodnoty z malého počtu jednoznačně izolovaných hodnot (počet let ve škole, počet dětí, počet elektronických zařízení)

Obrat (USD)



sloupcový graf

Počet zaměstnanců



histogram



Četnosti: Kolikrát se nějaká proměnná vyskytne v tabulce, nebo kolik %.

1) Doplň tabulku četností:

Barevné kuličky	n_i	N_i	p_i	P_i
červená	5			
modrá	2			
zelená	4			
žlutá	3			
oranžová	6			
Σ				

2) Doplň chybějící hodnoty:

Věk	n_i	p_i (%)	N_i
18	432		
19		4,176	
20			13405
21	10 501		
22	9 116		
23		14, 986	
24			
Σ	44 782		



1) Máme soubor sportovců, kteří budou reprezentovat ČR na Olympijských hrách. Zjistěte průměrný věk sportovců, medián a modus jejich věků.

Zjištěné věky jsou následující: 18, 19, 28, 24, 29, 31, 19, 23.

2) Řidič autobusu pražské MHD si měří svoji průměrnou rychlost na své pravidelné 10km dlouhé trase a získal následující výsledky: První jízdu jel rychlostí 57 km/h, druhou a třetí stejně dlouhou, jel rychlostí 43 km/h a 52 km/h. Jaká je jeho průměrná rychlost na základě získaných údajů?

b) Uvažujte, že úseky nebyly stejně dlouhé, ale měřily postupně: 15 km, 10 km a 12 km. Jaká je jeho průměrná rychlost na celkové trase?

3) HDP v USA posledních 5 let rostlo (jednou pokleslo) následovně: 4%, 3%, 1%, -2%, 1%. Jaký je průměrný růst HDP?



- vlastnosti aritmetického průměru
 - $\overline{x+k} = \bar{x} + k$
 - k = nějaká libovolná konstanta
 - jestliže ke každé hodnotě v souboru přičteme stejnou nenulovou konstantu, tak se o tuto konstantu zvýší také aritmetický průměr
 - $\bar{k} = k$
 - aritmetický průměr konstanty je konstanta
 - $\overline{x \cdot k} = \bar{x} \cdot k$
 - jestliže všechny hodnoty v souboru vynásobíme stejnou nenulovou konstantou k , tak se aritmetický průměr také zvýší k -násobně
 - - $\sum_{i=1}^n (x_i - \bar{x}) = 0$
 - součet odchylek hodnot od svého průměru je vždycky nulový
 - $\overline{x \pm y} = \bar{x} \pm \bar{y}$
 - aritmetický průměr součtu nebo rozdílu dvou znaků je roven součtu nebo rozdílu průměrů těchto znaků
- vlastnosti rozptylu
 - $s_{x+k}^2 = s_x^2$
 - jestliže ke každé hodnotě v souboru přičteme stejnou nenulovou konstantu k , rozptyl se nezmění
 - $s_k^2 = 0$
 - rozptyl konstanty je nulový
 - $s_{x \cdot k}^2 = k^2 \cdot s_x^2$
 - jestliže každou hodnotu v souboru vynásobíme stejnou nenulovou konstantou k , rozptyl se zvýší k^2 násobně
 - $s_{x \pm y}^2 = s_x^2 + s_y^2$
 - rozptyl součtu nebo rozdílu dvou nezávislých znaků je roven vždycky a pouze **součtu** rozptylů
- věta o rozkladu rozptylu
 - jestliže máme soubor rozdělen do k podskupin, u kterých známe průměr a rozptyl (a samozřejmě četnosti), anebo jsme schopni si je jednoduše dopočítat, pak rozptyl celého souboru spočítáme jako součet dvou složek – meziskupinové a vnitroskupinové variability

1) Jak se změní průměrná mzda pracovníka firmy 28 000 Kč, dojde-li k snížení mezd u každého pracovníka o 14 %?

2) Vypočítali jsme rozptyl 5,4. Vypočtěte nový rozptyl poté, co každou pozorovanou hodnotu:

- a) zvýšíme o 12
- b) ztrojnásobíme
- c) zvýšíme o 15%

3) Sledovali jsme proměnnou x a dodatečně jsme zjistili chyby u dvou jednotek. Místo 85 má být správně 95 a místo 120 má být správně 150. Ostatních 18 údajů je správných. Opravte vypočítané charakteristiky, byl-li aritmetický průměr 110 a rozptyl 800.



4) Průměrná mzda pracovníků je 23 000 Kč a variační koeficient je 0,4. Jak se změní průměrná mzda a variační koeficient, jestliže každý pracovník dostane přidáno 2 000 Kč?

5) 6 malých dětí nám odpovědělo na výši jejich měsíčního kapesného. Údaje o výši kapesného jsou následující: 80 Kč, 100 Kč, 90 Kč, 75 Kč, 80 Kč, 70 Kč.

Určete průměr, variační rozpětí, rozptyl, směrodatnou odchylku a variační koeficient.

[Výpočet rozptylu v Excelu: =VAR.P(...)]

DOMÁCÍ ÚKOL: Na naší škole máme volejbalové družstvo dívek a volejbalové družstvo chlapců. Dívky ve volejbalovém družstvu mají výšky v řadě od největší po nejmenší : 194 cm, 192cm, 175 cm, 175 cm, 174 cm, 172 cm, 171 cm, 170 cm.

Určete : a) aritmetický průměr; b) medián, c) modus d) směrodatnou odchylku, e) variační koeficient

[177,875; 174,5; 175; 8,908; 5%]



Individuální indexy (indexy stejnorodých ukazatelů):

- a) jednoduché - Zkoumáme jeden výrobek na jedné pobočce
- b) složené - Zkoumáme jeden výrobek na více pobočkách (za více poboček dohromady)

Souhrnné indexy (indexy nestejnorodých ukazatelů):

- a) Paascheho – fixuje druhou veličinu v běžném období
- b) Laspayresův – fixuje druhou veličinu v základním období
- c) Fishrerův – geometrický průměr Paascheho a Laspayresova

Ve 3 prodejnách byly zjištěny údaje o prodeji pomerančů za měsíc listopad a prosinec. Určete absolutní i procentní změnu v tržbách za pomeranče mezi listopadem a prosincem za všechny prodejny dohromady.

Prodejna	Cena za 1 kg (Kč) listopad	index ceny	Prodané množství (kg) listopad	index množství
Orange	21	1,04	430	1,05
Pomeranč	23	1,15	500	0,8
Oranžová	19	1,08	490	0,96

1) O kolik % se změnil celkové tržby?

- a) Vzrostly o 0,8%
- b) Vzrostly o 8%
- c) Nezměnily se
- d) Snížily se

2) Jaká byla absolutní změna tržeb?

- a) 283,56 Kč
- b) 253,368 Kč
- c) Těžko říct
- d) Tržby se snížily

3) Urči index tržby za prodejnu Orange:

- a) 1,092
- b) 0,92
- c) 9,2
- d) nelze určit

Prodejna	Cena za 1 kg (Kč) listopad	Cena za 1 kg (Kč) prosinec	index ceny	Prodané množství (kg) listopad	Prodané množství (kg) prosinec	index množství	p1q1	p0q0
Orange	21	21,84	1,04	430	451,5	1,05	9860,76	9030
Pomeranč	23	26,45	1,15	500	400	0,8	10580	11500
Oranžová	19	20,52	1,08	490	470,4	0,96	9652,608	9310
							30093,37	29840
							Tržba: 1,008491	
							Tržba(absolutní): 253,368	



Pohybuje se od 0-100%, 0 - jev nemožný, 100 - jev jistý.

Náhodný pokus = takový pokus, jehož výsledek je ovlivněn náhodou (loterie, hod mincí,..)

Závislé a nezávislé jevy:

Závislé jevy - Výsledek druhého jevu závisí na výsledku prvního jevu (tahání výherců, loterie).

Nezávislé jevy - Výsledek jednoho jevu neovlivní výsledek druhého jevu (hod kostkou, mincí).

Průnik a sjednocení jevů:

Průnik - $P(A \cap B)$

Nezávislé jevy:

Závislé jevy:

Jsem na párty a zítra v 8:00 půjdu do školy, pokud mě vzbudí budík a probudím se ve své posteli. Pravděpodobnost, že budík zazvoní je 0,7 a pravděpodobnost, že se probudím ve své posteli je 0,55 Jaká je pravděpodobnost, že do školy skutečně půjdu?

Sjednocení (nezávislé jevy) - $P(A \cup B)$

Slučitelné jevy:

Neslučitelné jevy:

Pravděpodobnost, že neudělám zkoušku z matematiky je 0,15, že neudělám z ekonomie je 0,25. Stačí, když jeden z nich neudělám a vyhodí mě ze školy. Jaká je pravděpodobnost, že ve škole skončím?



1) Házíme hrací kostkou. Určete pravděpodobnosti těchto jevů:

a) A při jednom hoďu padne šestka:

b) Při jednom hoďu padne sudé číslo:

c) Při dvou hoďech padnou obě šestky:

d) Při dvou hoďech padnou stejné počty:

e) Při dvou hoďech padne alespoň jednu šestku:

2) Elektronická součástka má výrobní vadu s pravděpodobností 0,001. Zákazník si v obchodě zakoupí 2 součástky. Jaká je pravděpodobnost, že:

- a) bude alespoň 1 vadná,
- b) bude právě jedna vadná,
- c) budou obě funkční

3) V osudí je 5 bílých a 3 černé koule. Opakovaně losujeme 3 koule tak, že je do osudí nevracíme. Jaká je pravděpodobnost, že:

- a) vytáhneme 2krát po sobě bílou kouli?
- b) vytáhneme 3krát po sobě kouli stejné barvy?
- c) při 3 losováních vytáhneme alespoň jednu černou kouli?



Nespojitá náhodná veličina:

Pravděpodobnostní funkce ($P(x)$) a Distribuční funkce ($F(x)$).

Spojitá náhodná veličina:

Distribuční funkce ($F(x)$) a hustota pravděpodobnosti ($f(x)$).

1) Pravděpodobnostní rozdělení náhodné veličiny X je dáno následující tabulkou:

x	$P(x)$
1	0,15
2	0,25
3	0,5
4	0,09
5	0,01

a) Spočítejte pravděpodobnosti:

$$P(X = 4)$$

$$P(1 < X \leq 3)$$

$$P(X > 1)$$

b) Spočítejte střední hodnotu, rozptyl a směrodatnou odchylku:

c) Určete distribuční funkci:

2) V osudí je 7 bílých a 3 černé míčky. Náhodně vytahujeme po jednom míčku tak dlouho, dokud nevytáhneme bílý míček. (Vytažené míčky do osudí nevracíme.) Náhodná veličina X nechť udává počet potřebných tahů.

Určete pravděpodobnostní funkci $P(x)$ a distribuční funkci $F(x)$.



3) Necht' náhodná veličina X má hustotu pravděpodobnosti:

$$f(x) = \frac{3}{8}x^2, 0 < x < 2$$

$$f(x) = 0, \text{ jinak}$$

Pro tuto náhodnou veličinu určete:

a) Distribuční funkci,

b) střední hodnotu,

c) rozptyl,

d) $P(1 < X < 1,5)$

DOMÁCÍ ÚKOL: Náhodná veličina X má pravděpodobnostní funkci:

$$P(x) = \frac{6-x}{15} \quad \text{pro } x = 1, 2, \dots, 5$$

$$P(x) = 0 \quad \text{jinak}$$

a) Vyjádřete pravděpodobnostní funkci:

b) Určete distribuční funkci, vypočítejte střední hodnotu, rozptyl a směrodatnou odchylku:

[suma prav. funkce musí být 1; 2,333; 1,557; 1,248]



○ alternativní rozdělení

- značení: $X \sim A(\pi)$
- **použití:** v případě jednoho pokusu, během kterého je sledován výskyt náhodného jevu, který může nastat s pravděpodobností π (nebo nenastane)
- $X \rightarrow$ počet výskytů jevu během jednoho pokusu
- pravděpodobnostní funkce
$$P(x) = \pi^x(1 - \pi)^{1-x} \quad x = 0, 1; 0 < \pi < 1$$
- **střední hodnota**
 $E(X) = \pi$
- **rozptyl**
 $D(X) = \pi(1 - \pi)$

○ binomické rozdělení

- označení: $X \sim Bi(n; \pi)$ (= tvrzení, že náhodná veličina X se řídí binomickým rozdělením se dvěma parametry)
- **použití:** u náhodné veličiny, která představuje počet výskytů sledovaného jevu během **n vzájemně nezávislých pokusů a parametr π představuje konstantní pravděpodobnost výskytu sledovaného jevu v každém z pokusů**
- $X \rightarrow$ počet výskytů jevu během n nezávislých pokusů
- pravděpodobnostní funkce
$$P(x) = \binom{n}{x} \pi^x(1 - \pi)^{n-x} \quad x = 0, 1, 2, \dots, n; n > 0, 0 < \pi < 1$$
- střední hodnota
 $E(X) = n\pi$
- rozptyl
 $D(X) = n\pi(1 - \pi)$

○ Poissonovo rozdělení

- označení: $X \sim Po(\lambda)$
- použití: sleduje se výskyt jevů v **časovém intervalu dané délky** (za určitých podmínek), u nezávislých pokusů pro tzv. řídké jevy s malou pravděpodobností výskytu ($n > 30 \wedge \pi \leq 0,1$)
- $X \rightarrow$ počet výskytů jevů během n pokusů (resp. v intervalu)
- pravděpodobnostní funkce ($\lambda =$ střední hodnota počtu výskytů náhodné veličiny $-\pi$)
$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x = 0, 1, \dots; \lambda > 0$$
- střední hodnota
 $E(X) = \lambda$
- rozptyl
 $D(X) = \lambda$

○ hypergeometrické rozdělení

- označení: $X \sim Hy(N, M, n)$ (= tvrzení, že náhodná veličina X se řídí přibližně hypergeometrickým rozdělením se **třemi parametry M, N, n** ($N =$ celkový počet prvků, $M =$ počet prvků se sledovanou vlastností, $n =$ rozsah výběru))
- použití: **v případě závislých pokusů** (například pokusů bez vracení)
- $X \rightarrow$ počet prvků se sledovanou vlastností během závislých pokusů
- pravděpodobnostní funkce
$$P(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$
- střední hodnota
 $E(X) = n \frac{M}{N}$
- rozptyl
$$D(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$$



1) Jaké je pravděpodobnost, že z deseti kol losování rulety obdržíme:

- a) dvakrát číslo červené barvy,
- b) 5 krát liché číslo,
- c) maximálně dvakrát číslo dělitelné 9?

a) =BINOMDIST(2;10;18/37;NEPRAVDA)

b) =BINOMDIST(5;10;18/37;NEPRAVDA)

c) =BINOMDIST(2;10;4/37;PRAVDA)

2) Na stůl bude volně vysypáno 10 kostek. Jaká je pravděpodobnost, že:

- a) na 4 z nich padne šestka;
- b) určete očekávaný počet kostek, na kterých padne šestka. Jaký je rozptyl počtu kostek, na kterých padne 6?

DOMÁCÍ ÚKOL: Šestka padne nejvýše na 3 z nich?

[0,161 + 0,323 + 0,2907 + 0,155 = 0,93]



V matematice psali studenti 6 testů. Naše náhodná veličina X udává počet úspěšně napsaných testů z matematiky. Pravděpodobnost úspěšnosti studenta v testu z matematiky je vždy 0,85. Zjistěte rozdělení veličiny X a určete pravděpodobnost toho, že:

- a) bude úspěšný právě 3-krát,
- b) bude úspěšný alespoň 2-krát.

DOMÁCÍ ÚKOL: V osudí jsou 4 koule červené barvy, zbylých 16 koulí má jinou barvu. Provedeme výběr 6 koulí s vrácením. Určete pravděpodobnost, že vybereme:

- a) tři červené,
- b) minimálně 3 červené koule,
- c) 1 nebo žádnou červenou kouli,
- d) alespoň 1 červenou kouli.

[a) 0,08192; b) 0,09888; c) 0,65536; d) 0,737856]



1) Do obchodu přijde průměrně 60 zákazníků za 1 hodinu. Jaká je pravděpodobnost, že do obchodu nepřijde žádný zákazník během 0,5 min., ve které je prodavač nepřítomen? Určete střední hodnotu a rozptyl.

2) Počet překlepů (před kontrolou pravopisu) na jedné straně textu je v průměru roven 2. S jakou pravděpodobností můžeme na 3 stranách textu očekávat:

- a) právě 6 překlepů,
- b) alespoň 2 překlepy,
- c) 2 nebo 3 překlepy

DOMÁCÍ ÚKOL: Předpokládejme, že realitní makléř jedná v průměru s pěti zákazníky za den. Zjistěte jaká je pravděpodobnost, že počet zákazníků makléře za jeden den bude větší než 4. Určete střední hodnotu a rozptyl.

[Excel: =1-POISSON.DIST(4;5;PRAVDA); 0,56]



HYPERGEOMETRICKÉ ROZDĚLENÍ

1) Z dodávky 500 výrobků je kontrolováno 5 výrobků. Víme, že v dodávce je 30 zmetků. Jaká je pravděpodobnost, že v 5 vybraných výrobcích nebude žádný zmetek?

Excel: =HYPGEOM.DIST(0;5;30;500;NEPRAVDA)

2) Dodávka 60 nosičů CD, které si objednala počítačová firma, obsahuje 4 vadná CD. Jaká je pravděpodobnost, že při vypalování 12 ks CD se:

- a) nepodaří vypálit právě 1 CD,
- b) podaří vypálit všech 12 CD,
- c) nepodaří vypálit minimálně 2 CD.

3) V souboru 10 000 lidí je 100 leváků. Pokud náhodně bez vracení vybereme 50 lidí, jaká je pravděpodobnost, že mezi nimi:

- a) nebude žádný levák,
- b) budou právě 3 leváci,
- c) určí střední hodnotu

DOMÁCÍ ÚKOL: Při slosování sportky je z osudí postupně vylosováno 6 čísel ze 49. Jaká je pravděpodobnost, že při správném tipování:

- a) šesti čísel, získává sázející výhru 1. pořadí,
- b) pěti čísel, získává sázející výhru 2. pořadí,
- c) tří čísel, získává sázející výhru 5. pořadí

[0,000000072; 0,000018021; 0,01765]



1) V osudí je 6 míčků černých, 6 bílých a 6 červených. Náhodně vybereme 6 míčků. Jaká je pravděpodobnost, že mezi vybranými míčky budou všechny míčky černé. Vybíráme bez vracení.

Řešení: 0,0497

2) V loterii je 200 losů, z nichž 10 vyhrává. Jaká je pravděpodobnost, že získáte alespoň jednu výhru, koupíte-li si 20 losů.

Řešení: 0,66

3) Telefonní ústředna zapojí během hodiny průměrně 15 hovorů. Jaká je pravděpodobnost, že během 4 minut zapojí ústředna

- a) právě jeden hovor,
- b) alespoň dva hovory?

Řešení: a) 0,368
b) 0,264

4) Student Ondra má potíže s ranním vstáváním. Proto někdy zaspí a nestihne přednášku, která začíná již v 7:30. Pravděpodobnost, že zaspí, je 0,3. V semestru je 13 přednášek - tzn. 13 nezávislých pokusů dorazit na přednášku včas. Nalezněte pravděpodobnost, že Pepe nestihne přednášku v důsledku zaspání ve více než polovině případů.

Řešení: 0,0624



- normální rozdělení (Gaussovo normální rozdělení)
 - označení: $X \sim N(\mu; \sigma^2)$
 - použití: u náhodné veličiny, jejíž kolísání je způsobeno součtem drobných vzájemně nezávislých jevů
 - Gaussova křivka
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty; -\infty < \mu < \infty; \sigma^2 > 0$$
 - střední hodnota
$$E(X) = \mu$$
 - rozptyl (udává tvar rozdělení)
$$D(X) = \sigma^2$$
 - kvantily
$$x_p = \mu + \sigma u_p$$
 - $P(x_1 \leq X \leq x_2) = P\left(\frac{x_1 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma}\right) = P(u_1 \leq U \leq u_2) = \Phi(u_2) - \Phi(u_1)$
- normované normální rozdělení
 - označení: $U \sim N(0; 1)$; (= normovaná náhodná veličina U má normované normální rozdělení s nulovou střední hodnotou a jednotkovým rozptylem)
$$U = \frac{X - \mu}{\sigma} \quad X \sim N(\mu; \sigma^2)$$
 - střední hodnota
$$E(U) = 0$$
 - rozptyl
$$D(U) = 1$$
 - pravděpodobnostní funkce

Další rozdělení:

- logaritmicko-normální rozdělení
- Chí-kvadrát rozdělení
- rozdělení t (Studentovo)
- rozdělení F (Fisherovo-Snedecorovo)

Normální rozdělení je nejdůležitější pravděpodobnostní rozdělení. Jedná se o model pro rozdělení náhodných chyb měření.

Vzorec je velmi složitý, a proto se používá jeho normovaný tvar. Hodnoty pro normovanou veličinu budeme hledat v tabulkách.



1) Hmotnost výrobních součástek je normálně rozdělená veličina se střední hodnotou 110 gramů a rozptylem 100. S jakou pravděpodobností bude hmotnost součástky menší nebo rovna 115-ti gramům?

2) V letech 1977 - 1982 byli všichni studenti přijati do 1. ročníku PedF UK v Praze psychologicky vyšetřováni. V Amthauerově testu struktury inteligence byl průměr hrubého skóre HS u studentů všeobecně vzdělávacích předmětů 115 a směrodatná odchylka 16. Za předpokladu normálního rozdělení statistického znaku HS určete pravděpodobnost, že:

- a) HS je menší nebo rovno 125,
- b) HS je větší než 100,
- c) HS je mezi 105 a 135.



3) Vyučující dá studentům neohlášený test. Výsledky tohoto testu mají normální rozdělení s průměrem 53%

a směrodatnou odchylkou 14%. Jaká je pravděpodobnost, že

a) student test neudělá, tj. bude mít méně než 40%?

DOMÁCÍ ÚKOL: student test napíše na výbornou, tj. bude mít alespoň 80%?

[0,027 = 2,7%]

4) Výrobky jsou považovány za prvotřídní, pokud odchylka od předepsané délky nepřekročí 3,6 mm. Jestliže odchylka má rozdělení $N(0;9)$, kolik prvotřídních výrobků lze čekat mezi 100 výrobky?

DOMÁCÍ ÚKOL: Předpokládejme, že výčetní tloušťky stromů v určitém porostu mají normální rozdělení. Střední tloušťka je 30 cm, směrodatná odchylka je 5 cm. Celkem bylo měřeno 500 stromů. Určete:

a) kolik stromů je silnějších než 36 cm

b) jaká je pravděpodobnost, že náhodným výběrem vybereme strom silnější než 36 cm

c) kolik stromů leží v rozmezí tlouštěk 25 – 36 cm

[57 stromů; 11,5%; 72,62% (363 stromů)]



NORMÁLNÍ ROZDĚLENÍ

- 5) Výška vysokoškolských chlapců má normální rozdělení s parametry $\mu = 180$ cm a směrodatná odchylka = 6.
- Určete dolní kvartil výšky chlapců.
 - Určete horní kvartil výšky chlapců.
 - V jakém výškovém intervalu symetrickém kolem střední hodnoty se nachází polovina vysokoškolských chlapců.
- 6) Zjistěte pravděpodobnost náhodné veličiny U , která má normované normální rozdělení a zakreslete.
- $P(U > 1)$
 - $P(U = 1)$
 - $F(-2)$
 - kvantil: $u_{0,95}$
 - kvantil: $u_{0,15}$



BODOVÝ A INTERVALOVÝ ODHAD

Bodový a intervalový odhad

Používáme tehdy, když je rozsah souboru tak velký, že je velmi obtížné zjistit skutečný stav.

Například

u testování součástí nemůžeme otestovat všechny, ale jen nějaký vzorek, tzv. výběrový soubor. V zásadě nám jde o to, zjistit buď rozptyl, nebo střední hodnotu základního souboru, pokud známe hodnoty nějakého výběru n prvků.

bodový odhad: odhaduje danou charakteristiku jako jedno číslo (například bodovým odhadem střední hodnoty může být průměr naměřených hodnot, bodovým odhadem rozptylu může být výběrový rozptyl).

$$\text{výběrový průměr: } \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \qquad \text{výběrová směrodatná odchylka: } S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

Intervalový odhad: není číslo, ale hned celý interval, ve kterém by měla daná charakteristika (například střední hodnota) ležet s určitou velkou pravděpodobností (pokud je tato pravděpodobnost například 0,95, pak hovoříme o 95% intervalu spolehlivosti). Přesnost, se kterou udáváme výsledek, nazýváme konfidenční interval, a značíme ho jako $1 - (\alpha)$.

$$P\left(\bar{X} - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) \mid P\left(P - u_{1-\alpha/2} \sqrt{\frac{P(1-P)}{n}} < \pi < P + u_{1-\alpha/2} \sqrt{\frac{P(1-P)}{n}}\right)$$

$$\text{standardní chyba: } \frac{\sigma}{\sqrt{n}} \mid \sqrt{\frac{P(1-P)}{n}}$$

$$\text{výběrová chyba: } u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \mid u_{1-\alpha/2} \sqrt{\frac{P(1-P)}{n}}$$

1) Při výrobě nových vlaků chce společnost orientačně zjistit, jaká je průměrná výška cestujících. Náhodně bylo vybráno 20 cestujících s následujícími údaji o výšce v cm. Předpokládejme, že výška cestujících má normální rozdělení. Odhadněte střední výšku a vypočítejte směrodatnou chybu tohoto odhadu.

169	168	174	181	160	155	201	175	165	164
151	177	185	155	172	197	192	184	156	167

EXCEL

Průměr: =PRŮMĚR(A1:J2)

Směrodatná odchylka: =SMODCH.VÝBĚR.S(A1:J2)



BODOVÝ A INTERVALOVÝ ODHAD

2) Útvar kontroly podniku Edison testoval životnost žárovek. Kontroloři vybrali z produkce podniku náhodně 50 žárovek a došli k závěru, že průměrná doba života (přesněji řečeno výběrový průměr doby života) těchto 50 žárovek je 950 hodin a příslušná směrodatná odchylka doby života je 100 hodin. (Předpokládejte, že životnost žárovek lze modelovat normálním rozdělením.)

- Se spolehlivostí 95% určete intervalový odhad střední životnosti žárovek firmy Edison.
- Se spolehlivostí 95% určete intervalový odhad střední životnosti žárovek firmy Edison, pokud směrodatná odchylka nebyla známá, ale zjistili jsme výběrovou hodnotu směrodatné odchylky, která je 80 hodin.

EXCEL

- `=CONFIDENCE.NORM(0,05;100;50)`
- `=CONFIDENCE.NORM(0,05;80;50)`

3) Při výrobě nových vlaků chce společnost orientačně zjistit, jaká je průměrná výška cestujících. Náhodně bylo vybráno 20 cestujících s následujícími údaji o výšce v cm. Předpokládejme, že výška cestujícího má normální rozdělení.

- Určete v jakých mezích lze se spolehlivostí 99% očekávat výšku cestujících.
 - Určete horní hranici 95% pravostranného intervalu spolehlivosti.
- DOMÁCÍ ÚKOL: Určete v jakých mezích lze se spolehlivostí 95% očekávat výšku cestujících.
pozn. Využijte vypočítané hodnoty z prvního příkladu.

169	168	174	181	160	155	201	175	165	164
151	177	185	155	172	197	192	184	156	167

Řešení: $P(165,7 < \mu < 179,1)$

EXCEL

- `=CONFIDENCE.T(0,01;14,28;20)`
- `=CONFIDENCE.T(0,1;14,28;20)`

BODOVÝ A INTERVALOVÝ ODHAD

4) Při kontrole data spotřeby určitého druhu masové konzervy ve skladech produktů masného průmyslu bylo náhodně vybráno 320 z 20 000 konzerv a zjištěno, že 59 z nich má prošlou záruční lhůtu. Stanovte se spolehlivostí 95% intervalový odhad podílu konzerv s prošlou záruční lhůtu.

EXCEL

0,184 → =CONFIDENCE.NORM(0,05;ODMOCNINA(0,184*0,816);320)

5) Podle nejmenovaného volebního modelu by v současnosti hnutí VŠEM získalo více jak třetí voličů (34,8%), oproti výsledku voleb v roce 2018 je to o 2,5% více. Dotázáno bylo 940 osob.

- Určete bodový odhad volebních preferencí hnutí VŠEM.
- Stanovte intervalový odhad pro volební preference hnutí VŠEM.

BODOVÝ A INTERVALOVÝ ODHAD

DOMÁCÍ ÚKOL:

Obchodní řetězec TESCO si v červnu 2018 zadal studii týkající se počtu zákazníků v prodejně TESCO Poruba každý pátek v měsíci. Předpokládejme, že sledovaný počet zákazníků má normální rozdělení. Po jednom měsíci sledování prodejny jsme získali údaje uvedené v tabulce.

Určete pro management řetězce TESCO intervalový odhad středního počtu zákazníků v prodejně TESCO Poruba v pátek odpoledne (se spolehlivostí 95%).

datum	Počet zákazníků v TESCO
01.06.2018	3756
08.06.2018	2987
15.06.2018	3042
22.06.2018	4206
29.06.2018	3597

Řešení: $\mu \in \langle 2883; 4153 \rangle$

DOMÁCÍ ÚKOL:

Průzkumem se zjistilo, že 90 z 800 smrků je napadených kůrovcem. Sestrojte 95% interval pro podíl napadených smrků v celém lese.

Řešení: $P(0,0906 < \mu < 0,1344)$



Hypotéza vychází z teorie, kterou je třeba doložit.
Hypotéza je pouhým odhadem, založeným na náhodném pozorování.

Srovnáváme dvě tvrzení:

H0 (nulová hypotéza) – většinou obhájí stávající stav věcí. Nulová hypotéza je hypotéza, kterou testujeme, neboli o které chceme rozhodnout, zda je pravdivá.

H1 (alternativní hypotéza) - odporuje nulové hypotéze. Zpravidla něco nového, alternativa stojící proti nulové hypotéze.

Jednoduchý a univerzální recept na řešení hypotéz:

- 1) Ze zadání si vypíšeme hodnoty, které známe a správně je označíme.
- 2) Formulujeme testovanou hypotézu H0 a alternativní hypotézu H1. (první sloupec)
- 3) Podíváme se do vzorců a vybereme vhodné **testové kritérium**. (druhý sloupec)
- 4) Vypočítáme hodnotu testového kritéria.
- 5) Zapišeme **kritický obor** (třetí sloupec), dosadíme do něj hodnotu test. kritéria, zvolenou **hladinu významnosti** a vyhledáme hodnotu v tabulkách.
- 6) Jestliže hodnota testové statistiky:
 - a) **Patří (nerovnost platí)** do kritického oboru, potom na dané hladině významnosti **ZAMÍTÁME H0 a přijímáme H1**.
 - b) **Nepatří (nerovnost neplatí)** do kritického oboru, potom na dané hladině významnosti **NEZAMÍTÁME H0. Výsledek je statisticky nevýznamný**. Pozor: Neznamená to, že zamítáme H1.
- 7) Zapišeme závěrečnou interpretaci.

1. Podle výsledných kritických oborů rozhodněte, zda zamítnete nebo nezamítnete nulovou hypotézu.

$$\text{a) } W_{0,01} = \{u; u \geq 2,326\} \qquad U = 5,42$$

$$\text{b) } W_{0,05} = \{t; t \geq -2,145\} \qquad t = -1,83$$

$$\text{c) } W_{0,05} = \{u; u \geq 1,96\} \qquad U = 1,53$$



TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

2) Z vyrobených koloběžek v určitém období bylo vybráno 30 modelů a byla zaznamenána jejich hmotnost. Šetření bylo prováděno firmou z Evropy a údaje jsou uvedené v gramech. Z těchto údajů byl spočten průměr 2452 gramů a směrodatná odchylka 510 gramů.

Proveďte statistický test na 5% hladině významnosti, že výběr pochází ze základního souboru se střední hodnotou 2400 gramů, a to vůči oboustranné alternativní hypotéze.

3) Na minci nám padlo 22 orlů ze 40 hodů. Preferuje tato mince orly? (řešte pro $\alpha=0.01$)

DOMÁCÍ ÚKOL: Vyzkoušej pro situaci, kdy nám na minci padlo 28 orlů ze 40 hodů. Preferuje tato mince orly? (řešte pro $\alpha=0.01$)

Řešení: test.kritérium = 2,53 - Na 1% hladině významnosti zamítáme H_0 a přijímáme H_1 .



4) Továrna vyrábí koberce o průměrné délce 205 cm. Náhodně jsme vybrali několik koberců o délkách 210, 217, 209, 216, 216, 215, 220, 214, 213 cm. Je známo, že rozptyl výběru i základní populace je 25. Otestujte, zda jsou náhodně vybrané koberce vyrobeny v továrně.

5) Hranice pro vstoupení politické strany do sněmovny je 5%. Byl proveden průzkum, kde z 400 dotázaných by sledovanou stranu volilo 35 lidí. Můžeme na 5% hladině významnosti předpokládat proniknutí strany do sněmovny ve volbách?

