

MA 0008 – Teorie pravděpodobnosti

VYUŽITÍ PROGRAMU EXCELL

Břetislav Fajmon, Petr Nezval

Obsah

1	Zpracování souboru měření – průměr	3
2	Zpracování souboru měření – popisná statistika	6
3	Kombinatorika a binomické psti	11
4	Základní pravděpodobnostní modely	14
5	Test chí kvadrát o tvaru rozdělení	15
6	Další příklady	17

Úvod

Tento text je doplňkovým textem k hlavním dvěma textům Ma008 přednáškový text a Ma008 cvičení s tím, že oba texty doplňuje odkazy na využití prostředí Excel ve statistice i pravděpodobnosti.

Obsah předmětu je rozdělen na tři části:

- Cvičení 01-02: Popisná statistika.
- Přednáška 01-08, cvičení 03-08: Teorie pravděpodobnosti.
- Přednáška a cvičení 09-12: Úsudková statistika.

autoři, září 2021

1 Zpracování souboru měření – průměr

- Statistická jednotka - elementární jednotka podrobená statistickému zpracování (např. zaměstnanec, student, firma, věc). Jedna zpracovávaná jednotka má zpravidla několik statistických znaků
- Statistický znak = (stat. proměnná) - konkrétní vlastnost stat. jednotky (věk zaměstnance, mzda zaměstnance, zaměstnancovo vzdělání,...)

1. kvantitativní znak (quantity = množství)

Lze vyjádřit číselně (počet členů v domácnosti, spotřeba vody,...)

Kvantitativní znaky často dále dělíme na:

- diskrétní znaky ... nabývají oddělených číselných hodnot (N), např. počet členů domácnosti, počet výrobků)
- spojité znaky ... nabývají hodnot z intervalu, např. spotřeba vody, doba čekání na příchod zákazníka

2. kvalitativní = nominální = kategoriální znak

Vyjádřuje se slovně (nomen = jméno, název) nebo subjektivním číslem

Kvalitativní znaky dělíme na:

- alternativní ... nabývají dvou hodnot (ano-ne, muž-žena, pravda-lež)
- množné ... nabývají 2 hodnot (vzdělání: ZŠ, SŠ, VŠ, Ph.D.)
- ordinální ... vyjádřené subjektivní stupnicí (míra spokojenosti s výrobkem vyjádřena na číselné škále 1- hodně, 5- vůbec ne, známka ve škole)

I kvalitativní znaky lze vyjádřit číslem a zpracovat počítačem

- alternativní ... např. ano = 1, ne = 0
- množné ... např. ZŠ = 1, SŠ = 2, VŠ = 3, Ph.D. = 4
- ordinální ... už je vyjádřeno číslem na stupnici (např. 1 až 5)

Ovšem tyto číselné hodnoty jsou subjektivní, např. 2 minus 1 je číselně to samé jako 3 minus 2, přesto nelze říci, že existuje stejný odstup mezi "rozhodně ano" a "spíše ano" jako mezi "nevím" a "spíše ano" (nebo ve škole nelze říci, že mezi známkou 1 a 2 je hodnotově stejný rozdíl jako mezi 4 a 5)

Příklad na ZŠ - práce v hodině

- kvantitativní znaky ... počet sourozenců, výška v cm, velikost nohy
- kvalitativní znaky ... barva očí, chlapec/dívka, oblíbený předmět

Jak získáme data? Měřením nebo dotazníkem.

- Základní soubor ... soubor všech jednotek, na kterých má smysl sledovat určité znaky = proměnné. Zpravidla je velmi obsáhlý, někdy nekonečný, tj. změřit všechny jednotky je často nákladné nebo neproveditelné.
A proto provádíme tzv.
- Výběrové šetření ... pro získání informací ze základního souboru vybereme jenom několik jednotek - měřením či dotazníkem. Získáme tzv. výběrový soubor

Tento výběrový soubor by měl být získán z tzv. reprezentativního vzorku populace = z takové množiny vybraných jedinců, který je věrnou zmenšeninou populace, má tedy stejné vlastnosti. Výběr nestranný, žádnému jednotlivci nebo skupině, tvoří ho jednotky pro základní soubor typické.
(Získat takový nestranný vzorek může být dost náročné)

Máme 3 druhy průměru, které můžeme počítat

1. Aritmetický průměr hodnot: $\bar{x} = \frac{1}{n} \cdot \sum_{i=0}^n x_i$
2. Geometrický průměr hodnot: $\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$
3. Harmonický průměr hodnot: $\bar{x}_H = \frac{n}{\sum_{i=0}^n \frac{1}{x_i}}$

Příklad 1.1 Určete u hodnot: 1, 5, 25, 125 a 625:

- a) aritmetický průměr
- b) geometrický průměr
- c) harmonický průměr

Řešení:

- a) V excelu pomocí funkce PRŮMĚR. =PRŮMĚR(1;5;25;125;625)

$$=PRŮMĚR(1;5;25;125;625) = 156,2$$

- b) V excelu pomocí funkce GEOMEAN. =GEOMEAN(1;5;25;125;625)

$$=GEOMEAN(1;5;25;125;625) = 25$$

- c) V excelu pomocí funkce HARMEAN. =HARMEAN(1;5;25;125;625)

$$=HARMEAN(1;5;25;125;625) = 4,0013$$

(Místo vypisování hodnot označíme políčka, ve kterých máme hodnoty napsány, abych tady nemusel vypisovat všechna čísla. Tak tomu bude již v dalších příkladech. Čili pokud bude v příkladu například: =PRŮMĚR(D6:D12), znamená to, že hodnoty, ze kterých počítáme průměr jsou v buňkách D6 až D12)

Příklad 1.2 50 studentů psalo test z pravděpodobnosti. Jejich bodové zisky byly následující. 5 studentů získalo 4 body, 10 studentů 6 bodů, 12 studentů 8 bodů, 15 studentů 11 bodů a zbylí studenti 12 bodů. Vypočtete v excelu aritmetický průměr bodů.

Řešení:

Sestavíme tabulku o dvou řádcích a 50 sloupcích. První řádek vyplníme postupně čísla 1-50. To nám udává číslo studenta. Do druhého řádku píšeme počty bodů. Tedy prvních 5 sloupců bude obsahovat číslo 4 dalších 10 sloupců číslo 6 a tak dále až k 50. sloupci. Jakmile máme tabulku stačí nám zadat název funkce a označit příslušné buňky.

Výsledek: =PRŮMĚR(B3:AY3) = 8,74

Příklad 1.3 Hodnoty růstu firmy v procentech za posledních 5 let byly 101,3%, 108,5%, 100,6%, 98,7% a 102,3%. Vypočtete průměrný roční nárůst za dané pětileté období.

Toto je typický příklad, kde použijeme geometrický průměr. Geometrický průměr využijeme právě u příkladů, kde se počítá průměrný roční (denní, měsíční) zisk, výdělek, produkce, ... nějaké veličiny za několik let (dní, měsíců). Všimněte si, že naše údaje v procentech jsou vlastně podíly. A to vždy hodnota růstu firmy za letošní rok lomeno hodnotou růstu, za rok předcházející

Pro výpočet tedy použijeme funkci GEOMEAN ve tvaru =GEOMEAN(C3:C7). Dostaneme výsledek 102,2271%. Výsledek nám tedy říká, že v průměru se každý rok prodalo o 2,2% více než předchozí rok.

Pokud byste chtěli příklad počítat podle vzorečku pro geo. průměr, tak by to vypadalo následovně: $\bar{x}_G = \sqrt[5]{101,3 \cdot 108,5 \cdot 100,6 \cdot 98,7 \cdot 102,3} = 102,2271 \%$

Příklad 1.4 Slon šel na vycházku po obvodu svého čtvercového výběhu o straně 2 km, po jedné straně šel rychlostí 1 km/h, druhou rychlostí 2 km/h, třetí rychlostí 4 km/h, a poslední opět 1 km/h. Jaká byla jeho průměrná rychlost během jeho vycházky?

Pro výpočet tohoto příkladu využijeme harmonický průměr. Harmonický průměr se typicky využívá při počítání hodnot, které jsou převrácenými hodnotami jiné veličiny.

Řešení: Víme tedy, že budeme počítat průměr harmonický. Zadáme tedy do excelu funkci =HARMEAN(1,2,4,1) a dostaneme že průměrná rychlost slona byla 1,4545 km/h.

Zde jsem nepoužil souřadnice buněk, ale čísla do funkce psal rovnou, jelikož u takto malého počtu hodnot, je to rychlejší.

Pokud byste chtěli počítat pomocí vzorečku, ten by vypadal následovně: $\bar{x}_H = \frac{4}{1 + \frac{1}{2} + \frac{1}{4} + 1} = 1,4545$

2 Zpracování souboru měření – popisná statistika

Příklad 2.1 V prodejně obleků prodali během týdne 46 obleků. Velikosti prodaných obleků byly následující: 39 41 40 42 41 40 42 42 40 43 42 41 43 39 42 41 42 39 41 37 43 41 38 43 42 41 40 41 38 40 40 39 41 40 42 40 41 42 40 43 38 39 41 41 42 a 45.

- sestavte histogram četností a polygon četností z těchto dat
- sestavte tabulku relativních četností, kumulativních absolutních četností, kumulativních relativních četností pro tato data
- určete modus a medián, průměr, rozptyl a směrodatnou odchylku velikostí obleků
- určete variační rozpětí a mezikvartilové rozpětí velikosti obleků
- určete 0,45 kvantil, 0,57 kvantil, 0,869 kvantil... (pomocí kumulativních relativních četností)

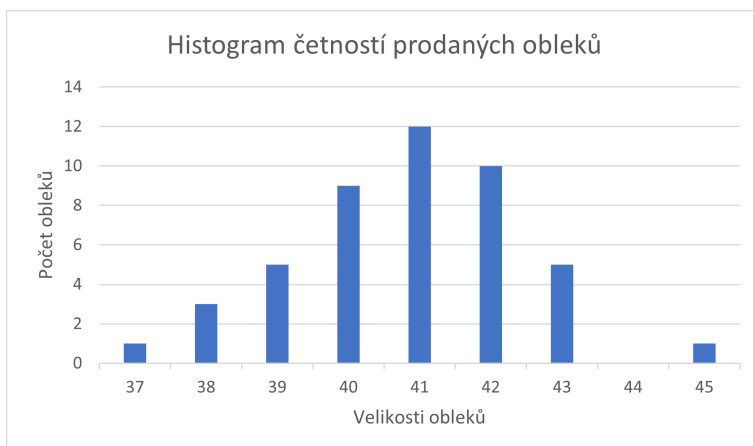
Řešení:

- V excelu sestavíme následující tabulku:

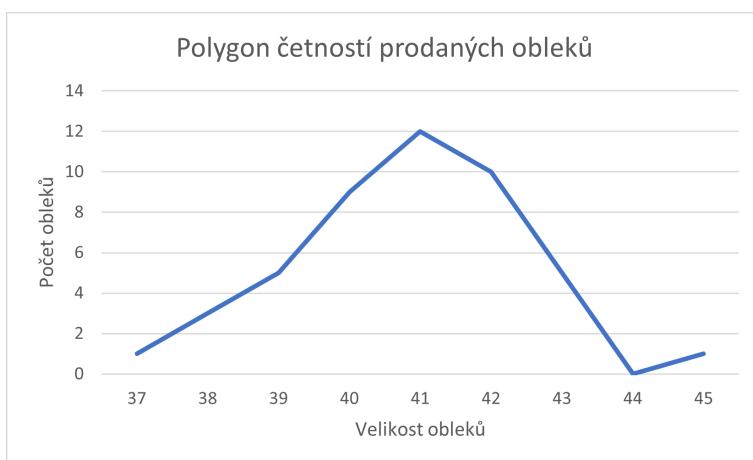
Tabulka 2.1: Tabulka velikostí prodaných obleků

Velikost obleku	Počet prodaných kusů
37	1
38	3
39	5
40	9
41	12
42	10
43	5
44	0
45	1

Hodnoty této tabulky poté označíme a klikneme na "vlození" dále "doporučené grafy" a vybereme "skupinový sloupcový". Tím nám excel vykreslí histogram četností velikostí obleků. Ještě najedeme na histogramu na "prvky grafu," zde necháme zaškrtně možnosti "osy, názvy os, název grafu a mřížka". Poté už jen napíšeme název grafu a názvy os. Výsledek by měl vypadat takto: Při vytváření polygonu pracujem téměř stejně jako u vytváření histogramu, akorát místo "skupinový sloupcový" vybereme "spojnicový". Excel nám vykreslí polygon. Opět v sekci "prvky grafu" doplníme polygon o názvy os a název grafu. Výsledek vypadá zhruba takto:



Obrázek 2.1: Histogram četností prodaných obleků.



Obrázek 2.2: Polygon četností prodaných obleků.

b) Zde naši tabulku doplníme od další tři sloupce, v jednom bude relativní četnost ta se počítá počet všech obleků celkem (tedy 46) děleno počet obleků dané velikost (1 u velikosti 37, 3 u velikosti 38 atd.) krát 100 a máme relativní četnost v %. Ve druhém sloupci bude kumulativní absolutní četnost - zde pouze sčítáme četnosti jednotlivých velikostí (např u velikosti 40 bude součet četností velikostí 37, 38, 39 a 40, tedy $1+3+5+9$ a to je 18). U komutativní relativní četnosti pracujeme velmi podobně jako u absolutní jen sčítáme již vypočítanou relativní četnost. Výsledkem je takováto tabulka:

Tabulka 2.2: Tabulka velikostí prodaných obleků

Velikost obleku	Počet prodaných kusů	Rel. četnost	Kum. četnost	Kum. rel. četnost
37	1	2,17%	1	2,17%
38	3	6,52%	4	8,69%
39	5	10,87%	9	19,57%
40	9	19,57%	18	39,13%
41	12	26,09%	30	65,22%
42	10	21,74%	40	86,96%
43	5	10,87%	45	97,83%
44	0	0%	45	97,83%
45	1	2,17%	46	100,00%

- c) *Modus - velikost, která má největší četnost - mod = 41. V excelu pomocí =MODE(B6:B51) = 41 (V buňkách B6 až B51 jsou velikosti obleků)*
Medián - hodnota, která dělí řadu na dvě stejné poloviny v našem případě, jelikož máme sudý počet prvků je medián arit průměr 23. a 24. prvku, jelikož 23. prvek i 24. prvek je velikost 41, tak i med = 41. V excelu pomocí =MEDIAN(B6:B51) = 41
Arit. průměr (zde bez komentáře) $\bar{x} = 40,83$
Rozptyl uděláme rovnou v excelu, protože to je nejjednodušší. =VAR.P(B6:B51) = 2,535
Stejně tak je jednodušší dělat směrodatnou odchylku rovnou v excelu, nebo prostě můžeme jen rozptyl odmocnit. =SMODCH(B6:B59) = 1,592
- d) *Variační rozpětí uděláme opět v excelu. Zde použijeme funkce MAX a MIN a odečteme je od sebe. Je to tedy rozdíl mezi největším a nejmenším prvkem. Lze ho tedy určit v tomto případě snadno. Pomocí excelu je to tedy =MAX(B6:B51)-MIN(B6:B51) = 8*
Mezikvartilové rozpětí představuje rozdíl mezi třetím a prvním kvantilem, tak ho tedy i spočítáme pomocí excelu. =QUARTIL(B6:B51;3)-QUARTIL(B6:B51;1) = 2.
Pokud necháme předešlé dva vzorečky osamotě a nebudeme je odečítat, máme rovnou i první a třetí kvartil.
- e) *Tyto kvantily vypočítáme pomocí funkce PERCENTIL.*
Takže =PERCENTIL(B6:B51;45%) = 41, =PERCENTIL(B6:B51;57%) = 41,

$$=PERCENTIL(B6:B51;86,9\%) = 42,105.$$

Příklad 2.2 V největších 27 městech České republiky jsou následující ceny bytů v korunách za m². 12 736, 12 975, 13 829, 14 316, 14 546, 14 897, 16 343, 16 369, 17 217, 17 327, 17 332, 18 200, 19 221, 20 162, 20 319, 20 864, 21 456, 21 794, 22 083, 22 215, 22 425, 22 768, 24 567, 25 078, 25 436, 29 031, 45 061. Proveďte intervalové rozdělení četností (relativní, kum. absolutní, kum. relativní), průměr, rozptyl, směrodatnou odchylku, variační rozpětí, 0,25-quantil a 0,85-quantil.

Řešení

Nejprve si sestrojíme intervaly, do kterých pak naše hodnoty zařadíme. Jak na to? Použijeme Sturgesovo pravidlo, to nám řekne, kolik budeme mít intervalů. Pravidlo vypadá takto: $1 + 3,3 \cdot \log n = V$ našem případě tedy $1 + 3,3 \cdot \log 27 = 5,72$ Zvolíme tedy 6 intervalů. Teď ještě šířku intervalu. Tu zjistíme podle vzorečku: největší hodnota minus nejmenší hodnota lomeno počet intervalů. V našem případě tedy. $\frac{45061-12736}{6} = 5387,5$. Zaokrouhlíme na 5 400 (je třeba vždy zaokrouhlit na číslo větší, jinak se může stát, že některá hodnota nespadne do žádného intervalu).

Nyní sestavíme tabulku s rozdělením četností.

Tabulka 2.3: Tabulka četností cen bytů

Cena bytu	Počet měst	Rel. četnost	Kum. četnost	Kum. rel. četnost
12 700 - 18 100	11	40,74%	11	40,74%
18 100 - 23 500	11	40,74%	22	81,48%
23 500 - 28 900	3	11,11%	25	92,59%
28 900 - 34 300	1	3,7%	26	96,29%
34 300 - 39 700	0	0%	26	96,29%
39 700 - 45 100	1	3,7%	27	100%

Následující výpočty budou dělány v excelu.

Výpočet průměru (bez komentáře): $\bar{x} = 20318,41$

Výpočet rozptylu: $=VAR.P(A2:A28) = 40\,475\,082,02$

Výpočet směr. odchylky: $=SMODCH(A2:A28) = 6\,362$

Výpočet variačního rozpětí: $=MAX(A2:A28)-MIN(A2:A28) = 32\,325$

Výpočet kvantilů je složitější. Nejprve musíme vynásobit hledaný kvantil číslem $n+1$. V našem případě $0,25 \cdot 28$. Dostaneme číslo 7. Hledáme tedy hodnotu 7. prvku intervalu. Tu zjistíme pomocí vzorce:

$$x_\alpha = a + \frac{(n+1) \cdot \alpha - c_a}{n(a;b)} \cdot (b-a)$$

Kde a - dolní mez intervalu, b - horní mez intervalu a c_a - kumulace v bodě dolní meze intervalu

V našem případě tedy:

$$x(0,25) = 12700 + \frac{28 \cdot 0,25 - 0}{11} \cdot 5400 = 16136$$

Nyní výpočet 0,85-kvantilu. Opět uděláme $0,85 \cdot 28$. Zde dostaneme hodnotu 23,8. To, že nám nevyšlo číslo celé nám nevadí a počítáme dále. Nalezneme tedy interval, kde je 23,8. hodnota a dosadíme do vzorečku.

$$x(0,85) = 23500 + \frac{28 \cdot 0,85 - 22}{3} \cdot 5400 = 26740$$

Úkolek 05: Pro příklad cen za metr čtvereční bytů proveďte intervalové rozdělení četností, průměr, rozptyl, odchylku, variační rozpětí, kvantily-percentily, zhruba podle toho, jak bylo počítáno na cvičení.

3 Kombinatorika a binomické psti

V této kapitole si nejdříve v rychlosti ukážeme, jak vypočítat faktoriál a kombinační číslo v excelu.

Příklad 3.1 Vypočítejte a) $10!$, b) kombinační číslo $\binom{10}{3}$

a) faktoriál vypočítáme jednoduše, stačí jen zadat `=FAKTORIÁL(10)`, to nám dá výsledek 3628800

b) kombinační číslo získáme podobně jednoduše, v případě kombinačního čísla $\binom{10}{3}$ zadáme funkci `KOMBINACE` a to následujícím způsobem: `=KOMBINACE(10;3)`, to nám dá výsledek 120

Nyní se již podíváme na binomické rozdělení psti v excelu.

Příklad 3.2 Vypočítejte binomické rozdělení psti a nakreslete jejich histogramy psti pro

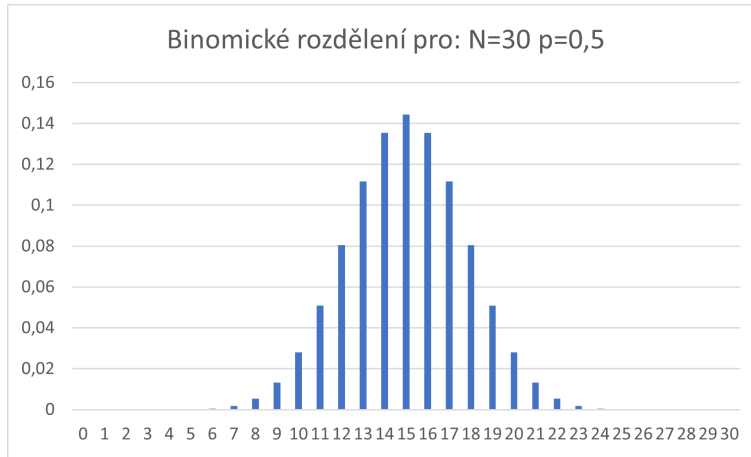
a) $N = 30$, $p = \frac{1}{2}$, b) $N = 30$, $p = \frac{1}{3}$, c) $N = 30$, $p = \frac{5}{6}$

Řešení:

a) $N = 30$, $p = \frac{1}{2}$,

Použijeme funkci `BINOM.DIST` a to následujícím způsobem: Nejdříve si do sloupečku napíšeme hodnoty 0-30, ty představují počet úspěchů. (Dále v textu je příklad řešen pro případ, kdy v buňce A13 je 0 A14 je 1 atd. až po A43, kde je 30). Následně sestrojíme funkci. `=BINOM.DIST(A13;30;0,5;NEPRAVDA)`. Tím získáme první hodnotu pro 0 úspěchů. Tuto funkci si v excelu "natáhneme" dolů až budeme mít všechny výsledky. Funkce tedy zůstává stejná, jen se nám mění A13 postupně na A14 A15,... až po A43. [`=BINOM.DIST(A13;30;0,5;NEPRAVDA)`, `=BINOM.DIST(A14;30;0,5;NEPRAVDA)`, `=BINOM.DIST(A15;30;0,5;NEPRAVDA)`,...`=BINOM.DIST(A43;30;0,5;NEPRAVDA)`]
Výsledky v tabulce úplně na konci textu až za zdroji

Zde je histogram tohoto rozdělení:



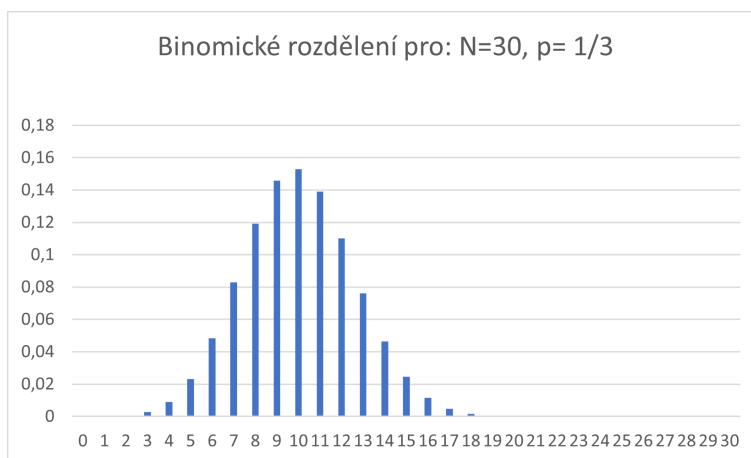
Obrázek 3.3: Binomické rozdělení pro $N = 30$, $p = \frac{1}{2}$.

b) $N = 30$, $p = \frac{1}{3}$

Zde použijeme úplně stejný postup jako v prvním případě. Pouze upravíme funkci do podoby $=\text{BINOM.DIST}(A13;30;1/3;NEPRAVDA)$ upravili jsme tedy pouze $\frac{1}{2}$ na $\frac{1}{3}$.

Celý postup opakujeme.

Výsledný histogram:

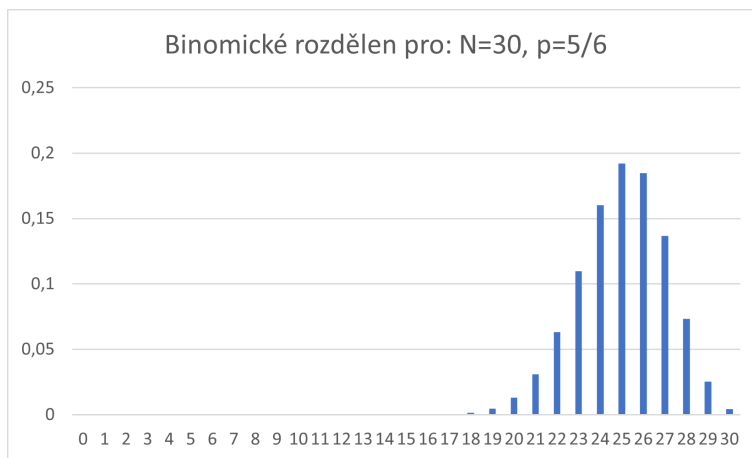


Obrázek 3.4: Binomické rozdělení pro $N = 30$, $p = \frac{1}{3}$.

c) $N = 30, p = \frac{5}{6}$

Postup stejný. Pouze místo `=BINOM.DIST(A13;30;1/3;NEPRAVDA)` použijeme `=BINOM.DIST(A13;30;5/6;NEPRAVDA)` a opět opakujeme.

Výsledný histogram:



Obrázek 3.5: Binomické rozdělení pro $N = 30, p = \frac{5}{6}$.

Příklad na oříšky, z přednáškového textu, str. 32, příklad 9.5, pouze výpočet pomocí binomického rozdělení, ale přesný, co dá Excel – hodila by se i tabulka příslušných dílčích pstí v Excelu, které se do toho výsledku sečtou.

4 Základní pravděpodobnostní modely

V podstatě obsah přednášky 08 (pptx) zpracovaný v Excelu, vyjma binomického modelu D3 zpracovaného už v předchozí kapitole, a vyjma D2 alternativního, který je příliš jednoduchý. Rozepisují do konkrétních úkolů pro Excel:

Příklad 4.1 *Vegenerujte pomocí Excelu 40 hodnot rozdělení $D1 =$ rovnoměrného $Ro(60, 61, \dots, 79, 80)$.*

Příklad 4.2 a) *Vegenerujte pomocí Excelu 40 hodnot počtu hodů správně vyváženou hrací kostkou, které jsou potřeba na padnutí první šestky.*

b) *Nakreslete histogram prvních deseti pravděpodobností veličiny, která měří počet hodů hrací kostkou potřebný na padnutí první šestky.*

Příklad 4.3 a) *Vegenerujte pomocí Excelu 40 hodnot Poissonova rozdělení pro $\lambda = 10$.*

b) *Nakreslete pomocí Excelu histogram prvních dvaceti pravděpodobností Poissonovsky rozdělené veličiny pro 1) $\lambda = 0,5$, 2) $\lambda = 3$, 3) $\lambda = 10$.*

Příklad 4.4 a) *Nakreslete pomocí Excelu graf funkce hustoty psti a graf distribuční funkce exponenciálního rozdělení $Exp(\lambda = 0,5)$, $Exp(\lambda = 3)$ a $Exp(\lambda = 10)$ (jedná se tedy o tři příklady v jednom, v každém příkladu dva grafy, tedy dohromady šest grafů – nakreslete ovšem každý graf do jiného obrázku, bude to tedy celkem šest obrázků).*

b) *Vygenerujte pomocí Excelu čtyřicet hodnot exponenciálně rozdělených veličin z části (a) pro 1) $\lambda = 0,5$, 2) $\lambda = 3$, 3) $\lambda = 10$.*

Příklad 4.5 *Vegenerujte pomocí Excelu 40 hodnot rozdělení $S1 =$ rovnoměrného spojitého na intervalu $\langle 70; 100 \rangle$.*

Příklad 4.6 a) *Nakreslete pomocí Excelu do jednoho obrázku (různými barvami) graf funkce hustoty psti normálního rozdělení $No(\mu = 75, \sigma = 5)$, a normálního rozdělení $No(\mu = 75, \sigma = 10)$.*

b) *Nakreslete pomocí Excelu do jednoho obrázku (různými barvami) graf distribuční funkce normálního rozdělení $No(\mu = 75, \sigma = 5)$, a normálního rozdělení $No(\mu = 75, \sigma = 10)$.*

c) *Vygenerujte pomocí Excelu čtyřicet hodnot normálně rozdělených veličin z části (a) pro 1) $No(\mu = 75, \sigma = 5)$, 2) pro $No(\mu = 75, \sigma = 10)$.*

5 Test chí kvadrát o tvaru rozdělení

Následující sérii příkladů proveďte za pomoci Excelu podle následujícího postupu: a) Odhadněte, jakým modelem popíšeme naměřená data, podle četností nebo intervalového rozdělení četností (nakreslete v Excelu histogram četností) – použijte v každém příkladu, co je vhodnější; b) Pak z měření vypreparejte parametry potřebné pro teoretický popis; c) pomocí teoretického modelu vypočtete teoretické četnosti (uved'te v tabulce v Excelu), d) a nakonec testem χ^2 (chí kvadrát) porovnejte, zda je navržený teoretický model ve shodě s naměřenými daty (test chí kvadrát proveďte také v Excelu, který tento test obsahuje).

Příklad 5.1 51.99882, 54.62071, 58.34618, 52.33503, 54.34023, 50.44404, 48.23762, 48.86250, 48.15060, 49.31562, 55.05544, 46.74591, 49.41650, 44.01911, 47.74186, 46.58175, 55.90116, 60.99310, 45.26392, 52.18025, 47.90630, 47.18151, 54.81286, 55.36082, 50.31988, 55.96361, 57.07858, 50.38156, 41.96963, 54.28295, 51.96203, 42.72137, 48.71775, 58.25657, 44.41251, 43.88321, 56.95593, 52.75391, 45.94371, 38.85037.

Nápověda: pravděpodobně se bude jednat o normální rozdělení $S3$ – zjistíte po provedení intervalového rozdělení četností.

Příklad 5.2 0.31679, 0.19195, 0.03202, 0.10703, 0.33120, 0.00305, 0.15606, 0.02107, 0.07481, 0.05017, 0.1332, 0.19609, 0.08193, 0.08473, 0.03614, 0.15838, 0.13192, 0.06462, 0.06559, 0.24249, 0.04222, 0.03124, 0.35355, 0.00125, 0.06360, 0.11149, 0.03475, 0.17893, 0.06980, 0.16811, 0.01520, 0.09588, 0.00583, 0.03656, 0.03999, 0.06500, 0.12474, 0.06415, 0.07629, 0.21434.

Nápověda: zkuste exponenciální rozdělení $Exp(\lambda)$.

Příklad 5.3 8.398068, 47.183415, 11.655249, 19.874558, 35.560654, 13.948998, 59.421469, 52.361322, 15.931113, 54.191444, 2.283628, 22.631475, 44.125797, 53.957976, 57.499843, 29.170605, 11.832367, 34.954316, 25.547414, 4.719695, 16.023275, 14.554789, 46.299197, 36.764952, 57.274647, 24.112847, 5.281371, 9.604469, 26.929707, 51.412021, 32.699513, 33.283787, 37.283912, 30.088545, 35.840679, 48.711673, 57.638311, 59.552026, 10.683609, 39.997975.

Nápověda: zkuste rovnoměrné spojité rozdělení, model $S2$. Je to dobře vidět z intervalového rozdělení četností.

Příklad 5.4 8, 8, 5, 12, 11, 15, 12, 9, 6, 7, 9, 7, 12, 7, 11, 12, 8, 6, 12, 7, 5, 7, 10, 10, 9, 13, 9, 11, 9, 6, 13, 11, 6, 10, 12, 6, 5, 8, 13, 10, 10, 13, 13, 7, 12, 10, 11, 7, 11, 8.

Nápověda: zkuste rovnoměrné diskrétní rozdělení, model $D1$. Bude to patrné z intervalového rozdělení četností.

Příklad 5.5 5, 6, 6, 11, 5, 5, 5, 2, 11, 5, 6, 3, 8, 4, 9, 5, 10, 7, 11, 8, 6, 8, 4, 3, 2, 7, 2, 4, 4, 4, 7, 7, 8, 7, 4, 2, 4, 4, 7, 6.

Nápověda: zkuste binomické rozdělení, model $D3$. Zvolte $N = 11$ (největší hodnotu), a pak z odhadu $\bar{X} \doteq N \cdot p$ odhadněte p ... bude asi blízke jedné polovině.

Příklad 5.6 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 2, 0, 0, 1, 0, 0, 3, 1, 0.

Nápověda: zkuste Poissonovo rozdělení, model D5.

Příklad 5.7 6, 0, 7, 2, 5, 11, 6, 0, 2, 2, 11, 0, 19, 21, 8, 0, 2, 14, 8, 5, 0, 21, 2, 6, 18, 6, 0, 6, 12, 6, 0, 0, 11, 3, 4, 0, 0, 6, 12, 2, 12, 0, 7, 4, 3, 6, 7, 0.

Nápověda: zkuste geometrické rozdělení, model D4. Zkuste nějak vypreparovat hodnotu p z odhadu střední hodnoty: $\bar{X} \doteq \frac{p}{1-p}$. Asi do čitatele odečtete a přičtete jedničku, rozdělte pak na dva zlomky, ten první zjednodušíte, a z rovnice vyjádříte p .

6 Další příklady

Vypočtete následující příklady = odpovězte na otázky:

- 3) Pro $\alpha = 0,05$ testujte hypotézu, že střední hodnota průměru měření veličiny je ve skutečnosti rovna 11:
9, 7, 12, 7, 11, 12, 8, 6, 12, 7, 5, 7, 10, 10, 9, 13, 9, 11, 9, 6, 13, 11, 6, 10, 12, 6, 5, 8, 13, 10, 10, 13, 13, 7, 12, 10, 11, 7, 11, 8.
- 4) Zjistěte p -hodnotu předchozího testu z otázky 3. Pomoc, kterou asi budete potřebovat: Pokud jste v předchozí otázce použili t -test (což jste udělali dobře), je možné, že budete nyní potřebovat spočítat obsah podgrafu hustoty t -rozdělení v jiných mezích, protože v tabulce t -testu ve skriptech se vyskytuje jen několik kritických hodnot – zkuste použít Excel k nalezení potřebného obsahu podgrafu.
- 5) Sestrojte 95%-ní interval spolehlivosti pro střední hodnotu průměru 40 hodnot veličiny z otázky 3. V jakém vztahu je tento interval s výsledkem testu v otázce 3?
- 6) Najděte kvantily následujících diskrétních rozdělání psti, ovšem nestačí výsledek – měli byste využít vektor dostatečného počtu¹ hodnot pstí $p(k)$, a také příslušných kumulativních pstí $c_p(k)$, které uveďte do tabulky jako součást vašeho řešení.
 - a) Najděte 0,38-quantil rozdělení $Bi(N = 40, p = 0,3)$.
 - b) Najděte 0,78-quantil rozdělení $Po(\lambda = 2)$, kde λ je průměrný počet výskytů jisté náhodné události za hodinu (o jakou náhodnou událost by se mohlo například jednat?).
- 7) Najděte kvantily následujících spojitých rozdělání psti pomocí distribuční funkce $F(x)$, ovšem nestačí výsledek, musíte provést a popsat celý postup – dále nemůžete využít žádný software, pouze kalkulačku, pokud se týká exponenciálního rozdělení, nebo tabulku distribuční funkce rozdělení U , pokud se týká normálního rozdělení.
 - a) Najděte 0,78-quantil rozdělení $Exp(\lambda = 2)$.
 - b) Najděte 0,38-quantil rozdělení $No(\mu = 50, \sigma = 10)$.
- 8)
 - a) Pomocí binomického rozdělení vypočtete: Experti odhadují, že o šampon V-protilup bude mít zájem jen 30% jeho bývalých uživatelů, protože po vánocích podražil o padesát procent. Vypočtete pst (za předpokladu správnosti odhadu expertů), že ze 40 bývalých uživatelů si jej v drogerii nadále koupí aspoň 15 lidí.
 - b) Pomocí náhrady binomického rozdělení normálním s korekcí vypočtete přibližně pst z části (a).

¹Pokud k nabývá nekonečně mnoha hodnot, nemusíte je všechny vypisovat, stačí ta správná část :-).

-
- 9) **a)** Veličina X má rozdělení Poissonovo s průměrným počtem výskytů náhodné události $\lambda = 2$ za hodinu. Jaká je pst, že za konkrétní hodinu měření veličiny dojde ke dvěma a více výskytům této události?
- b)** Pomocí náhrady Poissonova rozdělení normálním s korekcí vypočtete přibližně pst z části (a).
- 10) **a)** Náhodná veličina Y má rozdělení $Exp(\lambda = 2)$, kde λ je průměrný počet výskytů za hodinu. Vypočtete psti $P(Y < 30min)$, $P(Y > 40min)$... uveďte celý výpočet, nejen výsledek.
- b)** Náhodná veličina X má rozdělení $No(\mu = 50, \sigma = 8)$. Vypočtete psti $P(X < 55)$, $P(X > 40)$... uveďte celý výpočet, včetně využití tabulek distribuční funkce rozdělení U .

Seznam literatury:

(Brož, 2002) Mistrovství v Microsoft Excel 2000 a 2002.

(Fajmon, 2021) MA 008 – přednáškový text (Teorie pravděpodobnosti).

(Fajmon, Dufková, 2021) MA 008 – cvičení (Teorie pravděpodobnosti).

Tabulka 3.4: Tabulka binomického rozdělení pro $N = 30$, $p = \frac{1}{2}$

Počet úspěchů	Pravděpodobnost
0	9,31323E-10
1	2,79397E-08
2	4,05125E-07
3	3,78117E-06
4	2,55229E-05
5	0,000132719
6	0,000552996
7	0,001895986
8	0,005450961
9	0,013324572
10	0,027981601
11	0,050875638
12	0,080553093
13	0,111535052
14	0,13543542
15	0,144464448
16	0,13543542
17	0,111535052
18	0,080553093
19	0,050875638
20	0,027981601
21	0,013324572
22	0,005450961
23	0,001895986
24	0,000552996
25	0,000132719
26	2,55229E-05
27	3,78117E-06
28	4,05125E-07
29	2,79397E-08
30	9,31323E-10