Výpočet pravděpodobností

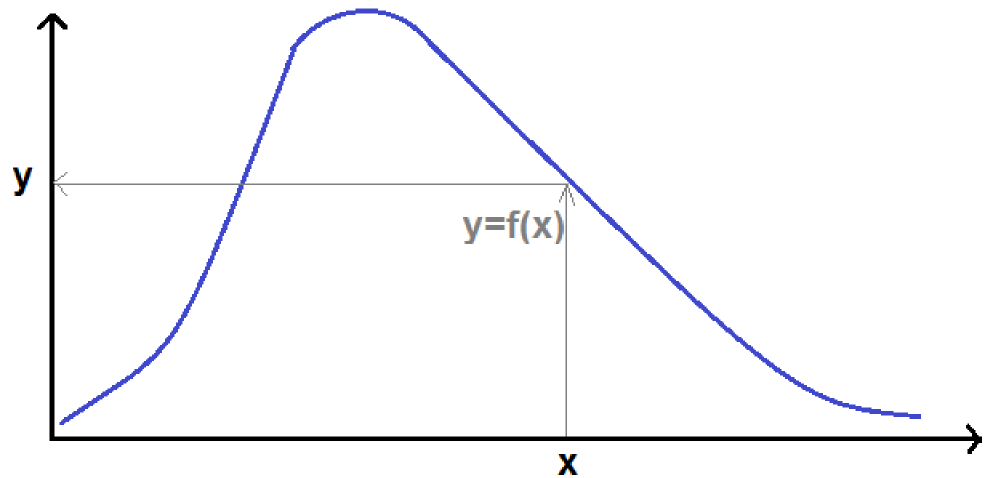
Pravděpodobnostní kalkulátor v programu STATISTICA

Cvičení 5
Statistické metody a zpracování dat 1 (podzim 2016)
Brno, říjen 2016
Ambrožová Klára

Trocha teorie

- Náhodné jevy mají své typické chování, které lze popsat teoretickými rozděleními

- Teoretické rozdělení
  - matematická funkce, která každému potenciálnímu výsledku přiřazuje pravděpodobnost, s nímž ho bude dosaženo
  - Umožňují zjistit, jak se chová základní soubor
  - Můžeme určit, s jakou pravděpodobností dostaneme určitý výsledek při realizaci jevu

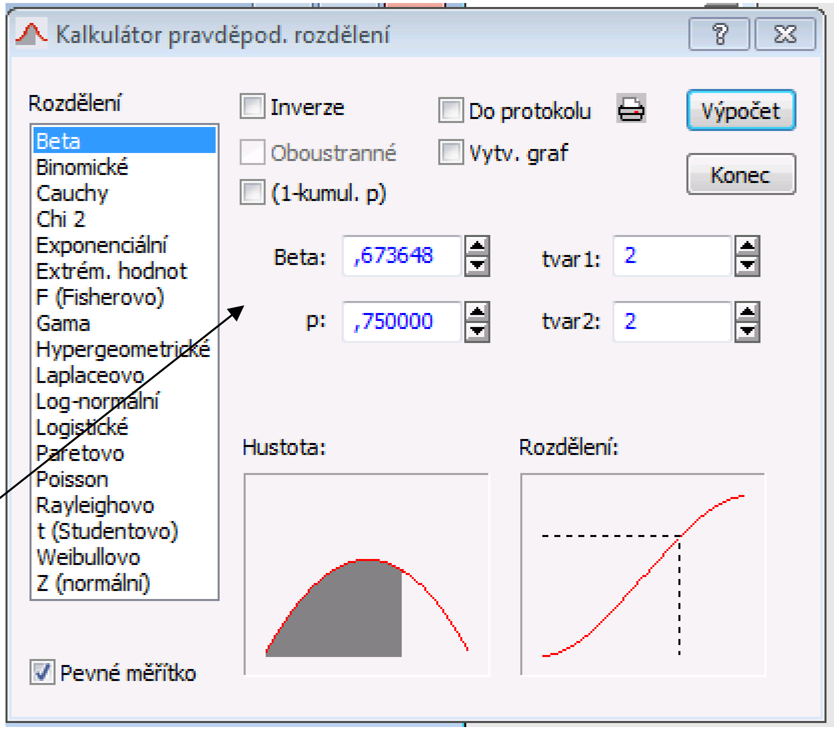


x: potenciální výsledky experimentu

y: pravděpodobnost jejich výskytu

Pravděpodobnostní kalkulátor

- Nástroj v programu Statistica (Statistiky – Základní statistiky – Pravděpobnostní kalkulátor)
- Umožní nám spočíst pravděpodobnost určitého jevu, známe-li:
  - Rozdělení výběrového souboru
  - Parametry výběrového souboru (záleží na rozdělení souboru)
  - Co chceme spočítat ☺



Předpříprava dat

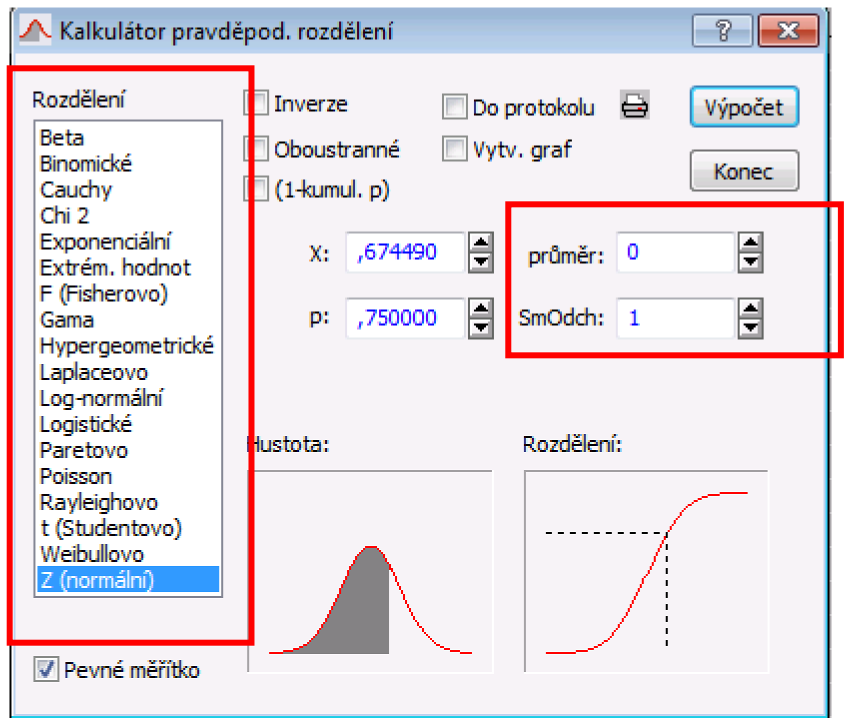
Než začneme počítat v pravděpodobnostním kalkulátoru:

1. **Importujeme do programu Statistica soubor „Klementinum_cv2.xls“**
   (Soubor – Otevřít – *najdeme správný soubor* – Importovat vybraný list do tabulky – *zaklikneme* „1.řádek jako názvy proměnných“)
2. **Vybereme našich 120 let**
   (Data – Podskupina – Případy – *zakliknout* „Povolit podm. výběru“ – Zahrnout případy: Některé, vybrané výrazem (*napíšeme podmínku*))
3. **Určíme rozdělení našeho souboru**
   (provedeno v předchozím cvičení ☺– vyšlo někomu, že nemá normální rozdělení?)
4. **Spočítáme popisné charakteristiky**
   (Statistiky – Základní statistiky – Popisné statistiky – Proměnné *(vybereme proměnnou s teplotními daty*) – Detailní výsledky: *zaklikneme proměnné)*

Pravděpodobnostní kalkulátor

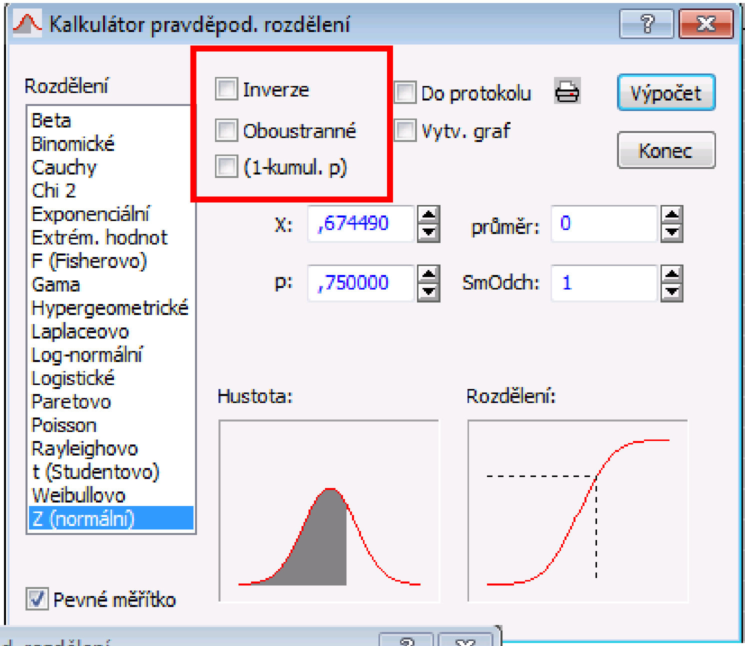
- Spustíme kalkulátor
  (Statistiky – Základní statistiky – Pravděpobnostní kalkulátor)

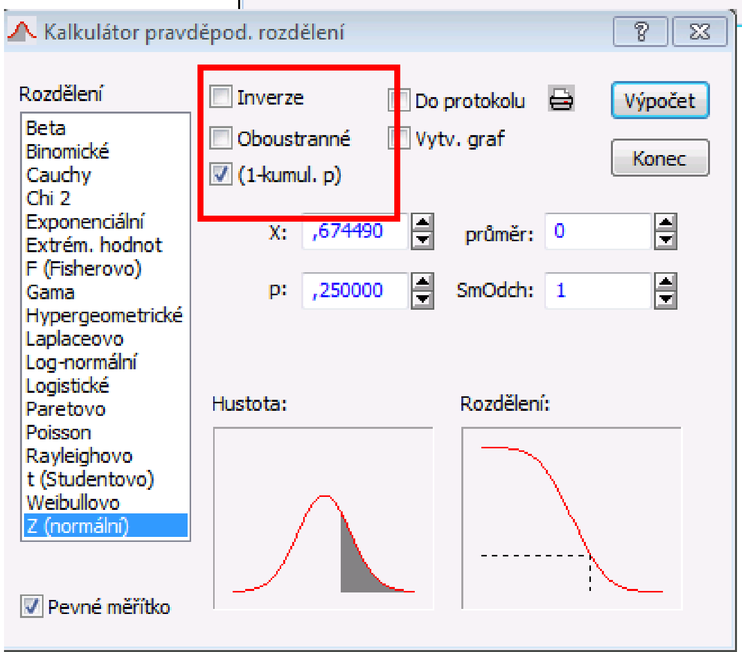
- Co je třeba specifikovat:
  1. Rozdělení výběrového souboru
     - Aktivní je to, které je modře podsvícené
     - Projdeme si ta, která byla probírána na přednášce
  2. Parametry rozdělení
     - Pro Normální rozdělení např. průměr a směrodatná odchylka
     - Nutno znát je dopředu



Pravděpodobnostní kalkulátor

- Co je třeba specifikovat:
  3. Jakou pravděpodobnost počítáme
     1. Pravděpodobnost jevu, že hodnota bude x nebo nižší (ponecháme nezaškrtnutá políčka vlevo nahoře)
     2. Pravděpodobnost jevu, že hodnota bude X nebo vyšší (zaškrtneme (1-kumul.p))

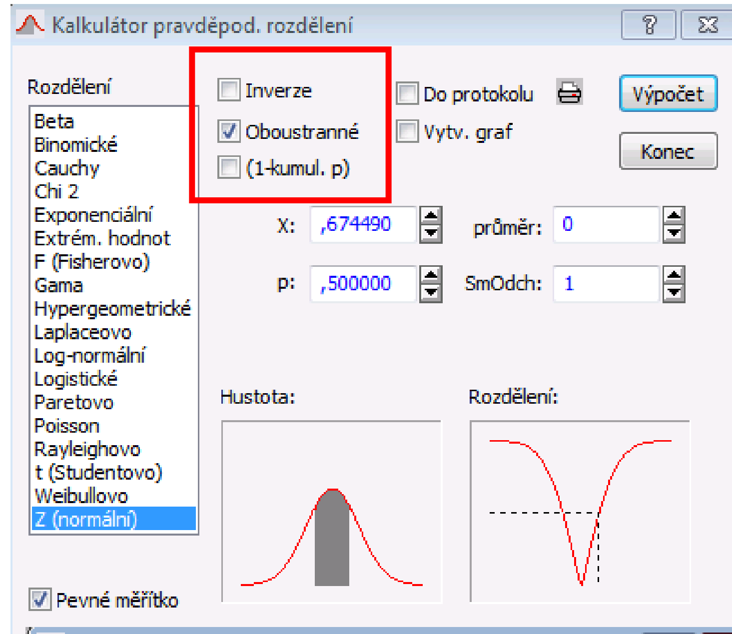


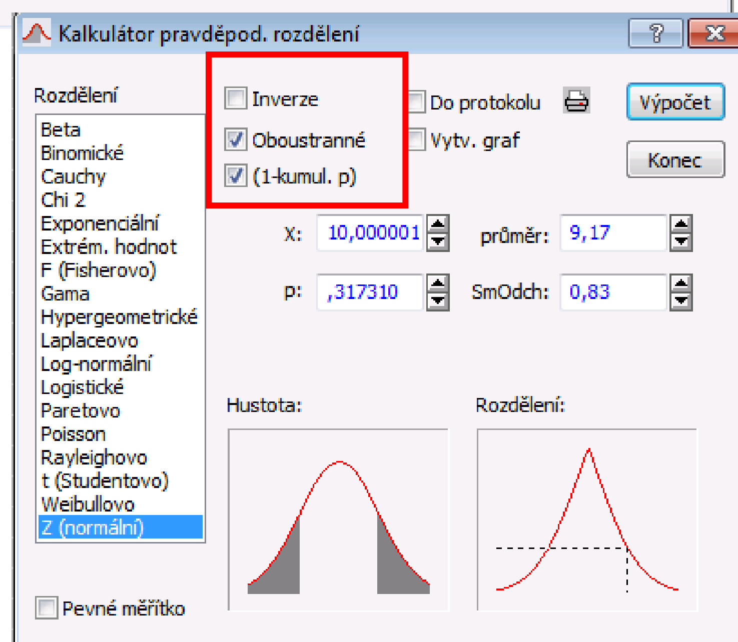


TENTO A NÁSLEDUJÍCÍ SLIDE PLATÍ PRO NORMÁLNÍ ROZDĚLENÍ!

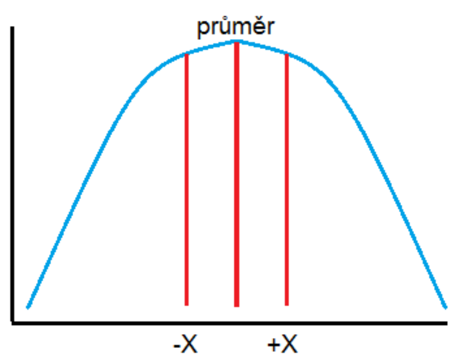
Pravděpodobnostní kalkulátor

- Co je třeba specifikovat:
  3. Jakou pravděpodobnost počítáme
     3. Pravděpdobnost jevu, že hodnota bude mezi –X a +X (zaškrtneme Oboustranné)
     4. Pravděpdobnost jevu, že hodnota bude menší než –X anebo větší než +X (zaškrtneme Oboustranné a (1-kumul.p))



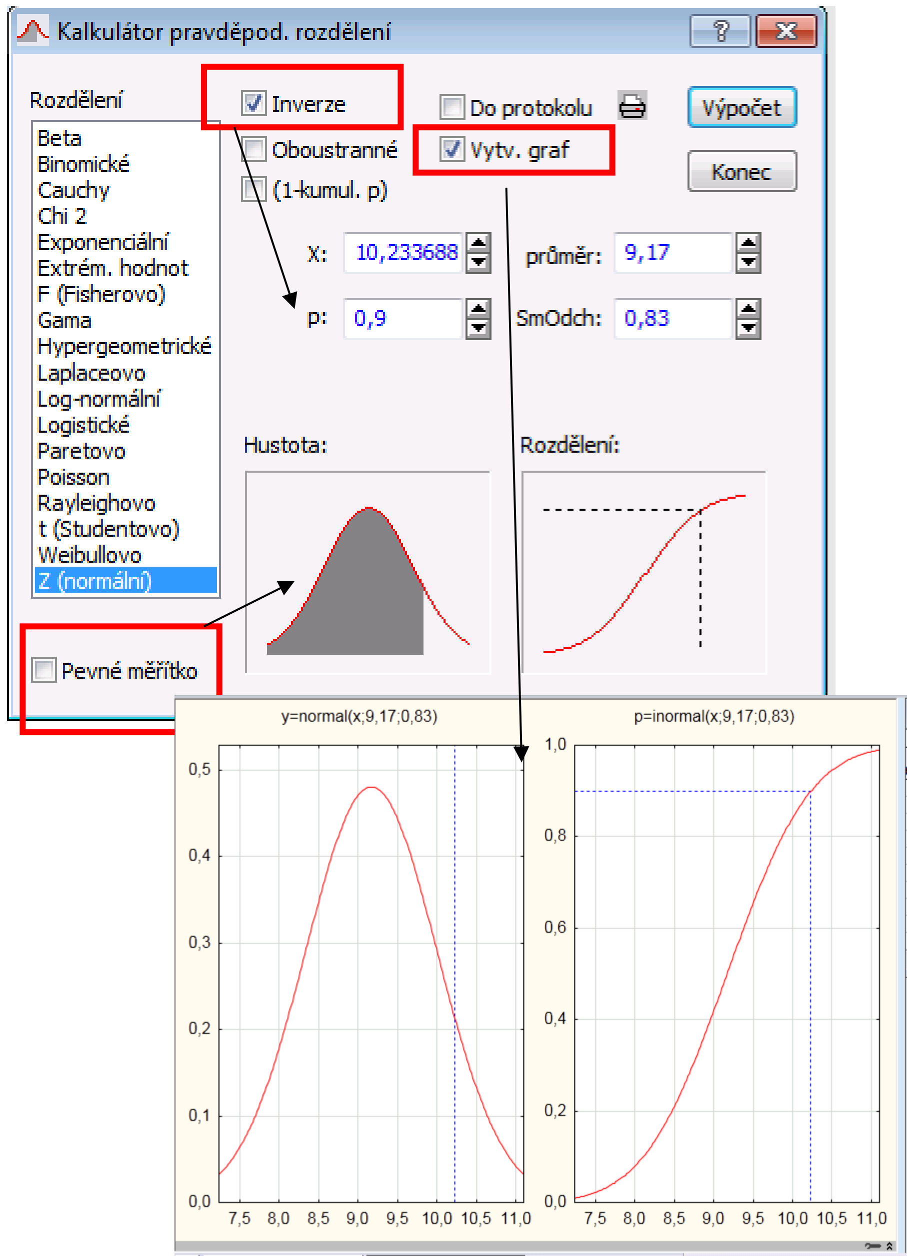


Pozn. Specifikuje-li např. X= 8.33, pak kalkulátor za –X nepovažuje -8.33, ale hodnotu stejně vzdálenou od průměru na opačnou stranu!



Pravděpodobnostní kalkulátor

- Další vychytávky:
  - odkliknout „Pevné měřítko“: grafy dole v okně nejsou fixovány na hodnotu průměru 0 a smodch 1
  - Zakliknout „Vytv.graf“: při zmáčknutí Výpočet se vytvoří větší grafy (které lze vkládat do protokolu…)
  - Zakliknout „Inverze“: chceme-li znát, pro které X bude pravděpodobnost taková, jakou chceme (např. 90%, tedy 0,9) = vyplňujeme v kalkulátoru „p“



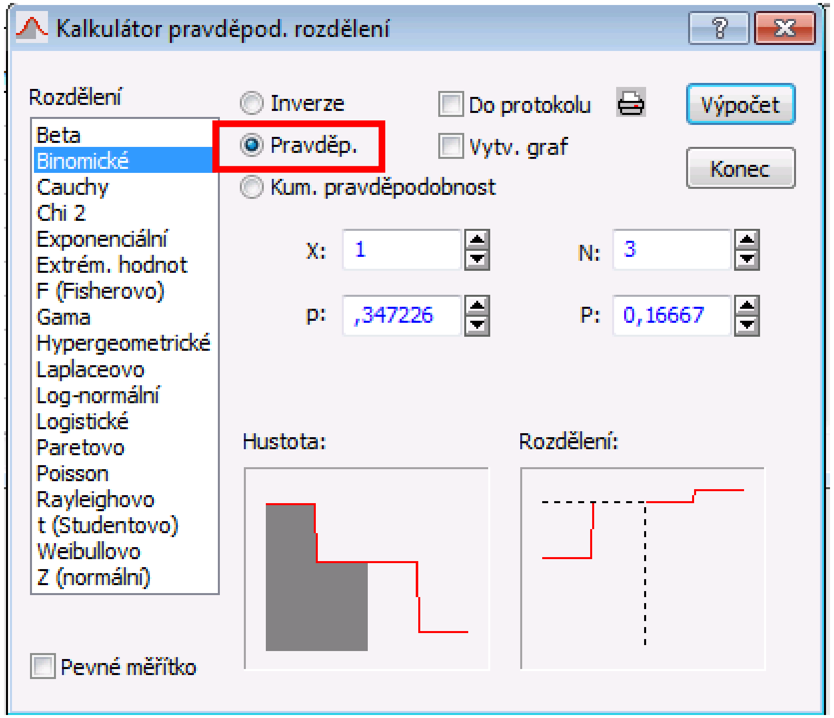
Vybraná rozdělení: použití a potřebné parametry

Binomické rozdělení

- Příklad: při hodu kostkou můžou padnout čísla 1–6 s pravděpodobností p=1/6 (je pro všechna čísla stejná!). Jaká je pravděpodobnost, že při hodu třemi kostkami (n=3) padne šestka?

- Pouze pro diskrétní proměnné!

- parametry:
  - p = pravděpodobnost
  - n = počet pokusů
  - X = 1 „správná“ možnost (pouze šestka je pro nás správná varianta)

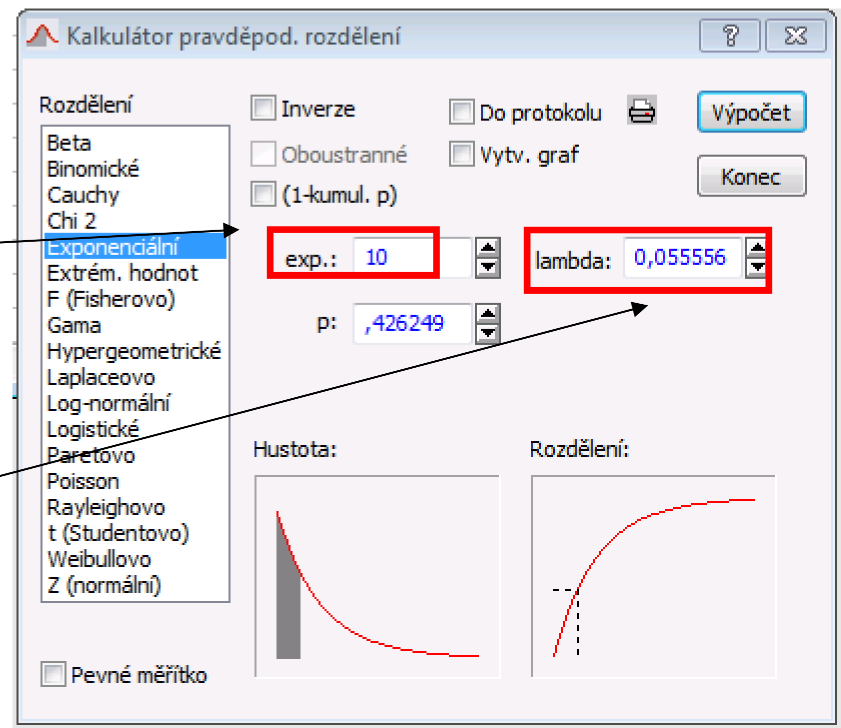


Vybraná rozdělení: použití a potřebné parametry

Exponenciální rozdělení

Lze také spočíst, dokdy vybuchne s p=0,5 (Inverze) nebo s jakou p vybuchne za více než 10 měsíců (1-kumul.p.)

- Pro spojité proměnné
- K výbuchu aktivní sopky může dojít kdykoliv (stejná pravděpodobnost pro všechny časové okamžiky). Jaká je pravděpodobnost, že vybuchne do 10 měsíců, je-li střední doba čekání 18 měsíců?
- parametry: λ, kde 1/ λ je střední doba čekání (na danou událost) – např. 18 = 1/ λ, tzn. λ = 1/18=0,055556

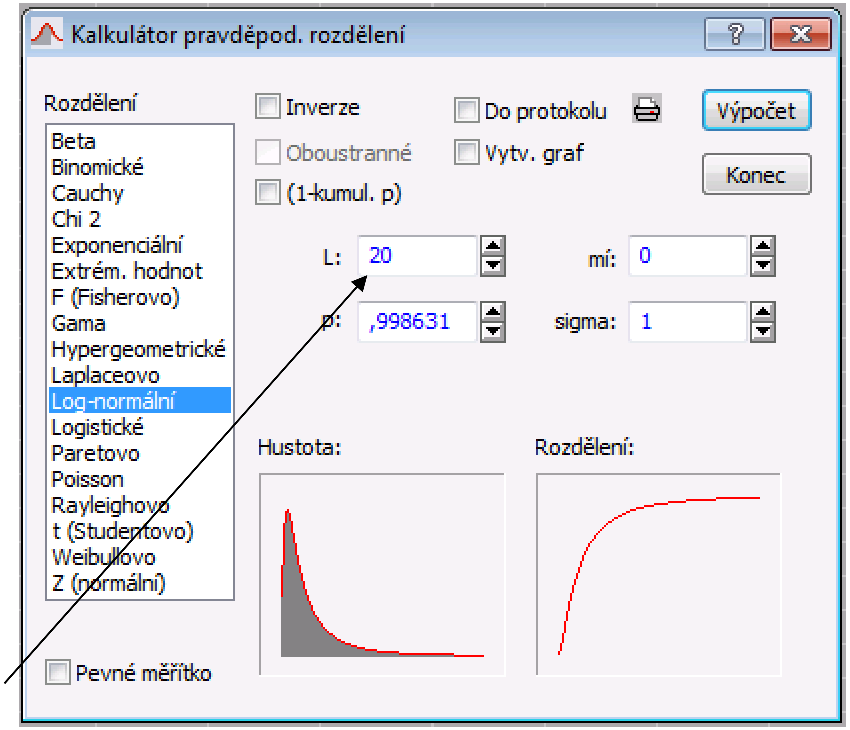


Vybraná rozdělení: použití a potřebné parametry

Lognormální rozdělení

- Pro spojité proměnné
- Provedeme-li logaritmizaci hodnot (y= ln(x)), tak získáme normální rozdělení
- typický jev: rozdělení věku obyvatelstva v populaci
- parametry: μ – střední hodnota zlogaritm. normálního rozdělení
- σ – směrodatná odchylka zlogaritm. normálního rozdělení

- Nikdy jsem to nepoužívala, netuším, jakých může nabývat hodnot (nejspíš by šlo zjistit pravděpodobnost, s níž bude věk prvního jedince, kterého potkáme na ulici, větší než 20 let)

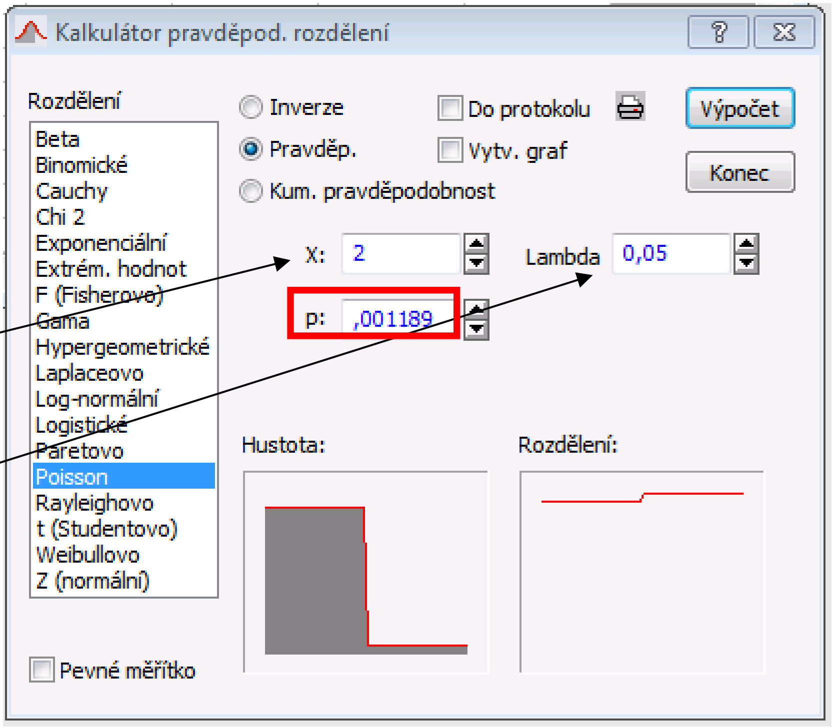


Vybraná rozdělení: použití a potřebné parametry

Poissonovo rozdělení

Lze také spočíst pravděpodobnost méně než 2 zemětřesení (1-kumul.p.) nebo kolik zemětřesení má p=0,01 (Inverze)

- Pouze pro diskrétní proměnné
- Podobné binomickému rozdělení, ale pro málo pravděpodobné jevy nabývající celých kladných čísel během dlouhého časového okamžiku
- Např. pravděpodobnost, že se v ČR daný den vyskytne více než 2 zemětřesení, je-li dlouhodobá pravděpodobnost výskytu zemětřesení = 0.05
- Parametry: λ - pravděpodobnost výskytu daného jevu (v daném dlouhém časovém intervalu)



Cvičení č. 5

- Na základě dat zpracovávaných ve cvičení 4 **předpokládejme**, že průměrné roční teploty vzduchu v Praze, Klementinu **mají normální rozdělení**.

Vypočtěte:

a. jaká je pravděpodobnost, že průměrná teplota vzduchu bude menší nebo rovna aritmetický průměr minus směrodatná odchylka **(x − s)**

b. jaká je pravděpodobnost, že průměrná teplota vzduchu bude menší než aritmetický průměr plus směrodatná odchylka **(x + s)**

c. jaká je pravděpodobnost, že průměrná teplota vzduchu bude větší než aritmetický průměr plus směrodatná odchylka **(x + s)**

d. jaká je pravděpodobnost, že průměrná teplota vzduchu bude nabývat hodnoty v intervalu **(x − s ; x + s)**

e. vypočtěte teploty vzduchu, které se mohou vyskytnout **s pravděpodobností 10, 50 a 90 procent** (pozn.: jde o pravděpodobnost, že se vyskytne daná nebo nižší teplota vzduchu)

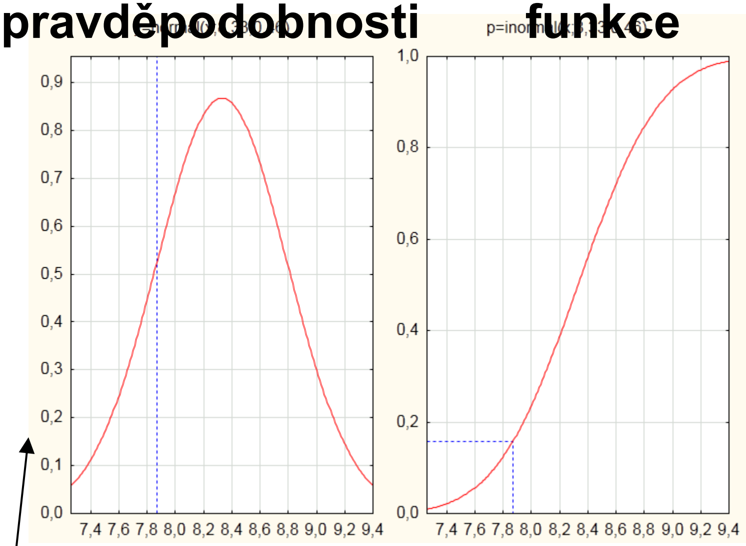
Cvičení 5

Výstup cvičení:

- V zadání uvést, které roky zpracováváte!
- Tabulka s vypočtenými údaji (viz vpravo)
- K bodům „a." a „e." obrázky grafů, které vytvoří program Statistica po zakliknutí „Vytv.graf" = 4 dvojobrázky
- Závěr:
  - „Vaše výsledky porovnejte s obecnými vlastnostmi frekvenční funkce normálního rozdělení (viz přednáška) a uveďte možné příčiny zjištěných rozdílů."
  - Tzn. prohlédněte si Vaši tabulku a spočtené hodnoty (pravděpodobnosti)
  - uvědomte si, že pracujete s průměrem směrodatná odchylka
  - Porovnejte spočtené pravděpodobnosti se spolužáky (máte je stejné/různé?)
  - Doplňují se třeba některé pravděpodobnosti (p1+p2 = 1)?

Průměr		8.33
Směrodatná odchylka		0.46
	Podmínka	Pravděpodobnost
a	T ≤ 7.87	
b	T ≤ 8.79	
c	T > 8.79	
d	7.87 ≤ T ≤ 8.79	
e	T ≤	0.1
	T ≤	0.5
	T ≤	0.9

Hustota pravděpodobnosti **Distribuční funkce**



Pravděpodobnost

Teplota vzduchu

(hodnoty, kterých může daný jev nabývat)

Zdroje

- BUDÍKOVÁ, Marie. Náhodné veličiny (přednáška). Brno: Masarykova univerzita, 26.9. 2016.

- DOBROVOLNÝ, Petr. Z1069 Statistické metody a zpracování dat: III. Pravděpodobnost, teoretická rozdělení (přednáška). Brno: Masarykova univerzita, 26.9.2016.

- PAVLÍK, Tomáš. Biostatistika pro matematickou biologii. <http://portal.matematickabiologie.cz/index.php?pg=aplikovana-analyza-klinickych-a-biologickych-dat--biostatistika-pro-matematickou-biologii--uvod-do-biostatistiky> 26.9.2016.

- STATSOFT. Rozdělení náhodné veličiny. <http://www.statsoft.cz/file1/PDF/newsletter/2012_11_12_StatSoft_Rozdeleni_nahodne_veliciny.pdf>. 26.9.2016.