**Vysoká škola báňská - Technická univerzita Ostrava**

**Fakulta elektrotechniky a informatiky**

**Katedra aplikované matematiky**

# Průvodce k programu Statgraphics

## Část 1

**Lenka Šimonová**

**Ostrava, 2006**

*Průvodce k programu Statgraphics* vznikl pro potřeby výuky předmětu Statistika I. na FEI VŠB-TU Ostrava, jak v prezenční tak v kombinované formě jako doplněk základní studijní opory, kterou je skriptum *Briš R., Litschmannová M.: Statistika I. pro kombinované a distanční studium, Ostrava 2004.*

*Průvodce k programu Statgraphics* ilustruje na příkladech řešených programem *Statgraphics* použití standardních statistických metod probíraných v předmětu Statistika I.. Podrobnější zdůvodnění použití odpovídajících statistických metod a vysvětlení jejich teoretického základu najde čtenář v již zmíněných skriptech *Briš R., Litschmannová M.: Statistika I. pro kombinované a distanční studium, Ostrava 2004* resp. jiné statistické literatuře – viz uvedený seznam literatury v závěru textu. *Průvodce k programu Statgraphics* není úplným manuálem k programu *Statgraphics*. Program *Statgraphics* obsahuje řadu dalších procedur, např. časové řady, které již nejsou v náplni předmětu Statistika I, tudíž nejsou ani zařazeny do tohoto textu.

Zdroje dat: v 1. a 2. kapitole jsou použita fiktivní data, v 3. kapitole jsou vyhodnocena data z balíku DataFile programu *Statgraphics*, 4. a 5. kapitola: Litschmannová M.: Statistika I. - Příklady, Ostrava 2000, 6. část vygenerovaná náhodná čísla programem *Statgraphics*, dále modifikovaná data z použité literatury, 7. kapitola ANOVA, příklady 1. a 2. : Friedrich V. : Statistika 1., Vysokoškolská učebnice pro distanční studium, Západočeská Univerzita, Plzeň 2002, 8. kapitola Regrese, příklad 1.: Novovičová J. : Pravděpodobnost a základy matematické statistiky, ČVUT Praha, 2002. Ostatní zdroje dat pro zpracování úloh ve Statgraphicsu byly internetové stránky statistického úřadu.

*Průvodce k programu Statgraphics část 1* obsahuje explorační analýzu dat a metody statistické dedukce, tj. hledání hodnot pravděpodobnostních, distribučních funkcí a kvantilů u daných typů rozdělení. *Průvodce k programu Statgraphics část 2* obsahuje metody statistické indukce, konkrétně testovaní parametrických a neparametrických hypotéz, konstrukce intervalových odhadů, jednofaktorovou analýzu rozptylu ANOVA a jednoduchou lineární regresi.

Autorka přeje studentům příjemné, ničím nerušené, studium předmětu Statistika I.

V Ostravě, 7.6.2006 Mgr. Lenka Šimonová

# 0. Jak spustit program

Program *Statgraphics* je na počítačových učebnách přímo nainstalován,  spustíte jej kliknutím na ikonu s názvem *sgwin*.

K demonstraci statistických pojmů a metod také můžete používat jiné softwary, např. program *Jump in* – starší verzi *Jmp3.01* nebo novější verzi *Jmp5.01*, český produkt *QCExpert*, *Statistica*, *SPSS*, … Příliš se nedoporučuje používat pouze program *Excel*, maximálně pouze k explorační analýze dat, tj. k první části učiva probíraného v předmětu Statistika I..

# 1. Vytvoření statistického souboru dat

Po spuštění programu *Statgraphics* se objeví okno **StatWizard**. Zvolme

**Analyze Existing Data or New Data/I Want To Enter New Data.**

Nejprve zadejme parametry sloupce, tj. typ proměnné. Označme název 1. sloupce např. „Student“a zvolme typ proměnné **Character** (slovní proměnná). Nyní vyplňme pole odpovídající 1. sloupci jmény studentů např. takto:

| **Student** |
|---|
| Martin |
| Jana |
| Rudolf |
| Jan |
| Petr |
| Jindřich |
| David |
| Libor |
| Bohumil |
| Oldřich |

Do 2. sloupce zadejme známky z matematické analýzy I. u zmíněných studentů. Sloupec nazvěme „Analýza“ (dvojité kliknutí na „Col 2“), typem proměnné bude **Integer**, neboť se jedná o celočíselnou proměnnou. Např.

| **Student** | **Analýza** |
|---|---|
| Martin | 2 |
| Jana | 3 |
| Rudolf | 2 |
| Jan | 2 |
| Petr | 3 |
| Jindřich | 3 |
| David | 3 |
| Libor | 2 |
| Bohumil | 3 |
| Oldřich | 2 |

Jako 3. proměnnou zvolme „Vzdálenost“ (vzdálenost bydliště studenta od školy), která je typem **Numeric** (číselná – nabývá hodnot z množiny reálných čísel). Např.

| **Student** | **Analýza** | **Vzdálenost** |
|---|---|---|
| Martin | 2 | 65 |
| Jana | 3 | 50 |
| Rudolf | 2 | 35 |
| Jan | 2 | 70 |
| Petr | 3 | 200 |
| Jindřich | 3 | 25 |
| David | 3 | 20 |
| Libor | 2 | 10 |
| Bohumil | 3 | 10 |
| Oldřich | 2 | 15 |

Dále přidejme 4. sloupec : typ mobilního operátora – „Mobil“, opět **Charakter** a 5. sloupec – provolané minuty za týden – „Hovory”: **Numeric.**

Například takto:

| **Student** | **Analýza** | **Vzdálenost** | **Mobil** | **Hovory** |
|---|---|---|---|---|
| Martin | 2 | 65 | T-mobile | 66 |
| Jana | 3 | 50 | Vodafone | 35 |
| Rudolf | 2 | 35 | O2 | 250 |
| Jan | 2 | 70 | O2 | 56 |
| Petr | 3 | 200 | Vodafone | 89 |
| Jindřich | 3 | 25 | O2 | 77 |
| David | 3 | 20 | T-mobile | 36 |
| Libor | 2 | 10 | O2 | 63 |
| Bohumil | 3 | 10 | T-mobile | 95 |
| Oldřich | 2 | 15 | O2 | 56 |

Vytvořený soubor uložte na lokálním disku pod názvem **Student**:

**Menu File/SaveAs/SaveDataFileAs**.

# 2. Jednorozměrná explorační analýza dat

Nyní budeme provádět samotnou explorační analýzu dat na vytvořeném souboru **Student**.

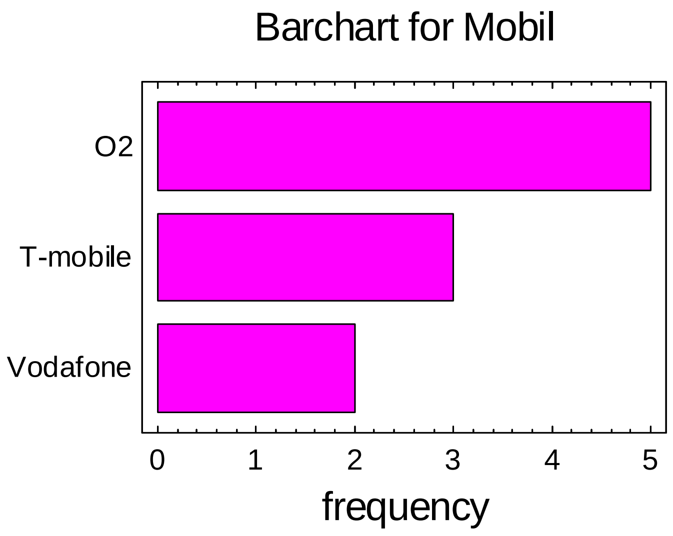
## a) Kategoriální proměnná

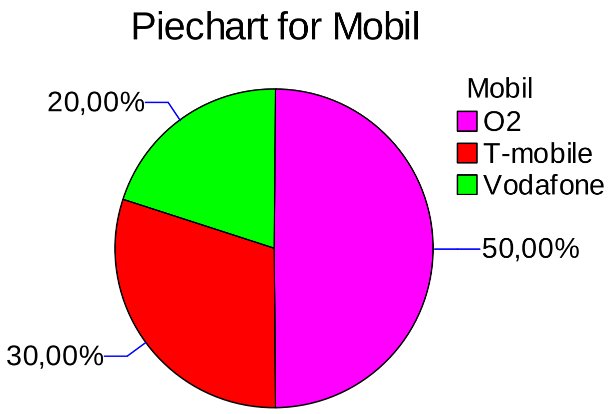
Začneme vyhodnocováním kategoriální proměnné, kterou je v našem případě např. proměnná „Mobil“.

**Menu Describe/Categorial Data/Tabulation**

Zvolit „Mobil” do **Data**.

Vykreslil se nám sloupcový (***barchart***) a koláčový graf (***piechart***), který udává počet (***frequency***) proměnné resp. procentuální zastoupení v jednotlivých kategorích této kategoriální proměnné - viz následující obrázek.



## b) Numerická proměnná

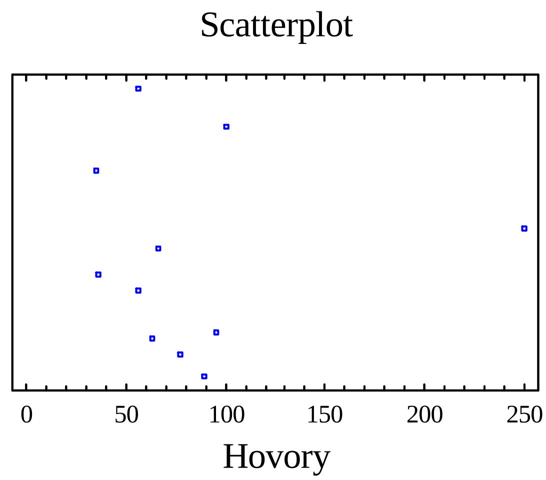
Analyzujme numerickou proměnnou „Hovory“.

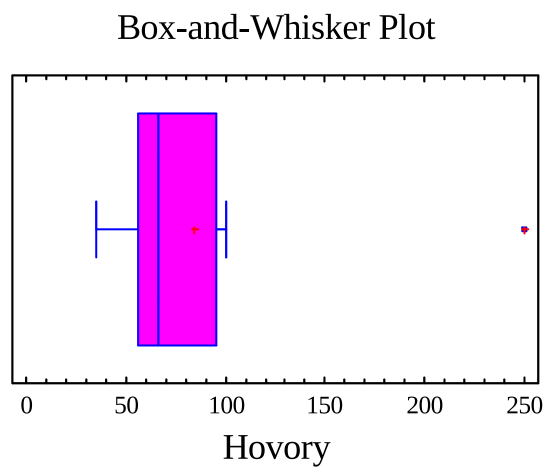
## i) Bodový a krabicový graf

**Menu Describe/Numeric Data/OneVariableAnalysis.**

Zvolit „Hovory” do **Data**.

Zobrazil se bodový graf (***scatter plot***) a krabicový graf (***box and whisker plot***) – viz následující obrázek.




Pokud chceme zobrazit více charakteristik proměnné „Hovory“, klikneme pravým tlačítkem myši na políčko zobrazených charakteristik, tj do levého dolního okna a zvolíme změnu údajů (**PaneOptions**). Přidáme například medián (**Median**), dolní (**Lower quartile**) a horní kvartil (**Upper quartile**), šikmost (**Standard skewness**) a špičatost (**Standard kurtosis**). Zobrazí se přehled charakteristik proměnné Hovory“:

***Summary Statistics for Hovory***
***Count = 11***
***Average = 83,9091***
***Median = 66,0***
***Variance = 3504,49***
***Standard deviation = 59,1987***
***Minimum = 35,0***
***Maximum = 250,0***
***Range = 215,0***
***Lower quartile = 56,0***
***Upper quartile = 95,0***
***Stnd. skewness = 3,43152***
***Stnd. kurtosis = 4,99702***

a slovní komentář v angličtině (pod heslem ***The StatAdvisor*** vždycky najdete slovní vysvětlení údajů, které jsou uvedeny v levém horním okně):

***The StatAdvisor***
***--------------------***
***This table shows summary statistics for Hovory. It includes measures of central tendency, measures of variability, and measures of shape. Of particular interest here are the standardized skewness and standardized kurtosis, which can be used to determine whether the sample comes from a normal distribution. Values of these statistics outside the range of -2 to +2 indicate significant departures from normality, which would tend to invalidate any statistical test regarding the standard deviation. In this case, the standardized skewness value is not within the range expected for data from a normal distribution. The standardized kurtosis value is not within the range expected for data from a normal distribution.***

Pokud neovládáte bravurně angličtinu, můžete komentáře pod heslem ***The StatAdvisor*** vynechávat a sami okomentovat číselné údaje, které se objevily v levém horním okně:

Počet pozorování (***count***) … 11
Průměr (***average***) … 83,9091 minut
Medián (***Median***) … 66,0 minut
Rozptyl (***variance***) .. 3504,49
Směrodatná odchylka (***standard deviation***) … 59,1987 minut
Minimum (***minimum***) … 35,0 minut
Maximum (***maximum***) … 250 minut
Interkvartilové rozpětí (***range***) … 215,0 minut
Dolní kvartil (***Lower quartile***) … 56,0 minut
Horní kvartil (***Upper quartile*** ) … 95,0 minut
Šikmost (***stnd. skewness***) … 3,43152
Špičatost (s***tnd. kurtosis***) … 4,99702

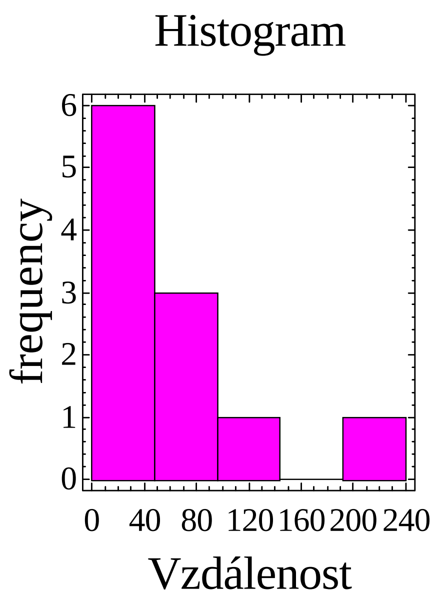
### ii) Histogram

Zobrazíme histogram rozložení četností pro numerickou proměnnou „Vzdálenost”.

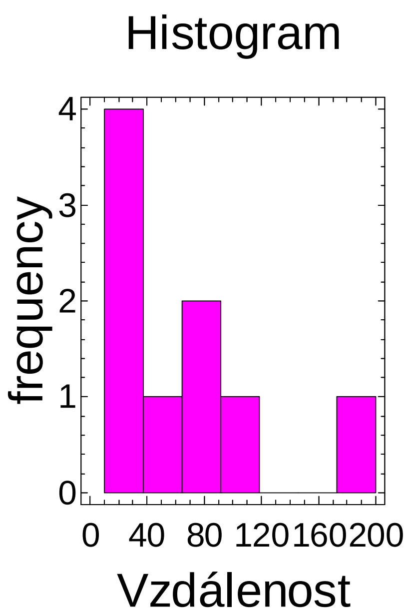
**Menu Plot/Exploratory Plot/Frequency Histogram**

(nebo Obrázková lišta: **Histogram**)

Zvolit: „Vzdálenost” do **Data**




Pokud chceme změnit počet dělících intervalů v histogramu, klikneme pravým tlačítkem myši na zobrazený histogram a pomocí volby **PaneOptions** nastavíme požadovaný počet dělících intervalů (**Numer of Classes**), např.7 a koncové body - dolní (**Lower Limit**), např. 10 a horní (**Upper Limit**) např. 200.



# 3. Dvourozměrná explorační analýza dat

Nyní na chvíli přerušíme práci s vytvořeným souborem **Student** a budeme se věnovat již vytvořenému souboru **Cardata** z datového balíku, který patří k softwaru *Statgraphics*, tj hledáte v adresáři **Statgra** na lokálním disku.

a) Pokud již máme program spuštěn, je vhodné nejprve zavřít doposud otevřené soubory a pak teprve otevřít nový soubor:

**Menu File/Open/OpenDataFile/TestData/Cardata.sf.**

b) Pokud program teprve spouštíme, můžeme použít tuto cestu:

Po spuštění programu *Statgraphics* se objeví okno **StatWizard**. Zvolme

**Analyze Existing Data or New Data/In an Existing StatFolio**

Oblast hledání: **Statgra/TestData**

Otevřít soubor: **Cardata**

Otevřeli jsme soubor dat s názvem **Cardata** (zde je uvedena zkrácená verze):

| mpg | cyl | dis | hp | accel | year | weight | origin | make |
|---|---|---|---|---|---|---|---|---|
| 43.1 | 4 | 90 | 48 | 21.5 | 78 | 1985 | 2 | Volkswagen |
| 36.1 | 4 | 98 | 66 | 14.4 | 78 | 1800 | 1 | Ford |
| 32.8 | 4 | 78 | 52 | 19.4 | 78 | 1985 | 3 | Mazda |
| 39.4 | 4 | 85 | 70 | 18.6 | 78 | 2070 | 3 | Datsun |
| 36.1 | 4 | 91 | 60 | 16.4 | 78 | 1800 | 3 | Honda |
| 19.9 | 8 | 260 | 110 | 15.5 | 78 | 3365 | 1 | Oldsmobile |
| 19.4 | 8 | 318 | 140 | 13.2 | 78 | 3735 | 1 | Dodge |
| 20.2 | 8 | 302 | 139 | 12.8 | 78 | 3570 | 1 | Mercury |
| 19.2 | 6 | 231 | 105 | 19.2 | 78 | 3535 | 1 | Pontiac |
| 20.5 | 6 | 200 | 95 | 18.2 | 78 | 3155 | 1 | Chevrolet |
| 20.2 | 6 | 200 | 85 | 15.8 | 78 | 2965 | 1 | Ford |
| 25.1 | 4 | 140 | 88 | 15.4 | 78 | 2720 | 1 | Ford |
| 20.5 | 6 | 225 | 100 | 17.2 | 78 | 3430 | 1 | Plymouth |
| 19.4 | 6 | 232 | 90 | 17.2 | 78 | 3210 | 1 | AMC |
| 20.6 | 6 | 231 | 105 | 15.8 | 78 | 3380 | 1 | Buick |
| 20.8 | 6 | 200 | 85 | 16.7 | 78 | 3070 | 1 | Mercury |
| 18.6 | 6 | 225 | 110 | 18.7 | 78 | 3620 | 1 | Dodge |
| 18.1 | 6 | 258 | 120 | 15.1 | 78 | 3410 | 1 | AMC |
| 19.2 | 8 | 305 | 145 | 13.2 | 78 | 3425 | 1 | Chevrolet |
| 17.7 | 6 | 231 | 165 | 13.4 | 78 | 3445 | 1 | Buick |

Procvičte si samostatně na souboru **Cardata** již procvičené procedury (1., 2.):

1. Analyzujte numerické proměnné **mpg** (mile per galon …kolik mil ujede na 1 galon pohonných hmot)**, horsepower** (koňská síla …výkon motoru)**.** Vykreslete box-plot a histogram rozdělení četností těchto numerických proměnných.

2. Analyzujte proměnnou **origin** (označení původu vyrobeného vozu …1 - Amerika, 2 - Evropa, 3 - Japonsko) jako kategoriální proměnnou. Vykreslete sloupcový a . koláčový graf.

Nyní přidejme nové úkoly:

3. Vyhodnoťme závislost mezi proměnnými **cylinders** a **origin** (dvě kategoriální proměnné).

4. Vyhodnoťme závislost mezi proměnnými **horsepower** a **origin** (numerická a kategoriální proměnná).

5. Vyhodnoťme závislost mezi proměnnými **horsepower** a **mpg** (dvě numerické proměnné).

6. Identifikujme odlehlá pozorování u proměnné **horsepower**.

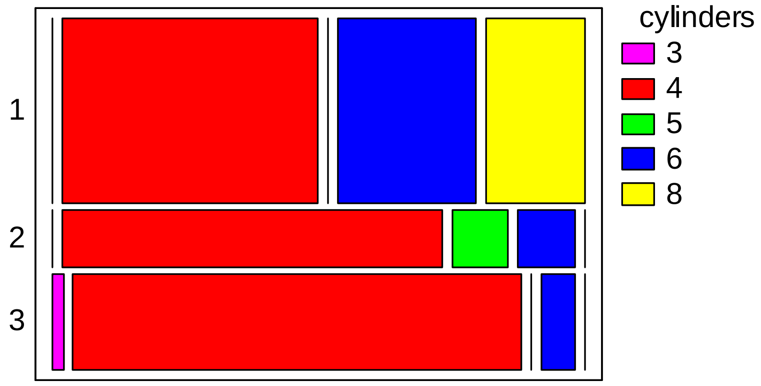
Ad 3. **Menu Describe/CategoricalData/Crosstabulation**

**Row Variable** … **origin**

**Column Variable** … **cylinders**

Grafickým výstupem je mozaikový graf (***mosaic chart***). Na první pohled vidíme rozdíl v umístění barev v jednotlivých pásech. Kdyby mozaikový graf obsahoval pouze svislé barevné pásy, tedy rozmístěné barev by bylo nezávislé na jednotlivých řádcích, znamenalo by to, že kategoriální proměnné jsou nezávislé. V našem případě vidíme, že proměnná **cylinders** závisí na proměnné **origin**. Tedy počet válců u auta závisí na zemi výroby. Americká auta (1) mají asi poloviční podíl čtyřválců a čtvrtinové podíly šesti a osmiválců, kdežto Evropská (2) a Japonská (3) auta mají jiné zastoupení aut co se týče počtu válců – viz zobrazený mozaikový graf.

Mosaic Chart for origin by cylinders



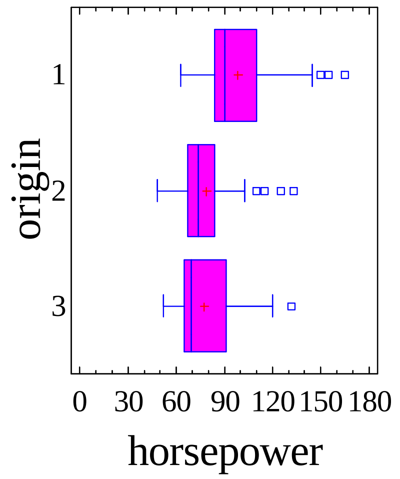

Ad 4. **Menu Plot/Exploratoryplots/Multi Box-and-Whiskers Plot**

**Data** … **horsepower**

**Level codes** … **origin**

Vykreslil se nám vícerozměrný krabicový graf. Na první pohled vidíme rozdíl v umístění krabicového grafu odpovídajícího země původu (origin) 1 a zemí původů 2, 3. Můžeme tedy předběžně říci, že **horsepower** (výkon motoru) je vyšší u automobilů vyrobených v Americe (1), než je u automobilů vyrobených v Evropě či Japonsku (2, 3). Podrobnější a přesnější analýze této závislosti se budeme věnovat v 3. části semestru v kapitole ***Analýza rozptylu (ANOVA).***

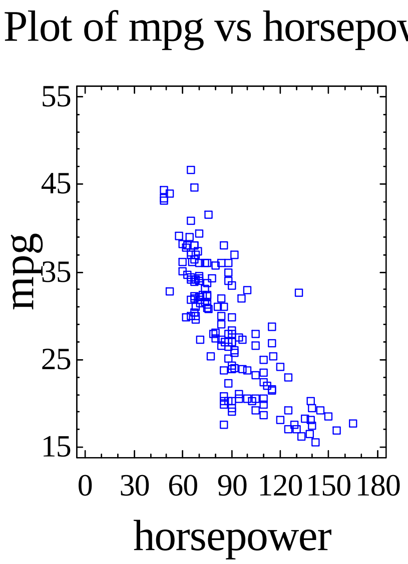
Box-and-Whisker Plot



Ad 5. **Menu Plot/Scatteplots/X-Yplot**

Y… **mpg,**

X… **horsepower.**




Zjišťujeme závislost mezi proměnnou **mpg (mile per galon)** a proměnnou **horsepower**. Vykreslil se dvourozměrný bodový graf. Již při prvním pohledu vidíme klesající tendenci. Čím je **horsepower** vyšší, tím je **mpg** nižší (čím je vyšší výkon motoru, tím více auto spotřebuje paliva a ujede méně mil s jednotkou pohonných hmot) a naopak. Bližší dvourozměrné analýze číselných proměnných se budeme věnovat v závěru semestru při učivu o ***Regresi.*** Prozatím tuto závislost můžeme vyhodnotit přibližně jako klesající.

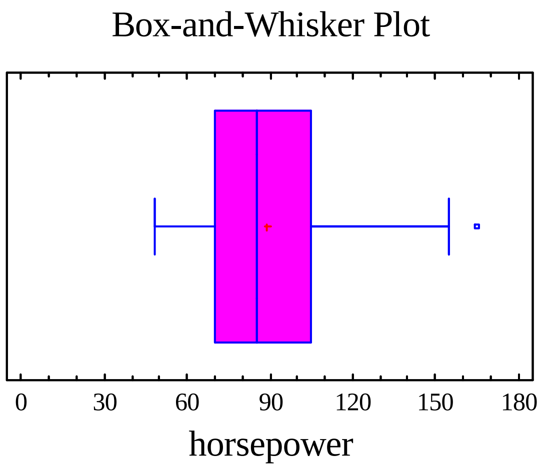
Ad 6. Identifikace odlehlých pozorování

Zaměřme se na jednorozměrnou proměnnou **horsepower**.

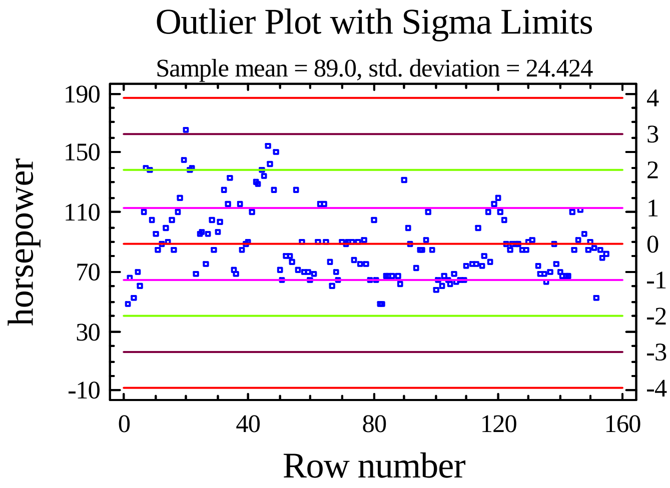
**Menu Describe/NumericData/OutlierIdentification**

**Data** … **horsepower**

Vidíme, že pozorování s hodnotou 165 je identifikováno jako odlehlé pozorování. Pokud se v souboru vyskytnou odlehlá pozorování máme dvě možnosti: buď budeme s odlehlým pozorováním dále pracovat anebo je vyjmeme z původního souboru, pokud si myslíme, že příliš ovlivní výsledky vyhodnocení anebo pokud se hodnota dostala do souboru nedopatřením (zmetek, španě údaj…).




V druhém obrázku vidíme podrobněji, že hodnota 165 leží mimo interval průměr plus minus třikrát směrodatná odchylka, tedy jedná se o odlehlé pozorování (pomocí z-souřadnice).



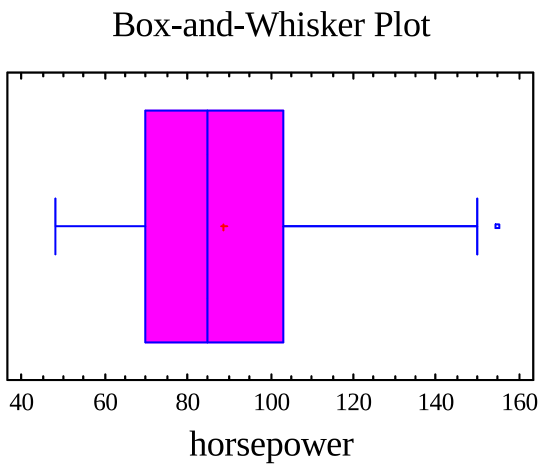

Zkusme v našem případě vyjmout ze souboru odlehlé pozorování. Postupujme následovně:

Označme řádek č. 20 (levou myší) – statistickou jednotku, u které se nachází údaj identifikovaný u této proměnné jako odlehlé pozorování. Pravou myší **Delete** smažeme tuto statistickou jednotku (řádek) s odlehlým pozorováním (celkem již není v souboru 155 pozorování, ale pouze 154). Na 20. místo se posunula další statistická jednotka.

Nově vytvořený soubor uložíme na lokálním disku pod názvem cardata_mdf:

**Menu File/SaveAs/SaveDataFileAs**

S novým souborem provedeme všechny předešlé analýzy (vykreslení box-plotu, histogramu, …). Software opět v souboru identifikoval odlehlé pozorování (tentokrát se jedná o hodnotu 155 – maximum, přesnou hodnotu najdeme vždy v textové části grafického výstupu), ale můžeme si všimnout, že rozdíl není již tak velký, jako byl v předchozím box-plotu. Opětovné vyjmutí odlehlého pozorování se již zpravidla neprovádí.



# Příklady k procvičení ke kapitolám 1 až 3

1. Proveďte analýzu dvourozměrných závislostí u Vámi vytvořeného souboru **Student**:

   a) závislost numerické proměnné na kategoriální proměnné,
   b) závislost dvou numerických proměnných,
   c) závislost dvou kategoriálních proměnných.

   Například můžete zkoumat ad a) zda je délka telefonního hovoru studenta závislá na typu jím používaného mobilního operátora, ad b) zda je délka telefonního hovoru studenta závislá na vzdálenosti jeho bydliště od školy, ad c) zda známka studenta z matematické analýzy závisí typu jím používaného mobilního operátora, …

2. Proveďte explorační analýzu datového souboru **Bodyfat.sf**, který se nachází v adresáři **\Stafgra\TestData** (**Open DataFile**).

3. Proveďte explorační analýzu datového souboru:

| Zaměstnanec | Pohlaví | Věk | Vzdělání | Funkce | Plat (v tis Kč) |
|---|---|---|---|---|---|
| 1 | Muž | 55 | VŠ | ředitel | 55 |
| 2 | Muž | 40 | VŠ | náměstek | 40 |
| 3 | Muž | 42 | VŠ | právník | 30 |
| 4 | Muž | 48 | SŠ | technik | 15 |
| 5 | Muž | 51 | SŠ | technik | 16 |
| 6 | Muž | 47 | SOU | dělník | 12 |
| 7 | Žena | 24 | SŠ | sekretářka | 15 |
| 8 | Žena | 45 | SOU | dělnice | 11 |
| 9 | Žena | 47 | SOU | dělnice | 12 |

   a) analyzujte rozložení platů,
   b) analyzujte rozložení vzdělání,
   c) zjistěte, zda má vzdělání vliv na plat,
   d) zjistěte, zda má věk vliv na plat.

4. Porovnejte „ruční“ výpočty provedené na úvodním cvičení a odpovídající výstupy ve *Statgraphicsu*.

5. Následující data představují platy zaměstnanců firmy XY:

14 659, 19 633, 15 899, 25 639, 56 496, 9 637, 12 567, 23 569, 19 639, 18 563.

Zkreslete graf stem and leaf

Nápověda: **Menu Describe/Numeric Data/OneVariableAnalysis**. V zobrazeném výstupu klikněte na žlutou ikonu (nahoře na liště, druhá zleva) a zadejte **Stem-and-Leaf Displey**.

# 4. Teoretická rozdělení pravděpodobnosti

## a) Binomické rozdělení

***Příklad 4.1.*** *Dlouhodobým pozorováním bylo zjištěno, že asi 30% uživatelů počítačů používá notebooky. Na školení nového softwarového produktu se sešlo 12 uživatelů počítačů. Předpokládejme, že všichni uživatelé, kteří používají notebook si jej vezmou s sebou na toto školení. Určete pravděpodobnost, že notebook s sebou budou mít:*

*a) ani jeden,*
*b) všichni,*
*c) právě jeden,*
*d) právě 3,*
*e) méně než tři,*
*f) více než tři.*

***Řešení:*** V zadání máme uvedeno, že $30\%$ uživatelů používá notebook. Označme těchto $30\%$ jako pravděpodobnost úspěchu $\mathrm{p} = 30\%$ (mezi všemi uživateli počítačů je $30\%$ „úspěšných”, tj. těch, kteří používají notebook). Celkem se školení má zúčastnit $12$ uživatelů, tedy označme tento počet $\mathrm{n} = 12$ jako celkový počet „pokusů“. Úkolem je určit počet (pravděpodobnost) „úspěšných pokusů“ mezi těmito *n* „pokusy“, jestliže známe pravděpodobnost úspěchu *p*.

Označme X … počet uživatelů, kteří budou mít s sebou na školení nového softwarového produktu notebook.

Úlohu můžeme zahrnout do tzv. <u>Bernoulliho pokusů</u>, tj. pokusů, které jsou

- nezávislé,
- každý z pokusů má pouze 2 možné výsledky úspěch/neúspěch,
- pravděpodobnost úspěchu *p* je konstantní.

Konvence: písmenem *n* budeme označovat celkový počet pokusů,
písmenem *k* budeme označovat počet úspěšných pokusů.

Nejprve si uvedeme přehled diskrétních rozdělení, které budeme dále používat:

| Rozdělení | Pravděpodobnostní funkce | Střední hodnota | Rozptyl | Interpretace |
|---|---|---|---|---|
| Binomické | $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ | $np$ | $np(1-p)$ | počet úspěchů v *n* pokusech |
| Geometrické | $P(X=n)=p(1-p)^{n-1}$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ | počet pokusů do 1. úspěchu |
| Negativně binomické | $P(X=n)=\binom{n-1}{k-1}p^k(1-p)^{n-k}$ | $\frac{k}{p}$ | $\frac{k(1-p)}{p^2}$ | počet pokusů do *k* úspěchů |
| Poissonovo | $P(X=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t}$ | $\lambda t$ | $\lambda t$ | počet událostí v době *t* |

Náhodná veličina X označuje počet úspěšných pokusů mezi *n* pokusy, bude tedy se řídit binomickým rozdělením. Její pravděpodobnostní funkce má tedy tvar

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

V našem případě (po dosazení za $n = 12$ a $p = 0{,}3$) dostáváme:

$$P(X = k) = \binom{12}{k} 0{,}3^k (1-0{,}3)^{12-k}$$

Hodnoty této pravděpodobnostní funkce pro jednotlivá *k* můžeme buď spočítat „ručně“anebo použít k výpočtu *Stagraphics*. V bodě a) si nejprve si naznačme postup při „ručním“ výpočtu, v ostatních bodech již budeme k výpočtu používat přímo *Statgraphics*.

Ad a) $P(X = 0) = ?$

Dosadíme do uvedeného vztahu za $k = 0$:

$$P(X = 0) = \binom{12}{0} 0{,}3^0 (1-0{,}3)^{12-0}.$$

Pomocí kalkulačky vypočteme

$$P(X = 0) = 0{,}0138413 \approx 1{,}4\%.$$

Jelikož výpočet pomocí kalkulačky je zdlouhavý, využijme dalším výpočtům již program *Statgraphics*. Aktivujme proceduru

**Menu Desribe/Distributions/ProbabilityDistributions**

a zvolme **Binomial.**

Nejprve nastavíme parametry binomického rozdělení (pravým tlačítkem myši: **Analysis Options**):

**Event Probability** … 0,3 - pravděpodobnost úspěchu p=30%

**Trials** … 12 - počet pokusů

Všimněme si výstupu v levém dolním okně:

***Cumulative Distribution***
***------------------------------***
***Distribution: Binomial***

| | ***Lower Tail Area (<)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***0*** | ***0,0*** | | | | |

***Probability Mass (=)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***0*** | ***0,0138413*** | | | | |

***Upper Tail Area (>)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***0*** | ***0,986159*** | | | | |

Hledanou pravděpodobnost najdeme na prostředním řádku:

***Probability Mass (=)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***0*** | ***0,0138413*** | | | | |

Tedy $P(X = 0) = \binom{12}{0} 0{,}3^{0} (1 - 0{,}3)^{12-0} = 0{,}0138413 \cong 1{,}4\%.$

Ad b) P(X=12)=?

Klikneme pravým tlačítkem myši na **variable**:

a změníme 0 na 12: (**Pane Options** … 12)

Dostaneme

***Cumulative Distribution***
***-------------------------------***
***Distribution: Binomial***

***Lower Tail Area (<)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***12*** | ***1,0*** | | | | |

***Probability Mass (=)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***12*** | ***5,31441E-7*** | | | | |

***Upper Tail Area (>)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***12*** | ***0,0*** | | | | |

Na prostředním řádku najdeme hledanou pravděpodobnost:

***<u>Probability Mass (=)</u>***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***12*** | ***5,31441E-7*** | | | | |

Tedy $P(X = 12) = \binom{12}{12} 0{,}3^{012} (1 - 0{,}3)^{12-12} = 5{,}31441\text{E-}7 \cong 0{,}0\%.$

Nyní již budeme postupovat rychleji:

Ad c) P(X=1)=?

**Variable** ... 1

**Probability Mass (=)** … 0,0711838

Tedy $P(X = 1) = \binom{12}{1} 0{,}3^1 (1 - 0{,}3)^{12-1} = 0{,}0711838 \cong 7{,}1\%.$

Ad d) P(X=3)=?

**Variable** ... 3

**Probability Mass (=)** … 0,2397

Tedy $P(X = 3) = \binom{12}{3} 0{,}3^3 (1 - 0{,}3)^{12-3} = 0{,}2397 \cong 24{,}0\%.$

Ad e) P(X<3)=?

***<u>Lower Tail Area</u> (<)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***3*** | ***0,252816*** | | | | |

Tedy $P(X < 3) = \sum_{k=0}^{2} \binom{12}{k} 0{,}3^k (1 - 0{,}3)^{12-k} = 0{,}252816 \cong 25{,}3\%.$

Ad f) P(X>3)=?

***<u>Upper Tail Area</u> (>)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***3*** | ***0,507484*** | | | | |

Tedy $P(X > 3) = \sum_{k=4}^{12} \binom{12}{k} 0{,}3^k (1 - 0{,}3)^{12-k} = 0{,}507484 \cong 50{,}7\%.$

## b) Geometrické rozdělení

***Příklad 4.2.*** *Jaká je pravděpodobnost, že aby padla šestka musíme hodit kostkou:*

*a) šestkrát,*

*b) jednou,*

*c) více než čtyřikrát.*

***Řešení:*** Označme X … počet pokusů potřebných k tomu, aby padla šestka.

Náhodná veličina X označuje počet pokusů nutných k dosažení 1.úspěchu, bude se tedy řídit geometrickým rozdělením s pravděpodobnostní funkcí

$P(X = n) = p(1-p)^{n-1}.$

Po dosazení za $p = \frac{1}{6} = 0{,}167$ dosatneme

$P(X = n) = 0{,}167(1-0{,}167)^{n-1}.$

Ad a) V šestém hodu má padnou šestka, tedy počet pokusů nutných k dosažení úspěšného hodu (počítáno včetně úspěšného hodu) má být šest:

$P(X = 6) = 0{,}167(1-0{,}167)^{6-1} \approx 6{,}7\%.$

**Pozor! Statgraphics je naprogramován tak, že proměnná *n* u Geometrického rozdělení znamená počet pokusů nutných k dosažení 1. úspěchu počítáno bez tohoto úspěchu, tedy počet pokusů předtím, než nastane úspěch.**

Tedy pokud k výpočtu použijeme *Statgraphics*, musíme nastavit počet pokusů před úspěšným hodem. *Statgraphics* vlastně používá alternativní vzorec pro pravděpodobnostní funkci geometrického rozdělení:

$P(X = n) = p(1-p)^{n}$, kde se symbolem *n* označuje počet pokusů před úspěšným pokusem.

**Menu Desribe/Distributions/ProbabilityDistributions**

Zvolíme **Geometric**

Nastavíme parametry geometrického rozdělení (pravé tlačítko myši: **Analysis Options**):

**Event Probability** … 0,167 - pravděpodobnost úspěchu $p=\frac{1}{6} \equiv 0{,}167\%$

V šestém hodu má padnou šestka, tedy počet pokusů před úspěšným hodem má být pět:

| | ***Probability Mass (=)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***5*** | ***0,0669794*** | | | | |

Tedy $P(X = 5) = 0{,}167(1-0{,}167)^{5} = 0{,}0669794 \cong 6{,}7\%.$

Ad b) P(X=1)=? V prvním hodu má padnou šestka, tedy <u>před</u> úspěšným hodem nemá být žádný hod:

| | ***Probability Mass (=)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***0*** | ***0,167*** | | | | |

$P(X=0)=0{,}167(1-0{,}167)^0 \approx 16{,}7\%$.

Ad c) P(X>3)=? Počet hodů než padne šestka má být větší než čtyři, tedy <u>před</u> prvním padnutím šestky má být více než 3 hody:

| | ***Upper Tail Area (>)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***3*** | ***0,481482*** | | | | |

Tedy $P(X>3)=1-P(X\leq 3)=\sum_{n=0}^{3} 0{,}167(1-0{,}167)^n \approx 48{,}1\%$.

## c) Negativně binomické rozdělení

***Příklad 4.3****. Jaká je pravděpodobnost, že aby nám padl 5x lev musíme hodit mincí:*

*a) desetkrát,*

*b) alespoň desetkrát,*

*c) nejvíce desetkrát.*

***Řešení:*** Označme X … celkový počet hodu mincí nutných k dosažení 5-ti úspěšných pokusů (padne lev).

Náhodná veličina X označuje celkový počet pokusů nutných k nastání *k* úspěšných pokusů, bude se tedy řídit <u>negativně binomickým rozdělením</u>, jehož pravděpodobnostní funkce je dána vztahem

$$P(X=n)=\binom{n-1}{k-1}p^k(1-p)^{n-k}.$$

V našem případě (po dosazení za k=5 a p=0,5) dostáváme:

$$P(X=n)=\binom{n-1}{5-1}0{,}5^5(1-0{,}5)^{n-5}.$$

Aktivujme proceduru

**Menu Desribe/Distributions/ProbabilityDistributions**

Zvolíme: **Negative Binomial**

Nastavíme parametry negativně binomického rozdělení (pravé tlačítko myši: **Analysis Options**):

**Event Probability** … 0,5 - pravděpodobnost úspěchu p=50%

**Success** … 5 - počet úspěchů

Ad a) P(X=10)=?

| | ***Probability Mass (=)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***10*** | ***0,123047*** | | | | |

Tedy $P(X=10)=\binom{10-1}{4}0{,}5^5(1-0{,}5)^{10-5}=0{,}123047\cong 12{,}3\%.$

Ad b) P(X$\geq$10)=P(X>9)=?

| | ***Upper Tail Area (>)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***9*** | ***0,5*** | | | | |

Tedy $P(X>9)=P(X\geq 10)=1-P(X<10)=1-\sum_{n=5}^{9}\binom{n-1}{4}0{,}5^5(1-0{,}5)^{n-5}=50\%.$

Ad c) P(X$\leq$10)=P(X<11)=?

| | ***Lower Tail Area (<)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***11*** | ***0,623047*** | | | | |

Tedy $P(X<11)=P(X\leq 10)=\sum_{k=5}^{10}\binom{n-1}{4}0{,}5^5(1-0{,}5)^{n-5}\approx 62{,}3\%.$

## d) Poissonovo rozdělení

Poissonovo a exponenciální rozdělení se řídí pravidlem Poissonovských pokusů, tj. pokusů, které jsou

- nezávislé,
- rychlost výskytu událostí $\lambda$ je v celém intervalu délky $t$ konstantní.

Konvence: písmenem $t$ budeme označovat délku časového intervalu,
písmenem $\lambda$ Budeme označovat počet událostí za časovou jednotku.

***Příklad 4.4.*** *Stroj vyrobí průměrně 2 zmetky za hodinu. Určete pravděpodobnost, že během 8-mi hodinové pracovní směny vyrobí stroj:*

*a) právě 16 zmetků,*
*b) právě 8 zmetků,*
*c) méně než 3 zmetky,*
*d) více než 10 zmetků.*

***Řešení:*** Označme X … počet vyrobených zmetků během 8-mi hodinové směny.

Náhodná veličina X označuje počet událostí, které nastanou během časového intervalu, bude se tedy řídit Poisssonovým rozdělením.

Zopakujme si předpis pro pravděpodobnostní funkci Poissonova rozdělení:

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

Dosadíme za $\lambda = 2$ (počet výskytů událostí za časovou jednotku) a $t = 8$ hodin (celková délka sledovaného časového intervalu), tedy $\lambda t = 16$. Dostaneme

$$P(X = k) = \frac{(16)^k}{k!} e^{-16}$$

Aktivujme proceduru

**Menu Desribe/Distributions/ProbabilityDistributions**

Zvolíme **Poisson**

**Mean** … 16 – střední hodnota počtu vyrobených zmetků během 8-mi hodinové směny

Ad a) P(X=16)=?

***Probability Mass (=)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***16*** | ***0,0992175*** | | | | |

$$P(X = 16) = \frac{(16)^{16}}{16!} e^{-16} \approx 9{,}9\% .$$

Zde si všimněte zajímavého výsledku: pravděpodobnost, že X nabude přesně střední hodnoty vyjde dosti malá oproti intuitivnímu očekávání.

Ad b) P(X=8)=?

| | ***Probability Mass (=)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| **8** | **0,0119875** | | | | |

$$P(X = 8) = \frac{(16)^{8}}{8!} e^{-16} \approx 1{,}2\% .$$

Ad c) P(X<3)=?

| | ***Lower Tail Area (<)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***3*** | ***0,0000163176*** | | | | |

$$P(X < 3) = \sum_{k=0}^{2} \frac{(16)^{k}}{k!} e^{-16} \approx 0{,}0\% .$$

.

V sumě je dolní mez nula, protože nejmenší možný počet událostí, které mohou nastat, je 0 (nemusí nastat žádná událost.).

Ad d) P(X>10)=?

| | ***Upper Tail Area (>)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***10*** | ***0,922604*** | | | | |

$$P(X > 10) = 1 - P(X \le 10) = 1 - \sum_{k=0}^{10} \frac{(16)^{k}}{k!} e^{-16} \approx 92{,}3\% .$$

.

## e) Exponenciální rozdělení

Exponenciální, Weibullovo a normální rozdělení jsou spojitá rozdělení, u kterých již nebudeme hledat pouze hodnoty pravděpodobnostních funkcí, ale převážně hodnoty distribučních funkcí a kvantilů.

***Příklad 4.5.*** *Výrobní zařízení má poruchu v průměru jednou za 2000 hodin. Jaká je pravděpodobnost, že přístroj bude pracovat bez poruchy déle než 550 hodin?*

***Řešení:*** Označme X … dobu mezi dvěma poruchami výrobního zařízení.

Jelikož X označuje dobu mezi dvěma po sobě jdoucími událostmi, bude se řídit <u>exponenciálním rozdělením.</u>

| | | | | |
|---|---|---|---|---|
| Exponenciální rozdělení | $F(t) = P(X < t) = 1 - e^{-\lambda t}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ | doba potřebná k uskutečnění *k* událostí |

Střední (průměrná) hodnota doby bezporuchového provozu je 2000 hodin:
$EX = \frac{1}{\lambda} = 2000$ hodin, tedy $\lambda = \frac{1}{2000}$ událostí za 1 hodinu.

Předpis pro distribuční funkci exponenciálního rozdělení:

$$F(t) = P(X < t) = 1 - e^{-\lambda t}$$

V našem případě:

$$F(t) = P(X < t) = 1 - e^{\frac{-t}{2000}}$$

Aktivujme proceduru

**Menu Desribe/Distributions/ProbabilityDistributions**

Zvolíme **Exponential**

Nastavíme parametry exponenciálního rozdělení (pravé tlačítko myši: **Analysis Options**):

**Mean** … 2000 - střední (průměrná) hodnota doby bezporuchového provozu

P(X>550)=?

***Upper Tail Area (>)***

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| ***550*** | ***0,759572*** | | | | |

$$P(X > 550) = 1 - F(550) = 1 - (1 - e^{-\frac{550}{2000}}) \approx 76{,}0\% .$$

## f) Weibullovo rozdělení

***Příklad 4.6.*** *Předpokládejme, že doba do poruchy určitého systému je modelována Weibullovým rozdělením s klesající intenzitou poruch, parametry: $\lambda = 0.02$; $\beta = 0.5$.*

*a) Jaká je pravděpodobnost, že systém bude pracovat bez poruchy během prvních 10-ti hodin?*

*b) Jaká je pravděpodobnost, že systém bude pracovat bez poruchy během prvních 200 hodin?*

***Řešení:*** Označme X … dobu do poruchy (délku bezporuchového provozu).

V zadání máme přímo uvedeno, že náhodná veličina X je modelována Weibullovým rozdělením, které je zobecněním exponenciálního rozdělení.

**Menu Desribe/Distributions/ProbabilityDistributions**

Zvolíme **Weibull**

Nastavíme parametry Weibullova rozdělení (pravé tlačítko myši: **Analysis Options**):

**Scale** … $\frac{1}{\lambda} = \frac{1}{0{,}02} = 50$ - parametr měřítka

**Shape** … $\beta = 0.5$ - parametr tvaru

Označme X … doba do poruchy sledovaného systému

Ad a) Určujeme pravděpodobnost, že systém bude pracovat bez poruchy během prvních 10-ti hodin, tedy že doba do poruchy bude delší než těchto 10 hodin:

P(X>10)=?

***Upper Tail Area (>)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***10*** | ***0,639407*** | | | | |

P(X>10)= 0,639407 =63,9%.

Ad b) Určujeme pravděpodobnost, že systém bude pracovat bez poruchy během prvních 200 hodin, tedy že doba do poruchy bude delší než těchto 200 hodin:

P(X>200)=?

***Upper Tail Area (>)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***200*** | ***0,135335*** | | | | |

P(X>200)= 0,135335 $\cong$ 13,5%.

## g) Normální rozdělení

***Příklad 4.7.*** *Nechť X je náhodná veličina s normálním rozdělením se střední hodnotou 6 a rozptylem 49. Určete:*
*a) P(X<7) ,*
*b) P(X>9) ,*
*c) P(5<X<10).*

***Řešení:*** V zadání je přímo uvedeno, že se jedná o náhodnou veličinu s normálním rozdělením.

**Menu Desribe/Distributions/ProbabilityDistributions**

Zvolíme **Normal**

Nastavíme parametry normálního rozdělení (pravé tlačítko myši: **Analysis Options**):

**Mean** … 6 - střední hodnota

**Std Dev** … 7 - směrodatná odchylka, tj. druhá odmocnina z rozptylu

Ad a) P(X<7)=?

***Cumulative Distribution***
***-------------------------------***
***Distribution: Normal***

***Lower Tail Area (<)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***7*** | ***0,556801*** | | | | |

Tedy $P(X<7)= F(7)=0{,}556801 \cong 55{,}7\%$ .

Ad b) P(X>9)=?

***Upper Tail Area (>)***

| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
|---|---|---|---|---|---|
| ***9*** | ***0,334116*** | | | | |

Tedy $P(X>9)= 1-F(9)=0{,}334116 \cong 33{,}4\%$.

Ad c) P(5<X<10)=?.

$P(5<X<10)= P(X<10) - P(X \le 5)=P(X<10) - [1 - P(X>5)]=[P(X<10)+P(X>5) - 1]$,
(udělejte si náčrtek časové osy)

neboli z pravděpodobnostních pravidel plyne:
$P(A \cap B)=P(A)+P(B) - P(A \cup B)$, dosadíme $A=(X>5)$, $B=(X<10)$ a dostaneme
$P(5<X<10)=P(X>5 \wedge X<10)=P(X>5)+P(X<10) - P(X>5 \vee X<10)= [P(X<10)+P(X>5) - 1]$.

Nejprve určíme P(X<10) a P(X>5) a potom vypočteme hodnotu výrazu P(X<10) +P(X>5) –1.

i) P(X<10)=?

| | ***Lower Tail Area (<)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***10*** | ***0,716147*** | | | | |

Tedy P(X<10)= 0,716147,

ii) P(X>5)=?

| | ***Upper Tail Area (>)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***5*** | ***0,556801*** | | | | |

Tedy P(X>5)= 0,556801,

iii) Tedy P(5<X<10)= [P(X<10) +P(X>5) –1]= [0,716147+ 0,556801–1]=0,272948 $\cong$ 27,3%.

***Příklad 4.8**. Nechť X je náhodná veličina s normálním rozdělením se střední hodnotou 5 a rozptylem 4.*

*Najděte: a)* $x_{0,1}$ *- 10 % kvantil ,*

*b)* $x_{0,5}$ *- medián,*

*c)* $x_{0,75}$ *- 75 % kvantil.*

***Řešení:*** V zadání je přímo uvedeno, že se jedná o náhodnou veličinu s normálním rozdělením.

**Menu Desribe/Distributions/ProbabilityDistributions**

Zvolíme **Normal**

Nastavíme parametry normálního rozdělení (pravé tlačítko myši: **Analysis Options**):

**Mean** … 5 - střední hodnota,

**Std Dev** … 2 - směrodatná odchylka (druhá odmocnina z rozptylu).

Nyní máme „opačný úkol“ než doposud. Hledali jsme hodnoty pravděpodobnostní nebo distribuční funkce. Nyní tyto hodnoty distribučních funkcí známe (je to těch 10 %, 50 % a 75 %) a budeme dohledávat proměnnou *x*, pro kterou platí: *F(x)=p*, kde p=0,10 (resp. 0,5, resp. 0,75).

V otevřeném okně **Probability Distributions** klikneme na žlutou ikonu: objeví se okno **Tabular Options,** ve kterém aktivujeme **Inverse CDF** (Cumulative Distribution Funkction)

Dostaneme následující tabulku kvantilů:

***Inverse CDF***
***----------------***
***Distribution: Normal***

| *CDF* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| *0,01* | *0,347296* | | | | |
| *0,1* | *2,43689* | | | | |
| *0,5* | *5* | | | | |
| *0,9* | *7,56311* | | | | |
| *0,99* | *9,6527* | | | | |

Ad a) $x_{0,1}$=?

10 % kvantil se v zobrazené tabulce nachází ve 2. řádku:

***Inverse CDF***
***------------------***

| *CDF* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| *0,1* | *2,43689* | | | | |

Tedy $x_{0,1}=2{,}43689 \cong 2{,}44$

Ad b) $x_{0,5}$=?

Medián se nachází na prostředním řádku tabulky:

***Inverse CDF***
***------------------***

| *CDF* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| *0,5* | *5* | | | | |

Tedy $x_{0,5}=5$ (u normálního rozdělení je medián vždy roven střední hodnotě).

Ad c) $x_{0,75}$=?

V tabulce řádek pro 75 % kvantil není explicitně zadán, změníme tedy např. zadaný 90 % kvantil na 75 % kvantil … pravým tlačítkem myši … **Pane Options**

Dostaneme novou tabulku

***Inverse CDF***
***----------------***

| *CDF* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| *0,01* | *0,347296* | | | | |
| *0,1* | *2,43689* | | | | |
| *0,5* | *5* | | | | |
| *0,75* | *6,34898* | | | | |
| *0,99* | *9,6527* | | | | |

ve které je hledaný 75 % kvantil na 4. řádku, tedy

$x_{0,75}=6{,}34898 \cong 6{,}35$.

***Příklad 4.9.*** *Doba potřebná k objevení a odstranění poruchy stroje se řídí normálním rozdělením se střední hodnotou 40 minut a směrodatnou odchylkou 30 minut. Jaká je pravděpodobnost, že doba potřebná k objevení a odstranění poruchy stroje nepřekročí 1 hodinu?*

***Řešení:*** V zadání je přímo uvedeno, že se jedná o náhodnou veličinu s normálním rozdělením.

**Menu Desribe/Distributions/ProbabilityDistributions**

Zvolíme **Normal**

Nastavíme parametry normálního rozdělení (pravé tlačítko myši: **Analysis Options**):

**Mean** … 40 - střední hodnota (v minutách),

**Std Dev** … 30 - směrodatná odchylka (v minutách).

P(X<60 minut)=?

| | ***Lower Tail Area (<)*** | | | | |
|---|---|---|---|---|---|
| ***Variable*** | ***Dist. 1*** | ***Dist. 2*** | ***Dist. 3*** | ***Dist. 4*** | ***Dist. 5*** |
| ***60*** | ***0,747509*** | | | | |

P(X<60)= F(60)=0,747509 ≅ 74,8%.

# Generování náhodných čísel

Na závěr 4. kapitoly si ukážeme, jak program Statgraphics umožňuje vygenerovat náhodná čísla podléhající zvolenému typu rozdělení.

***Příklad 4.10.*** *Vygenerujte ve Statgraphicsu náhodná čísla podléhající následujícím typům rozdělení:*

*a) N(8,9)*
*b) Exp (4)*
*c) Weibull (50,4)*

***Řešení:*** ad a) **Menu Desribe/Distributions/ProbabilityDistributions**

Zvolíme **Normal** s parametry $\mu = 8, \sigma = 3$.

Dále klikneme na žlutou ikonu (**Tabular Options**), aktivujeme **Random Numbers.**

Nyní klikneme na čtvrtou ikonu s obrázkem diskety (**Save Results Options**), zvolíme označení sloupce „Normal“ a potvrdíme („odfajfkujeme“) **Save**. V původní tabulce zadaných hodnot (doposavad prázdné) se objeví v 1. sloupci vygenerovaná náhodná čísla odpovídající normálnímu rozdělení s parametry $\mu = 8, \sigma = 3$.

Dále postupujeme analogicky:

ad b) **Menu Desribe/Distributions/ProbabilityDistributions**

Zvolíme **Exponential** s parametrem $\lambda = 4$ (pozor! volíme **Mean**=0,25).

Sloupec s vygenerovanými náhodnými čísly podléhající exponenciálnímu rozdělení s parametrem $\lambda = 4$ označíme „Exp“.

ad c) **Menu Desribe/Distributions/ProbabilityDistributions**

Zvolíme **Weibull** s parametry $\Theta = 50, \lambda = 4$.

Sloupec s vygenerovanými náhodnými čísly podléhající Weibullovu rozdělení s parametry $\Theta = 50, \lambda = 4$ označíme „Weibull“.

Dostaneme tabulku o třech sloupcích (Normal, Exp, Weibull) a sto řádcích podobnou následující tabulce (zkrácená verze). Uložte ji pod názvem „random_numbers“.

| *Normal* | *Exp* | *Weibull* |
|---|---|---|
| *10,2402* | *0,447763* | *4,00039* |
| *4,42069* | *0,113514* | *4,06081* |
| *12,2046* | *0,840781* | *3,85664* |
| *13,2463* | *0,206974* | *3,97788* |
| *6,94946* | *0,0230306* | *3,9478* |
| *9,08386* | *0,152326* | *3,95748* |
| *5,20644* | *0,028362* | *4,01216* |
| *5,83704* | *0,123523* | *4,06196* |
| *10,6849* | *0,02282* | *3,97438* |
| *4,18501* | *0,210294* | *3,86028* |
| *7,35329* | *0,607451* | *4,01181* |
| *9,23364* | *0,121714* | *3,95713* |
| *7,08214* | *0,163455* | *3,77711* |
| *8,08223* | *0,12948* | *3,9599* |
| *7,33378* | *0,427409* | *3,91462* |
| *7,52937* | *0,0809073* | *3,69909* |
| *11,4342* | *0,584463* | *4,02485* |
| *8,53363* | *0,0385625* | *3,95242* |
| *3,38411* | *0,0361806* | *3,85564* |
| *10,9857* | *0,0223349* | *4,08423* |
| *7,29247* | *0,6203* | *3,99103* |
| *11,8511* | *0,351747* | *3,96093* |
| *5,66421* | *0,0803284* | *3,82474* |
| *9,98491* | *0,0282136* | *3,73782* |

# Příklady k procvičení ke kapitole 4

1. Při provozu balícího automatu vznikají během směny náhodné poruchy. Ze zkušenosti víme, že během směny dochází v průměru ke 2 poruchám. Jaká je pravděpodobnost, že během 24 hodin (třísměnného provozu) nedojde ani jednou k poruše?

2. Průměrná doba mezi příjezdy nákladních automobilů s betonovou směsí je 10 minut. Jaká je pravděpodobnost, že doba mezi příjezdy dvou vozidel bude kratší než 7 minut?

3. Student složí zkoušku, jestliže v testu odpoví správně alespoň na čtyři z pěti otázek. U každé otázky jsou čtyři možné odpovědi, z nichž jediná je správná. S jakou pravděpodobností student složí zkoušku, jestliže se vůbec nepřipravoval a odpovědi volil náhodně?

4. Student se má ke zkoušce naučit 60 otázek. Z nedostatku času se naučil jen 40. U zkoušky si vylosuje 3 otázky. S jakou pravděpodobností:
   a) bude umět alespoň dvě otázky?
   b) nebude umět ani jednu otázku?

5. Revizor ze zkušenosti ví, že zhruba v 26% tramvají při kontrole najde černého pasažéra. Kolik tramvají musí zkontrolovat, aby alespoň s 95% pravděpodobností našel alespoň jednoho černého pasažéra?

6. V jednom mililitru určitého dokonale rozmíchaného roztoku se v průměru nachází 15 určitých mikroorganismů. Určete pravděpodobnost, že při náhodném výběru vzorku o objemu 1/2 mililitru bude ve zkumavce méně než 5 těchto mikroorganismu.

7. Výrobní zařízení má poruchu v průměru jednou za 2000 hodin. Jaká je pravděpodobnost, že přístroj bude pracovat déle než 550 hodin?

8. Odhadujeme, že střední životnost určitého přístroje je 110 dnů. S jakou pravděpodobností bude životnost náhodně vybraného přístroje mezi 100 a 150 dny?

9. Hodinová dopravní intenzita na určitém místě dálnice v určitou denní dobu je 300 vozidel. S jakou pravděpodobností projede tímto místem během jedné minuty více než 6 vozidel?

10. Počet návštěvníků Fitness Centra VŠB je v průměru 10 na hodinu. Jaká je pravděpodobnost, že bude během určité hodiny ve Fitness Centru VŠB přesně 10 lidí?

11. Doba do vybití baterie se řídí exponenciálním rozdělením.
   a) Jaká je střední doba do vybití, víme-li, že 1% těchto baterií vydrží déle než 4000 hodin?
   b) Je-li střední doba do vybití 3.150 hodin, kolik procent těchto baterii vydrží déle než 4000 hodin?

12. Doba potřebná k vypracování písemky ze statistiky má normální rozdělení se střední hodnotou 45 minut a směrodatnou odchylkou 10 minut.
   a) Kolik procent studentů dokončí test do jedné hodiny?
   b) Jak dlouho by měl test trvat, aby jej dokončilo 99 % studentů?

# Slovníček některých anglických termínů

| **Anglicky** | **Česky** |
|---|---|
| Variable | Proměnná |
| Observation | Pozorování |
| Plot | Graf, vykreslit |
| Scaterplot | Bodový graf |
| Describe | Popsat |
| Compare | Porovnat |
| Relation | Závislost |
| Relate | Najít závislost |
| Simple regression | Jednoduchá regrese |
| Frequency | Četnost |
| Frequency histogram | Histogram rozdělení čeností |
| Average, SampleMean | Průměr |
| Standard deviation (St Dev) | Směrodatná odchylka |
| Count | Počet |
| Skewness | Šikmost |
| Kurtosis | Špičatost |
| z-score | z-souřadnice |
| Reject $H_0$ | Zamítáme $H_0$ |
| Do not reject $H_0$ | Nezamítáme $H_0$ |

# Literatura


1. Anděl J. : Matematická statistika, Praha, SNTL, 1978

2. Briš R., Litschmannová M. : Statistika I. Pro kombinované a distanční studium, VŠB-TU Ostrava, 2004,

3. Cyhelský L., Kalounová J., Hindls R. : Elementární statistická analýza, Management Press Praha, 1996,

4. Dupač V., Hušková M. : Pravděpodobnost a matematická statistika, Karolinum, Praha, 2001

5. Dummer M. : Introduction to Satistical Sience, VŠB-TU Ostrava, 1998,

6. Dummer M., Klímková M. : Statistika I. (cvičení), VŠB-TU Ostrava, 1997,

7. Friedrich V. : Statistika 1., Vysokoškolská učebnice pro distanční studium, Západočeská Univerzita, Plzeň 2002,

8. Hebák P., Kahounová J. : Počet pravděpodobnosti v příkladech, SNTL Praha, 1988

9. Hebák P., Hustopecký J., Jarošová E., Pecáková I. : Vícerozměrné statistické metody (1), (2), (3), Informatorium Praha, 2004

10. Hindls R., Hronová S., Seger J. : Statistika pro ekonomy, Professional Publishing Praha, 2004

11. Kunderová P.: Úvod do teorie pravděpodobnosti a matematické statistiky, Olomouc, 1997,

12. Křivý I. : Úvod do teorie pravděpodobnosti, Ostravská Univerzita, 1983,

13. Křivý I. : Základy matematické statistiky, Ostravská Univerzita, 1985,

14. Likeš J., Cyhelský L., Hindls R. : Úvod do statistiky a pravděpodobnosti, VŠE Praha, 1994

15. Likeš J., Machek J. :Počet pravděpodobnosti, SNTL Praha, 1982,

16. Likeš J., Machek J. : Matematická statistika, SNTL Praha, 1988,

17. Litschmannová M. : Statistika I. - příklady, VŠB-TU Ostrava, 2000,

18. Novovičová J. : Pravděpodobnost a základy matematické statistiky, ČVUT Praha, 2002

19. Riečan B. : Pravděpodobnost a matematická statistika, Bratislava

20. Riečan B, Neubrunn T. : Teória miery, Bratislava, 1992