



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Výuka předmětu Matematika 4 s využitím statistického software Minitab

Odbor statistiky a optimalizace

ÚM FSI VUT v Brně

Brno

2022





EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



Studium moderní a rozvíjející se techniky VUT

CZ.02.2.69/0.0/0.0/18_056/0013325

Toto je učební opora rozšiřuje skriptu KARPÍŠEK, Zdeněk. Matematika IV. 4., přeprac. vyd. Brno: Akademické nakladatelství CERM, 2014. ISBN 9788021448582 o využití statistického software Minitab. Kapitoly a čísla příkladů jsou převzaty z výše uvedené publikace.



1 Popisná statistika

Řešený příklad 1.1

Měřením délky X (mm) 10 válečků byly získány hodnoty: 5,38; 5,36; 5,35; 5,40; 5,41; 5,34; 5,29; 5,43; 5,42; 5,32. Určete rozsah, variační obor, variační rozpětí, aritmetický průměr, rozptyl, směrodatnou odchylku, variační koeficient a medián statistického souboru.

Ř e š e n í:

Rozsah daného souboru je $n = 10$, takže nemá smysl jej třídit. Protože $x_{(1)} = 5,29$ mm a $x_{(10)} = 5,43$ mm, je variační obor $<5,29; 5,43>$ mm a variační rozpětí je $5,43 - 5,29 = 0,14$ mm. Dále je:

$$\bar{x} = (5,38 + \dots + 5,32)/10 = 53,70/10 = 5,37 \text{ mm} \dots \text{průměrná délka,}$$

$$s^2 = (5,38^2 + \dots + 5,32^2)/10 - 5,37^2 = 288,388/10 - 28,8369 = 0,0019 \text{ mm}^2,$$

$$s = \sqrt{0,0019} \approx 0,0435889894 \approx 0,044 \text{ mm,}$$

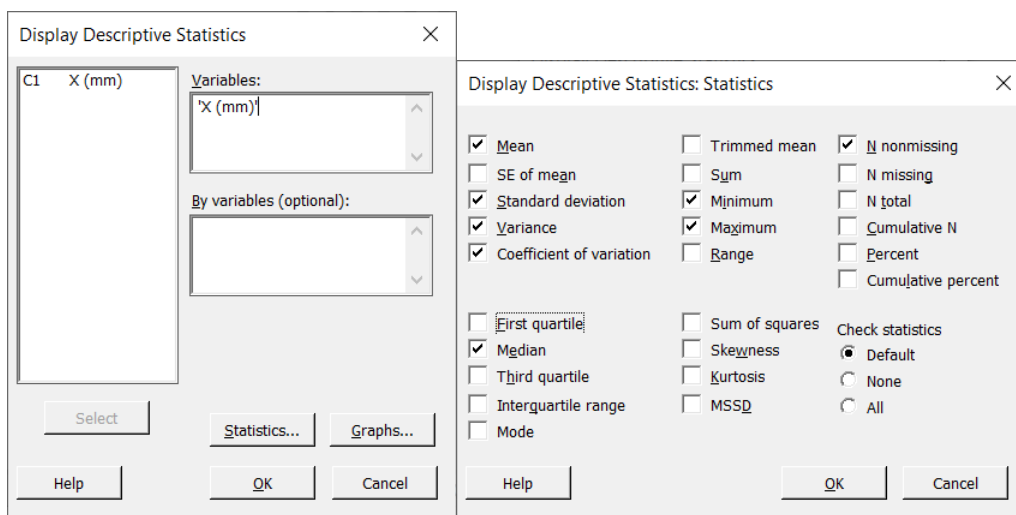
$$v = \sqrt{0,0019}/5,37 \approx 0,0435889894/5,37 \approx 0,00811713 \approx 0,8117 \%,$$

$$\tilde{x} = (5,36 + 5,38)/2 = 5,37 \text{ mm} \dots \text{medián délky.}$$

Pro grafické vyjádření tohoto statistického souboru by byl vhodný krabicový graf.

Postup v Minitabu:

Stat > Basic Statistics > Display Descriptive Statistics





EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MSMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Display Descriptive Statistics: Graphs

☐ Histogram of data

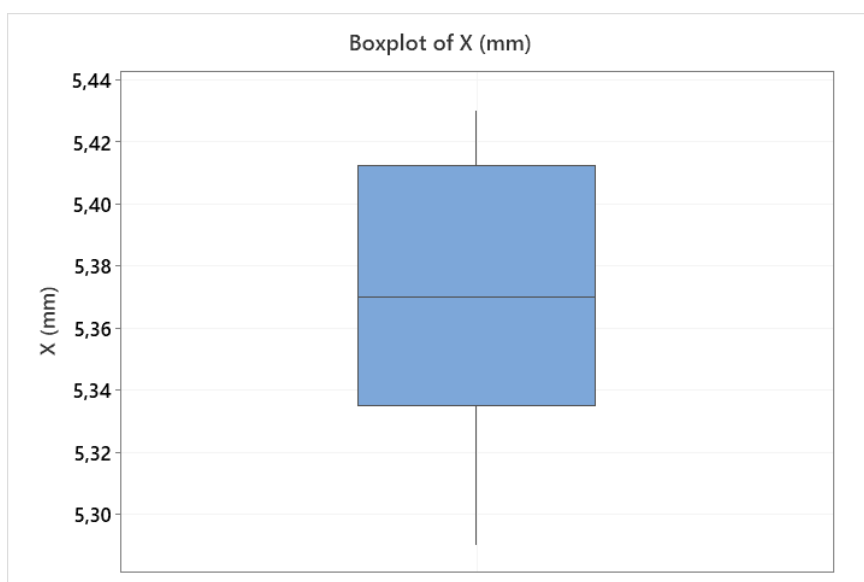
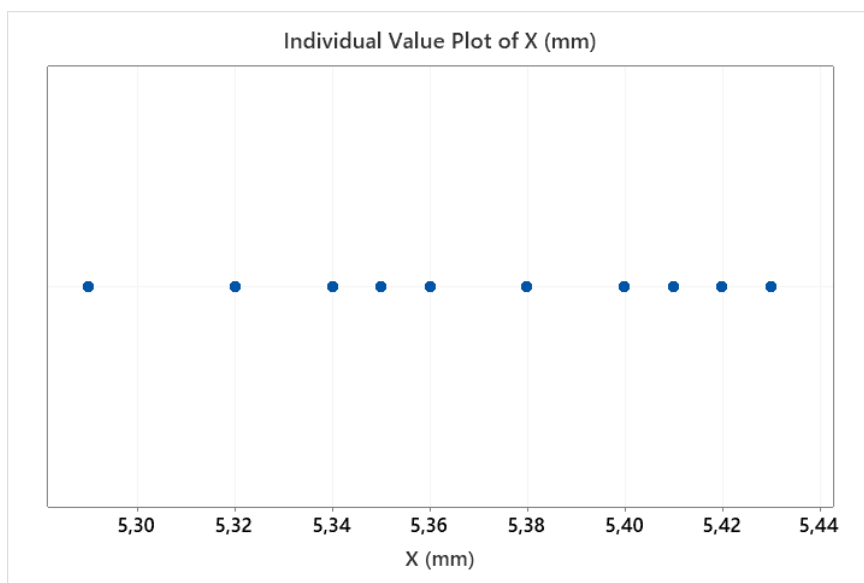
☐ Histogram of data, with normal curve

☒ Individual value plot

☒ Boxplot of data

Help OK Cancel

Grafický výstup:





Tabulkový výstup:

Statistics

Variable	N	Mean	StDev	Variance	CoefVar	Minimum	Median	Maximum
X (mm)	10	5,3700	0,0459	0,0021	0,86	5,2900	5,3700	5,4300

Při výpočtech se také užívá jiný vzorec pro rozptyl, když výraz $\frac{1}{n}$ zaměníme výrazem $\frac{1}{n-1}$. Takto vypočtený rozptyl je roven číslu $\frac{n}{n-1}s^2 > s^2$ (pro $s^2 \neq 0$). Zdůvodnění výrazu $\frac{1}{n-1}$ plyne z požadavků uvedených v kapitole 6 a 7.

Řešený příklad 1.2

Při kontrole byl zjišťován objem nápoje X v 50 lahvích a byly naměřeny následující odchylky (ml) od hodnoty na etiketě:

1,2; 2,1; 1,7; 0,9; 0,3; 2,0; -1,3; -0,1; 3,2; 2,8;
0,8; 4,4; 2,9; 1,2; 0,0; -2,3; 1,2; 0,9; 2,3; -0,2;
0,1; 1,9; -1,9; -0,2; -1,3; 1,5; 0,5; 2,0; -1,3; 3,7;
0,9; 1,0; 0,4; 1,9; 1,4; -1,3; 1,6; 1,4; 3,1; -0,1;
1,8; 0,0; 4,1; 1,3; 3,0; 0,4; 3,8; -0,8; 3,1; 0,9.

Roztřídte daný statistický soubor, graficky jej znázorněte a vypočtěte \bar{x} , s^2 , s , \hat{x} , A .

Ř e š e n í:

Rozsah souboru $n = 50$; $x_{(1)} = -2,3$ ml a $x_{(50)} = 4,4$ ml, takže variační obor je $<-2,3; 4,4>$ ml a rozpětí je $4,4 - (-2,3) = 6,7$ ml. Volíme počet tříd $m = 7$ (tj. asi $\sqrt{50}$) a délku třídy $h = 1$ (tj. asi $6,7/7$). Volba tříd a jejich středů, roztrídění do tříd a výpočet absolutních a kumulativních četností je v následující tabulce, kde např. // značí 2 hodnoty a /// značí 5 hodnot ležících v dané třídě:

j	třída	x_j^*	zařazení do tříd	f_j	F_j
1	-2,5; -1,5	-2	//	2	2
2	-1,5; -0,5	-1	///	5	7
3	-0,5; 0,5	0	/// /// /	11	18
4	0,5; 1,5	1	/// /// ///	13	31
5	1,5; 2,5	2	/// ///	9	40
6	2,5; 3,5	3	/// /	6	46
7	3,5; 4,5	4	////	4	50



Histogramy a polygony tohoto statistického souboru jsou na obr. 1.4 a 1.5. Další výpočty jsou pro přehlednost znázorněny v následující tabulce, ze které dostaneme:

$$\bar{x} = 56/50 = 1,12 \text{ ml}; s^2 = 180/50 - 1,12^2 = 2,3456 \text{ ml}^2; s = \sqrt{2,3456} \approx 1,532 \text{ ml};$$

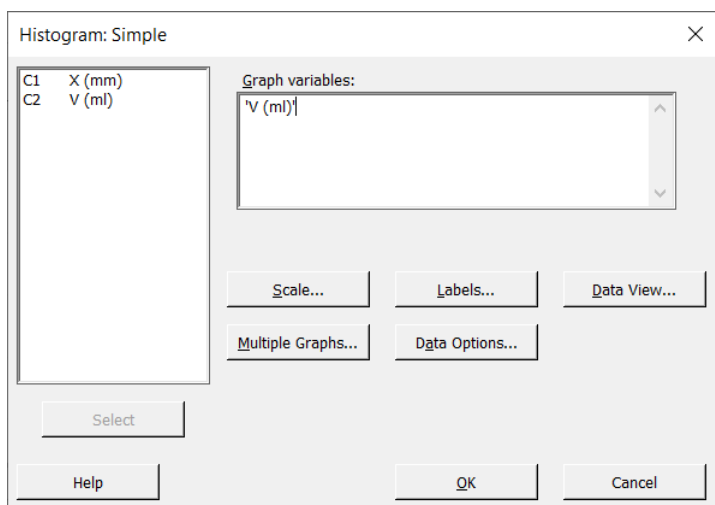
střed třídy s největší četností $\hat{x} = 1 \text{ ml}$; dalším výpočtem obdržíme $A \approx 0,098502$.

j	x_j^*	f_j	$f_j x_j^*$	$f_j x_j^{*2}$
1	-2	2	-4	8
2	-1	5	-5	5
3	0	11	0	0
4	1	13	13	13
5	2	9	18	36
6	3	6	18	54
7	4	4	16	64
Σ	—	50	56	180

Postup v Minitabu:

Minitab číselné charakteristiky roztříděného statistického souboru nepočítá, statistický soubor zpracovává neroztříděný, ale lze vytvořit histogramy.

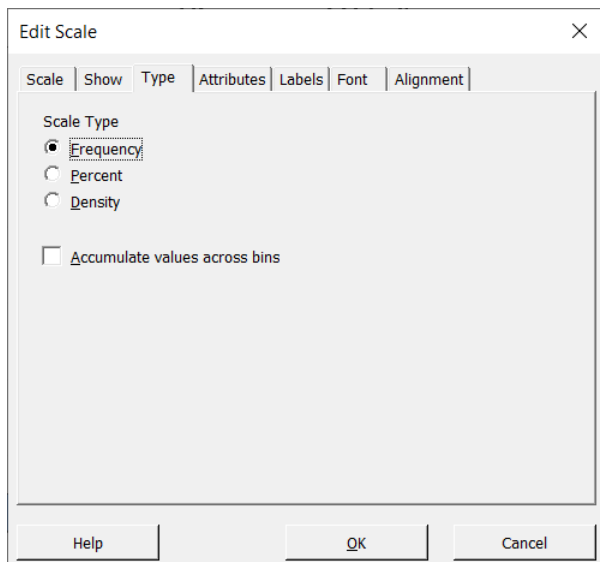
Graph > Histogram > choose Simple



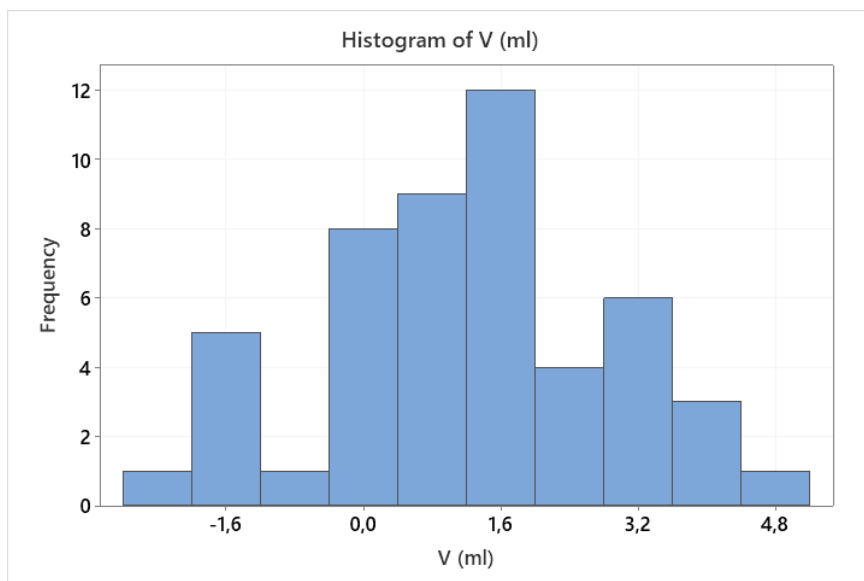


... > Scale > Y-Scale Type

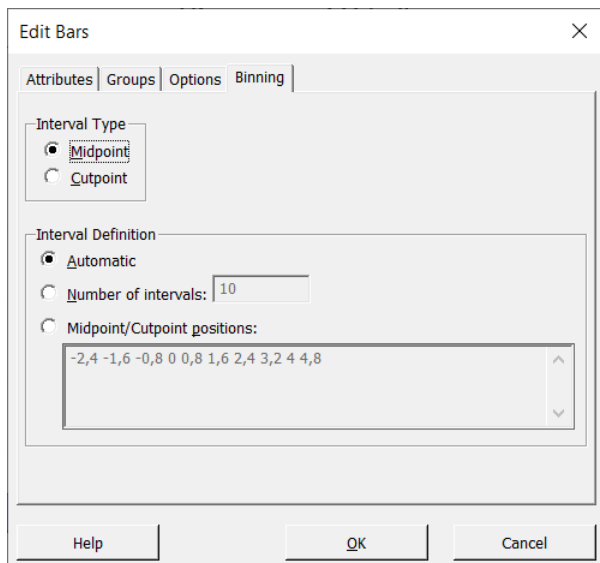
lze měnit o jaký typ histogramu jde.



Výstup:



Select bars > **Editor** > **Edit Bars** > **Binning** lze měnit parametry třídění.



Řešený příklad 1.3

Statistickým šetřením nákladů X (Kč) a cen Y (Kč) pro stejný výrobek u 10 výrobců byl získán dvourozměrný statistický soubor:

(30,18; 50,26), (30,19; 50,23), (30,21; 50,27), (30,22; 50,25), (30,25; 50,22),
(30,26; 50,32), (30,26; 50,33), (30,28; 50,29), (30,30; 50,37), (30,33; 50,42).

Vypočtete \bar{x} , \bar{y} , $s^2(x)$, $s^2(y)$, $s(x)$, $s(y)$, c , r .

Ř e š e n í:

Vzhledem k malému rozsahu $n = 10$ soubor netřídíme. Použitím výše uvedených vztahů dostaneme:

$$\bar{x} = (30,18 + \dots + 30,33)/10 = 30,248 \text{ Kč} \dots \text{průměrné náklady,}$$

$$\bar{y} = (50,26 + \dots + 50,42)/10 = 50,296 \text{ Kč} \dots \text{průměrná cena,}$$

$$s^2(x) = (0,18^2 + \dots + 30,33^2)/10 - 30,248^2 = 0,002096 \text{ Kč}^2,$$

$$s^2(y) = (50,26^2 + \dots + 50,42^2)/10 - 50,296^2 = 0,003684 \text{ Kč}^2,$$

$$s(x) = \sqrt{0,002096} \approx 0,0457821 \text{ Kč} \approx 0,0458 \text{ Kč},$$

$$s(y) = \sqrt{0,003684} \approx 0,0606960 \text{ Kč} \approx 0,0607 \text{ Kč},$$

$$cov = (30,18 \cdot 50,26 + \dots + 30,33 \cdot 50,42)/10 - 30,248 \cdot 50,296 = 0,002292 \text{ Kč}^2,$$

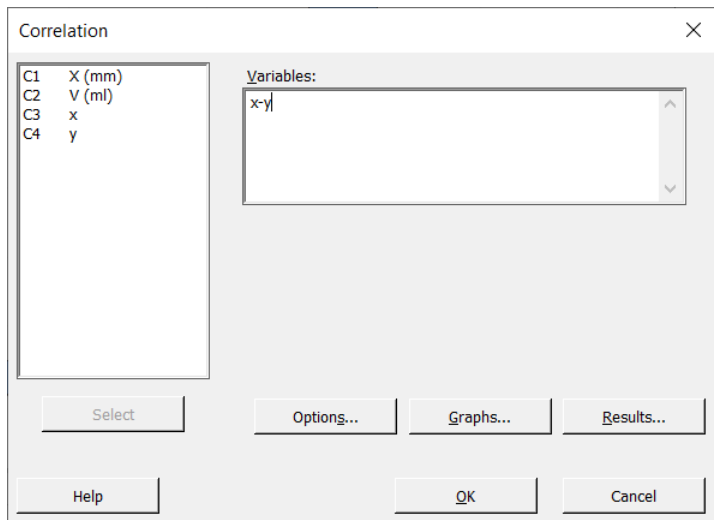
$$r = 0,002292 / (0,0457821 \cdot 0,0606960) = 0,82481996263 \approx 0,8248.$$

Vzhledem k velikosti koeficientu korelace r lze předpokládat, že mezi oběma znaky X a Y (náklady a cenou) je závislost víceméně blízká lineární. Jeho kladná hodnota odpovídá tomu, že s rostoucími náklady roste cena výrobku.

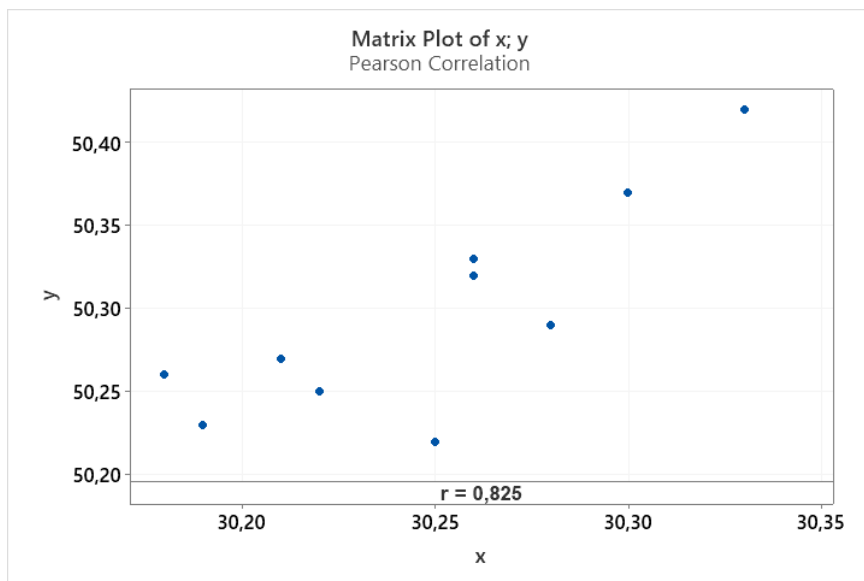


Postup v Minitabu:

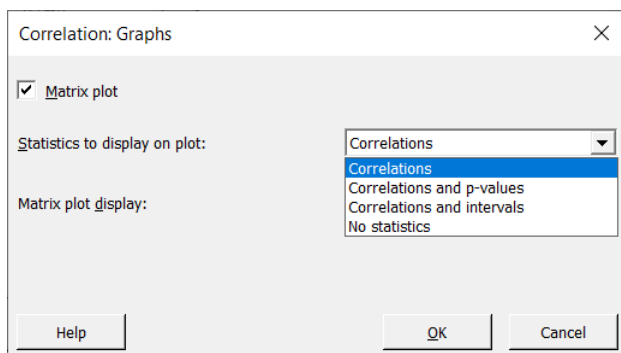
Stat > Basic Statistics > Correlation



Výstup:

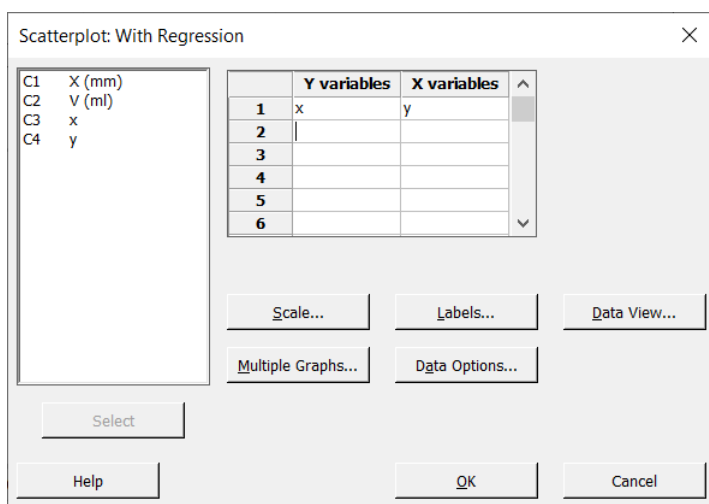


Kromě koeficientu korelace Minitab může vypočítat i intervalový odhad koeficientu korelace za předpokladu výběru z dvourozměrného normálního rozdělení, který bude popsán v sedmé kapitole, nebo p-hodnotu k testu nezávislosti, jehož význam je popsán v osmé kapitole. My jsme zvolili pouze výpočet koeficientu korelace, jak je vidět na následujícím obrázku.

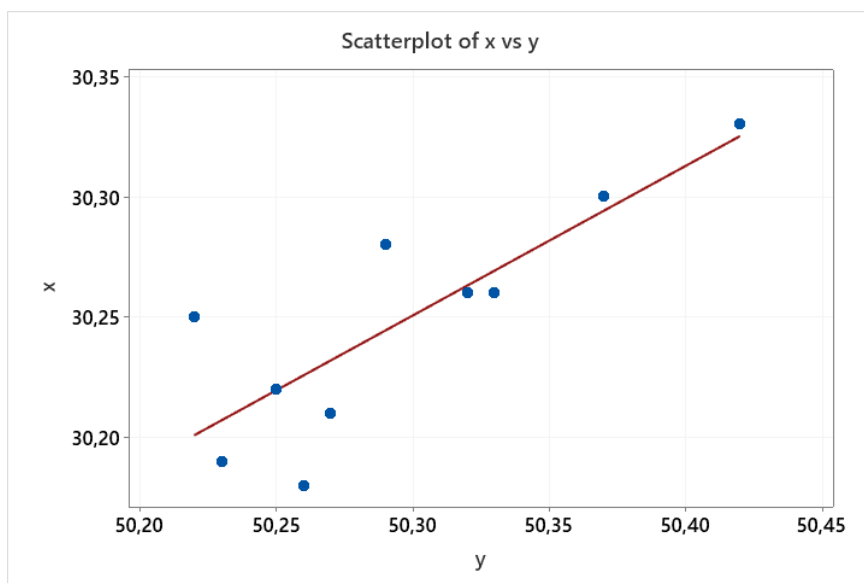


Grafické znázornění regresní přímky v Minitabu:

Graph > Scatterplot > > choose Simple or With Regression



Výstup:





5 Rozdělení pravděpodobnosti pro aplikace

Řešený příklad 5.1

V dodávce 50 výrobků je 5 zmetků. Z dodávky jsou náhodně vybrány 3 výrobky. Počet zmetků mezi vybranými výrobky je náhodná veličina X . Určete typ jejího rozdělení pravděpodobnosti, její pravděpodobnostní funkci $p(x)$, střední hodnotu $E(X)$, rozptyl $D(X)$, směrodatnou odchylku $\sigma(X)$, koeficient šikmosti $A(X)$, medián $x_{0,5}$, modus \hat{x} a $P(1 < X \leq 3)$. Předpokládejte, že každý vybraný výrobek se vrátí nazpět do dodávky, takže jde o náhodný výběr s vracením.

Ř e š e n í:

Náhodná veličina X má rozdělení $Bi(n, p)$, kde $n = 3$ a $p = 5/50 = 0,1$. Náhodná veličina X nabývá hodnot $x = 0, 1, 2, 3$ a její pravděpodobnostní funkce je

$$p(x) = \binom{3}{x} 0,1^x \cdot 0,9^{3-x} \text{ pro } x = 0, 1, 2, 3.$$

Střední hodnota je $E(X) = np = 3 \cdot 0,1 = 0,3$,

rozptyl je $D(X) = np(1 - p) = 3 \cdot 0,1 \cdot 0,9 = 0,27$,

směrodatná odchylka je $\sigma(X) = \sqrt{D(X)} = \sqrt{0,27} \approx 0,51962$,

koeficient šikmosti je $A(X) = \frac{1 - 2p}{\sqrt{np(1 - p)}} = \frac{1 - 2 \cdot 0,1}{\sqrt{0,27}} \approx 2,7245$,

medián $x_{0,5} = 0$, neboť $F(x) = p(0) = \binom{3}{0} 0,1^0 \cdot 0,9^{3-0} = 0,729 > 0,5$ pro $x(0;1)$,

modus $\hat{x} = 0$, neboť $(n + 1)p - 1 = -0,6$ a $(n + 1)p = 0,4$,

$P(1 < X \leq 3) = p(2) + p(3) = 0,027 + 0,001 = 0,028$.

Minitab nepočítá číselné charakteristiky rozdělení, ale může spočítat danou pravděpodobnostní funkci.

Calc > Probability Distributions > > choose Binomial

Ve sloupci x jsou uvedeny hodnoty 0 až 3, jak je patrné z tabulkového výstupu.



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

Binomial Distribution

☒ Probability
☐ Cumulative probability
☐ Inverse cumulative probability

Number of trials: 3
Event probability: 0,1

☒ Input column: x
Optional storage:
☐ Input constant:
Optional storage:

Select

Help OK Cancel

Tabulkový výstup:

Binomial with $n = 3$ and $p = 0,1$

x	$P(X = x)$
0	0,729
1	0,243
2	0,027
3	0,001

Grafické znázornění příkladu:

Graph > Probability Distribution Plot

Probability Distribution Plots

View Single Vary Parameters Two Distributions View Probability

Help OK Cancel

>> *choose* Binomial



Probability Distribution Plot: View Probability

Distribution Shaded Area

Distribution: Binomial

Number of trials: 3

Event probability: 0,1

Select

Help OK Cancel

Probability Distribution Plot: View Probability

Distribution Shaded Area

Define Shaded Area By

☐ Probability

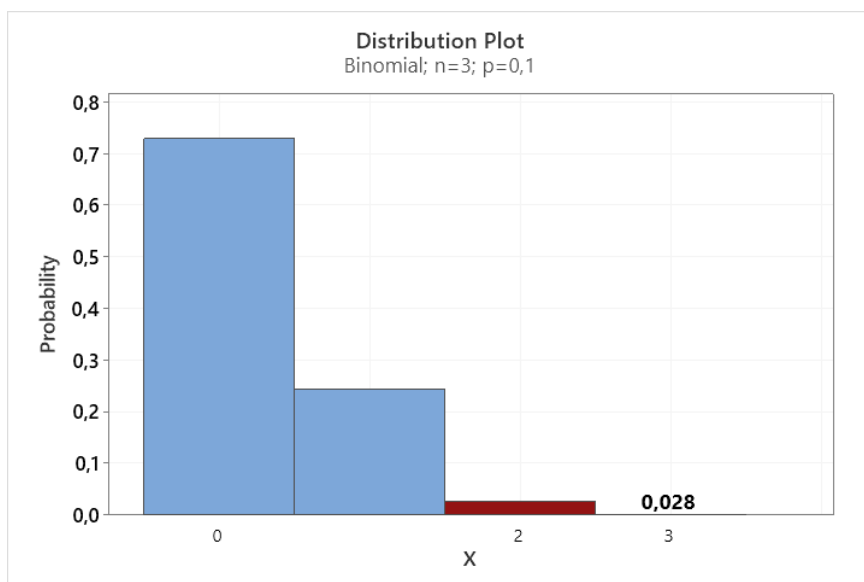
☒ X Value

Right Tail Left Tail Both Tails Middle

X value 1: 2 X value 2: 3

Help OK Cancel

Grafický výstup:



Řešený příklad 5.2

V dodávce 50 výrobků je 5 zmetků. Z dodávky jsou náhodně vybrány 3 výrobky. Počet zmetků mezi vybranými výrobky je náhodná veličina X . Určete typ jejího rozdělení pravděpodobnosti, její pravděpodobnostní funkci $p(x)$, střední hodnotu $E(X)$, rozptyl $D(X)$, směrodatnou odchylku $\sigma(X)$, medián $x_{0,5}$, modus \hat{x} a $P(1 < X \leq 3)$. Předpokládejte (na rozdíl od řešeného příkladu 5.1), že se vybraný výrobek nevrací nazpět do dodávky, takže jde o náhodný výběr bez vracení.

Ř e š e n í:

Náhodná veličina X má rozdělení $H(N, M, n)$, kde $N = 50$, $M = 5$ a $n = 3$. Náhodná veličina X nabývá hodnot $x = 0, 1, 2, 3$ a její pravděpodobnostní funkce je



$$p(x) = \frac{\binom{5}{x} \binom{45}{3-x}}{\binom{50}{3}} \text{ pro } x = 0, 1, 2, 3.$$

Střední hodnota je $E(X) = n \frac{M}{N} = 3 \cdot 0,1 = 0,3$,

rozptyl je $D(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1} = 3 \cdot 0,1 \cdot 0,9 \cdot (47/49) \approx 0,25898$,

směrodatná odchylka je $\sigma(X) = \sqrt{D(X)} \approx \sqrt{0,25898} \approx 0,50890$,

medián $x_{0,5} = 0$, neboť $F(x) = p(0) = \frac{\binom{5}{0} \binom{45}{3-0}}{\binom{50}{3}} \approx 0,72398 > 0,5$ pro $x(0;1)$,

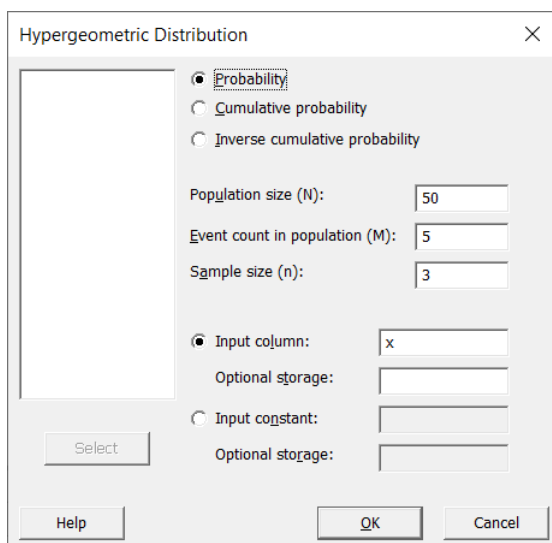
modus $\hat{x} = 0$, neboť $a = \frac{(M+1)(n+1)}{N+2} \approx 0,46154$, takže $a - 1 \approx -0,53846$,

$P(1 < X \leq 3) = p(2) + p(3) \approx 0,02296 + 0,00051 = 0,02347$.

Minitab nepočítá číselné charakteristiky rozdělení, ale může spočítat danou pravděpodobnostní funkci.

Calc > Probability Distributions > choose Hypergeometric

Ve sloupci x jsou uvedeny hodnoty 0 až 3, jak je patrné z tabulkového výstupu.





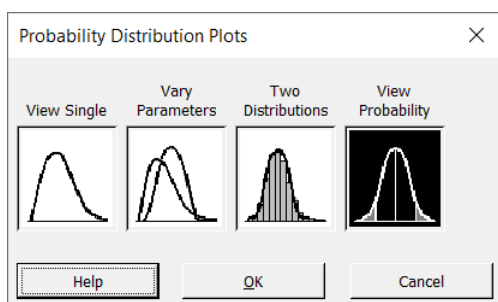
Tabulkový výstup:

Hypergeometric with $N = 50$, $M = 5$, and $n = 3$

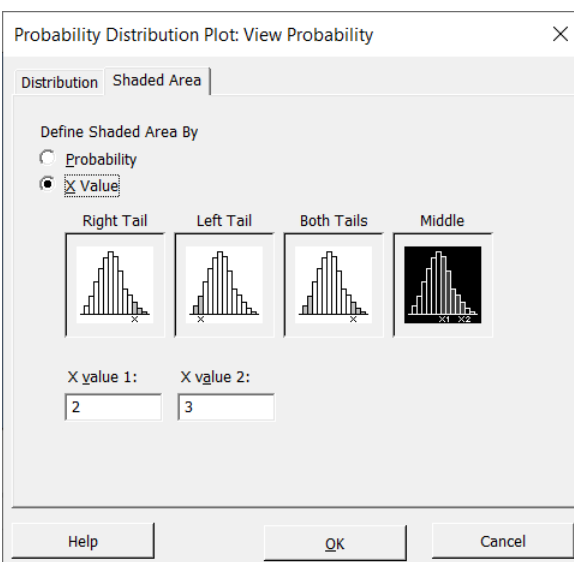
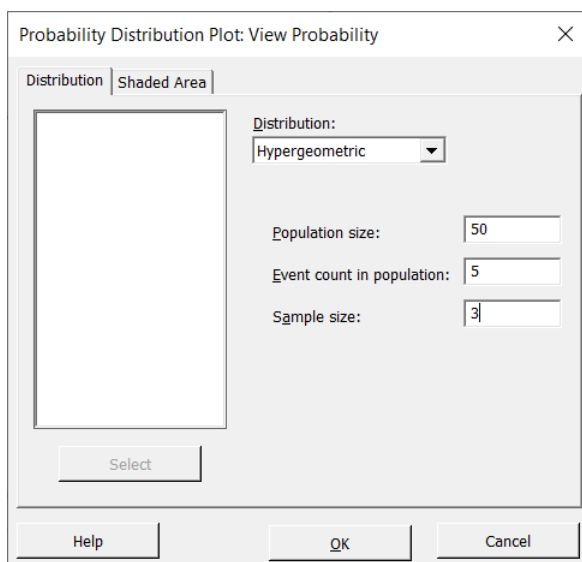
x	$P(X = x)$
0	0,723980
1	0,252551
2	0,022959
3	0,000510

Grafické znázornění příkladu:

Graph > Probability Distribution Plot

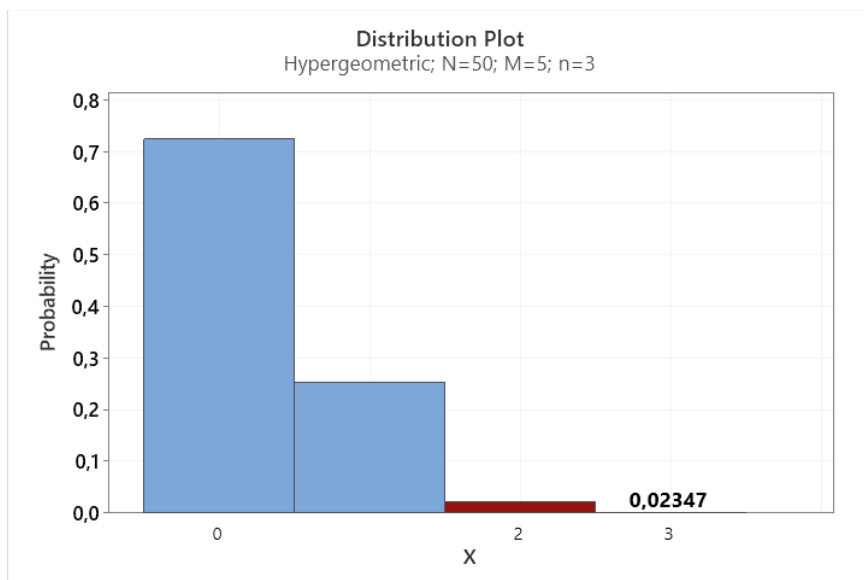


> > *choose Hypergeometric*





Grafický výstup:



Řešený příklad 5.3

Statistickým průzkumem bylo zjištěno, že během jedné minuty navštíví prodejnu průměrně 3 zákazníci. Najděte vhodný typ rozdělení pravděpodobnosti náhodné veličiny X vyjadřující počet zákazníků, kteří navštíví prodejnu během jedné minuty. Určete její pravděpodobnostní funkci $p(x)$, střední počet zákazníků $E(X)$, rozptyl $D(X)$ a směrodatnou odchylku $\sigma(X)$ počtu zákazníků, koeficient šikmosti $A(X)$ a nejpravděpodobnější počet zákazníků za jednu minutu. Určete dále pravděpodobnost, že během jedné minuty přijde a) právě 1 zákazník, b) aspoň 1 zákazník, c) medián $x_{0,5}$ počtu zákazníků.

Ř e š e n í:

Nahradíme střední počet zákazníků, kteří navštíví prodejnu během jedné minuty, jejich průměrným počtem, tj. položíme $E(X) = \bar{x}$. Vzhledem k tomu, že nemáme další informace o rozdělení pravděpodobnosti náhodné veličiny X (např. o rozptylu $D(X)$ a koeficientu šikmosti $A(X)$), použijeme Poissonovo rozdělení pravděpodobnosti $Po(\lambda)$ s pravděpodobnostní funkcí

$$p(x) = \frac{3^x}{x!} e^{-3}, \quad x = 0, 1, \dots$$

Střední hodnota je $E(X) = \lambda = 3$,

rozptyl je $D(X) = \lambda = 3$,

směrodatná odchylka je $\sigma(X) = \sqrt{D(X)} = \sqrt{3} \approx 1,73205$,

koeficient šikmosti je $A(X) = \sqrt{\lambda} = \sqrt{3} \approx 1,73205$,



pro modus (nejpravděpodobnější počet zákazníků) je $\lambda - 1 \leq \hat{x} \leq \lambda$, takže $\hat{x} = 2$ a 3.

a) $P(X = 1) = p(1) = \frac{3^1}{1!} e^{-3} \approx 0,14936$,

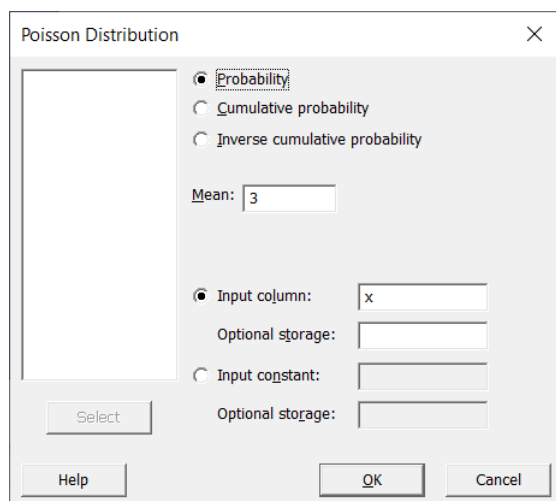
b) $P(X \geq 1) = p(1) + p(2) + \dots = 1 - p(0) = 1 - \frac{3^0}{0!} e^{-3} \approx 1 - 0,04979 = 0,95021$,

c) medián je $x_{0,5} = 3$, neboť $p(0) + p(1) + p(2) \approx 0,42319 < 0,5$ a $p(0) + p(1) + p(2) + p(3) \approx 0,64723 > 0,5$.

Minitab nepočítá číselné charakteristiky rozdělení, ale může spočítat danou pravděpodobnostní funkci.

Calc > Probability Distributions > choose Poisson

Ve sloupci x jsou uvedeny hodnoty 0 až 12, jak je patrné z tabulkového výstupu (Poissonovo rozdělení není nijak omezeno, tedy X nabývá nekonečně mnoha hodnot, ale jejich pravděpodobnost je velmi malá, proto je dále nepočítáme).



Tabulkový výstup:

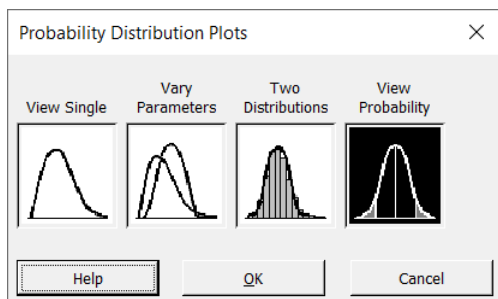
Poisson with mean = 3

x	P(X = x)
0	0,049787
1	0,149361
2	0,224042
3	0,224042
4	0,168031
5	0,100819
6	0,050409
7	0,021604
8	0,008102
9	0,002701
10	0,000810
11	0,000221
12	0,000055

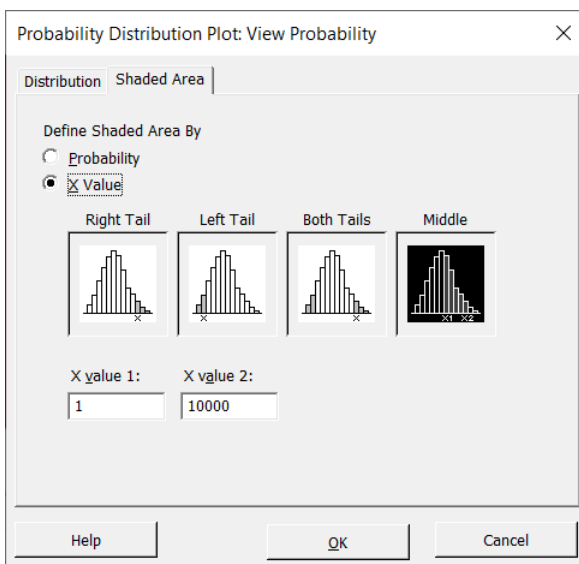
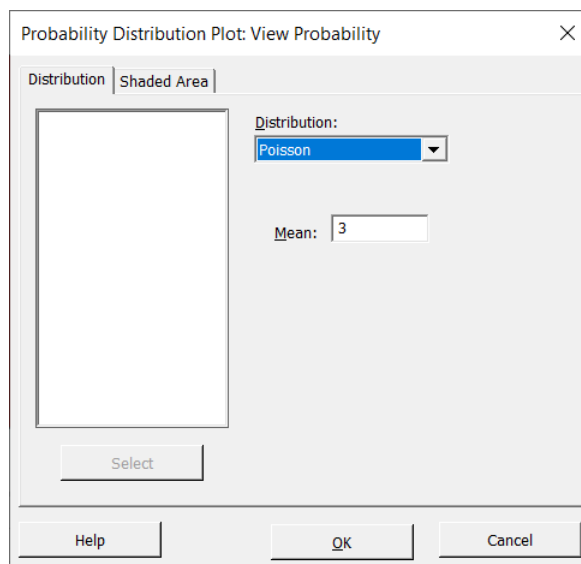


Grafické znázornění příkladu:

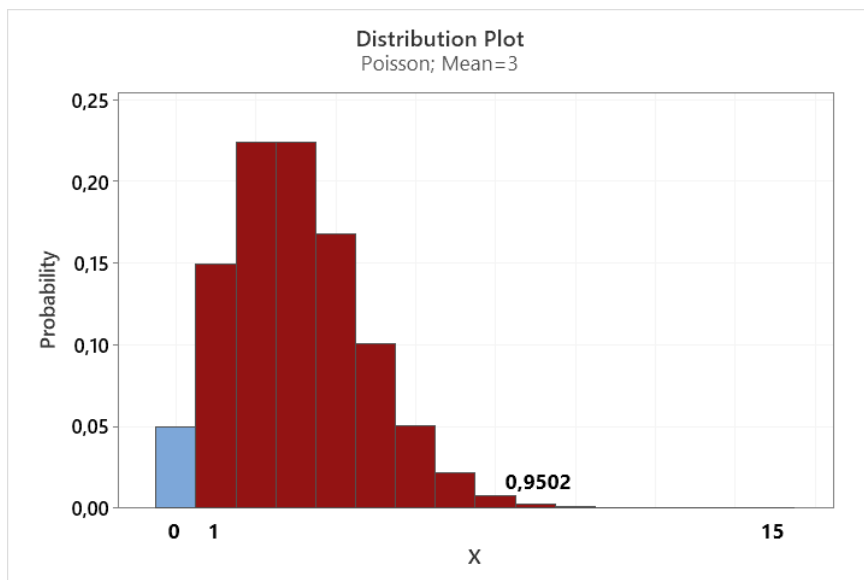
Graph > Probability Distribution Plot



> > *choose Poisson*



Grafický výstup:





Řešený příklad 5.4

K přerušení optického kabelu v délce 500 m může dojít v libovolné vzdálenosti od jeho počátku, přičemž pravděpodobnost náhodného jevu, že dojde k přerušení v nějakém úseku je přímo úměrná délce úseku a nezávisí na jeho poloze. Určete rozdělení pravděpodobnosti náhodné veličiny X vyjadřující vzdálenost místa přerušení od počátku, její hustotu pravděpodobnosti a základní číselné charakteristiky a pravděpodobnost, že k přerušení kabelu dojde v úseku od 300 m do 400 m.

Ř e š e n í:

Náhodná veličina X má rozdělení $R(a, b)$, kde $a = 0$ a $b = 500$ s hustotou pravděpodobnosti

$$f(x) = \begin{cases} \frac{1}{500} & \text{pro } x \in \langle 0; 500 \rangle, \\ 0 & \text{pro } x \notin \langle 0; 500 \rangle. \end{cases}$$

Střední hodnota a medián vzdálenosti je $E(X) = x_{0,5} = \frac{0 + 500}{2} = 250$ m,

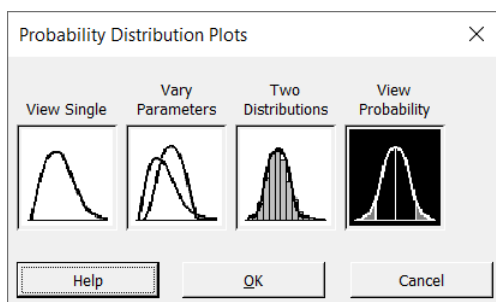
rozptyl je $D(X) = \frac{(500 - 0)^2}{12} \approx 20833,3 \text{ m}^2$,

směrodatná odchylka je $\sigma(X) = \sqrt{D(X)} \approx \sqrt{20833,3} \approx 144,34$ m,

pravděpodobnost $P(300 \leq X \leq 400) = F(400) - F(300) = \frac{400}{500} - \frac{300}{500} = 0,2$.

Protože výpočet pravděpodobnosti je elementární, uvádíme pouze grafický výstup:

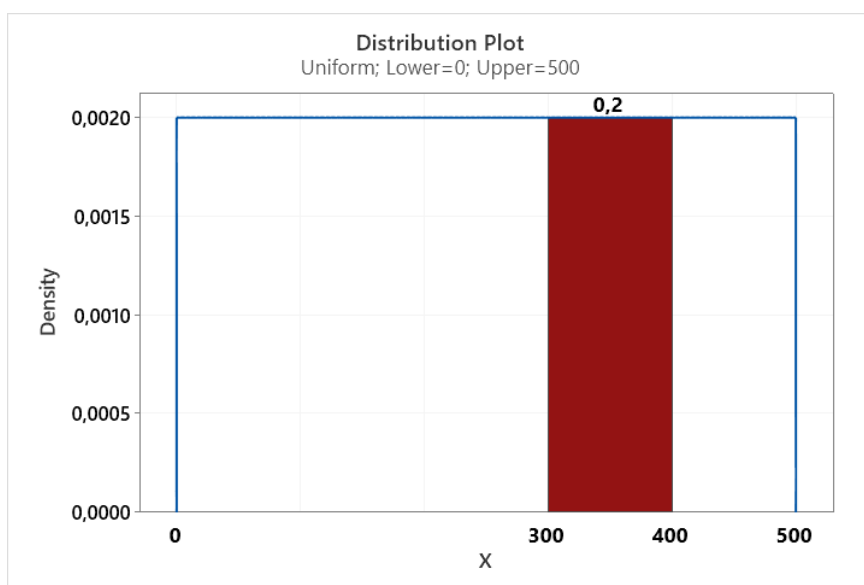
Graph > Probability Distribution Plot



>> *choose Uniform*



Grafický výstup:



Řešený příklad 5.5

Určete pravděpodobnost, že náhodná veličina X s normálním rozdělením pravděpodobnosti $N(20; 16)$, nabude hodnotu:

- menší než 16,
- větší než 20,
- v mezích od 12 do 28,
- menší než 12 nebo větší než 28.



Ř e š e n í:

Ze vztahu $F(x) = \Phi\left(\frac{x-20}{4}\right)$ a tabulky **T1** dostaneme:

- a) $P(X < 16) = F(16) = \Phi((16 - 20) / 4) = \Phi(-1) = 1 - \Phi(1) = 1 - 0,84135 = 0,15865$;
b) $P(X > 20) = 1 - P(X \leq 20) = 1 - F(20) = 1 - \Phi((20 - 20) / 4) = 1 - \Phi(0) = 1 - 0,5 = 0,5$;
c) $P(12 \leq X \leq 28) = F(28) - F(12) = \Phi((28 - 20) / 4) - \Phi((12 - 20) / 4) = \Phi(2) - \Phi(-2) =$
 $= \Phi(2) - (1 - \Phi(2)) = 2\Phi(2) - 1 = 2 \cdot 0,97725 - 1 = 0,95450$;
d) $P((X < 12) \vee (X > 28)) = 1 - P(12 \leq X \leq 28) = 1 - 0,9545 = 0,0455$.

Pro úspěšné vyřešení příkladu jsou potřeba hodnoty distribučních funkcí $F(12)$, $F(16)$, $F(20)$ a $F(28)$. Tyto argumenty napíšeme do Minitabu do sloupce x vypočítáme hodnoty distribučních funkcí a postupujeme jako při ručním výpočtu.

Calc > Probability Distributions > > choose Normal

Normal Distribution

C1 X (mm)
C2 V (ml)
C3 xx
C4 yy
C6 x

☐ Probability density
☒ Cumulative probability
☐ Inverse cumulative probability

Mean: 20
Standard deviation: 4

☒ Input column: x
Optional storage: |
☐ Input constant:
Optional storage:

Select

Help OK Cancel

Tabulkový výstup:

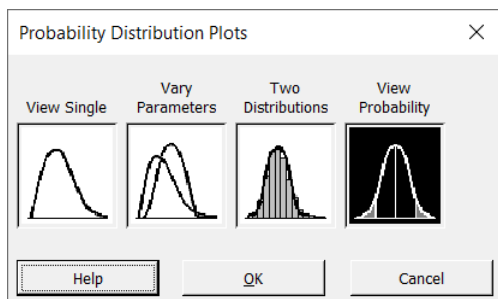
Normal with mean = 20 and standard deviation = 4

x	P(X ≤ x)
12	0,022750
16	0,158655
20	0,500000
28	0,977250

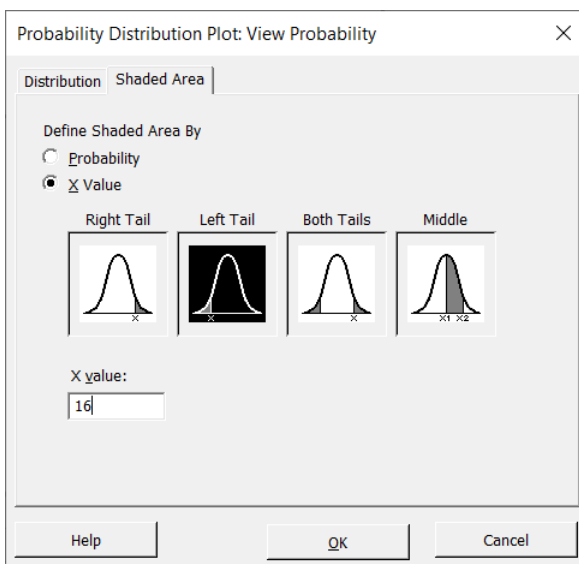
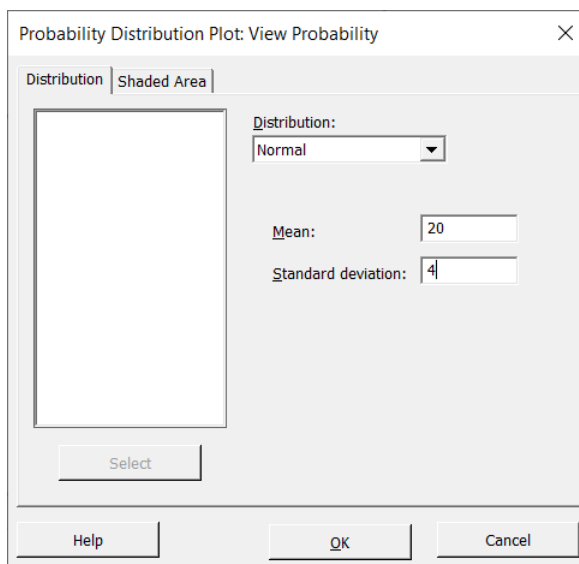


Grafický výstup například pro úlohu a):

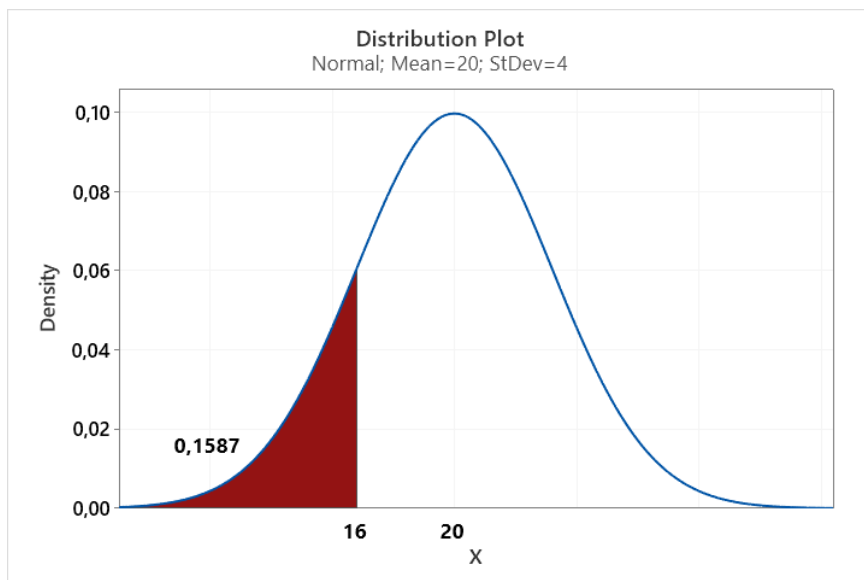
Graph > Probability Distribution Plot



> > *choose* Normal



Grafický výstup:





Řešený příklad 5.6

Měření délkového rozměru je zatíženo systematickou chybou 0,5 mm a náhodnou chybou s normálním rozdělením pravděpodobnosti s rozptylem 0,09 mm². Určete, pro jakou hodnotu δ bude celková chyba jednoho měření v mezích 0,5 – δ až 0,5 + δ s pravděpodobností 0,95.

Ř e š e n í:

Chyba jednoho měření X má normální rozdělení s parametry $\mu = 0,5$ a $\sigma^2 = 0,09$, neboť u náhodné chyby předpokládáme, že má nulovou střední hodnotu, takže

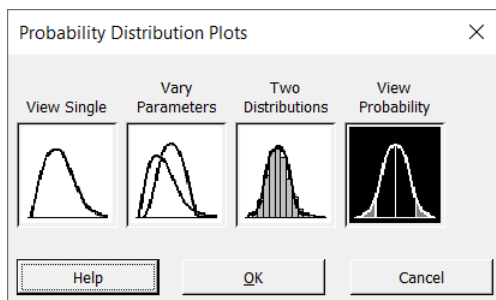
$$\begin{aligned} P(0,5 - \delta \leq X \leq 0,5 + \delta) &= F(0,5 + \delta) - F(0,5 - \delta) = \\ &= \Phi\left(\frac{\delta}{0,3}\right) - \Phi\left(-\frac{\delta}{0,3}\right) = 2\Phi\left(\frac{\delta}{0,3}\right) - 1 = 0,95. \end{aligned}$$

Odtud je $\Phi\left(\frac{\delta}{0,3}\right) = 0,975$, takže $\frac{\delta}{0,3} = u_{0,975}$. Protože z tabulky **T1** je $u_{0,975} = 1,960$, je

$\delta = 0,3 \cdot 1,960 = 0,588$. S pravděpodobností 0,95 bude celková chyba jednoho měření v intervalu $\langle -0,088; 1,088 \rangle$ mm.

Při řešení tohoto příkladu v Minitabu je vhodné využít:

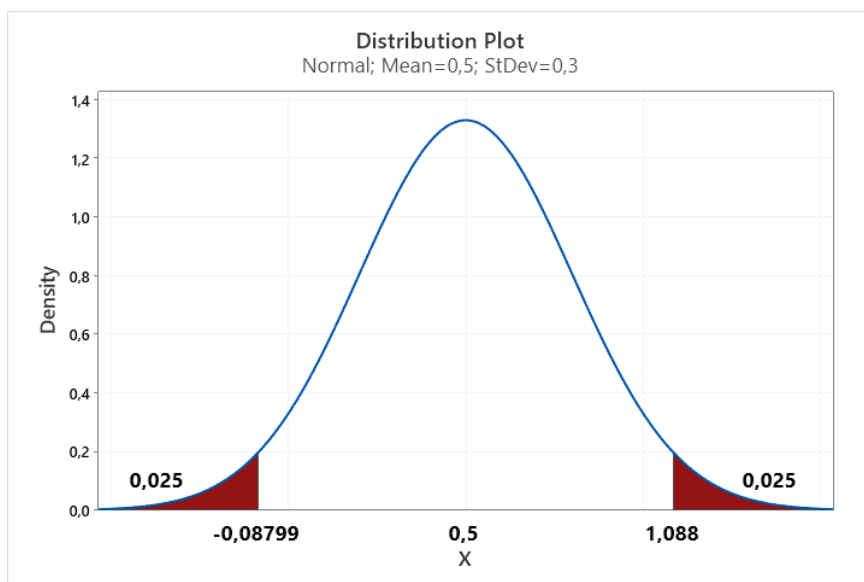
Graph > Probability Distribution Plot



>> **choose Normal**



Grafický výstup:



7 Odhady parametrů

Odhady parametrů normálního rozdělení

Řešený příklad 7.2

Měřením délky 10 válečků byl získán statistický soubor s empirickými charakteristikami $\bar{x} = 5,37$ mm, $s^2 = 0,0019$ mm² a $s = 0,044$ mm (viz řešený příklad 1.1). Určete bodové odhady střední hodnoty, rozptylu a směrodatné odchylky. Za předpokladu, že naměřená délka X má normální rozdělení pravděpodobnosti, určete intervalové odhady těchto číselných charakteristik se spolehlivostí 0,95.



Ř e š e n í:

Bodové odhady jsou:

střední délka válečku $\mu = 5,37$ mm,

rozptyl délky válečku $\sigma^2 = \frac{10}{9} 0,0019 = 0,00211$ mm²,

směrodatná odchylka délky válečku $\sigma = \sqrt{0,00211} \approx 0,046$ mm.

Intervalový odhad střední délky válečku μ se spolehlivostí 0,95 je, neboť $t_{0,975} = 2,262$ pro 9 stupňů volnosti z tabulky **T2**,

$$\mu \in <5,37 - 2,262 \frac{\sqrt{0,0019}}{\sqrt{10-1}}; 5,37 + 2,262 \frac{\sqrt{0,0019}}{\sqrt{10-1}}> \approx <5,337; 5,403> \text{ mm.}$$

Intervalový odhad rozptylu délky válečku σ^2 se spolehlivostí 0,95 je, neboť $\chi^2_{0,025} = 2,700$ a $\chi^2_{0,975} = 19,023$ pro 9 stupňů volnosti z tabulky **T3**,

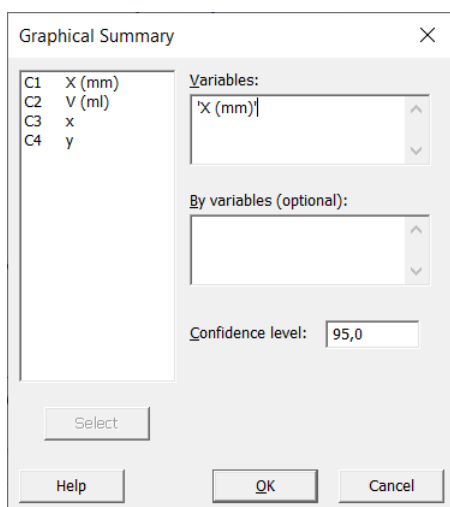
$$\sigma^2 \in <\frac{10 \cdot 0,0019}{19,023}; \frac{10 \cdot 0,0019}{2,700}> \approx <0,00100; 0,00704> \text{ mm}^2,$$

takže intervalový odhad směrodatné odchylky délky válečku σ je

$$\sigma \in <\sqrt{0,00100}; \sqrt{0,00704}> \approx <0,0316; 0,0839> \text{ mm.}$$

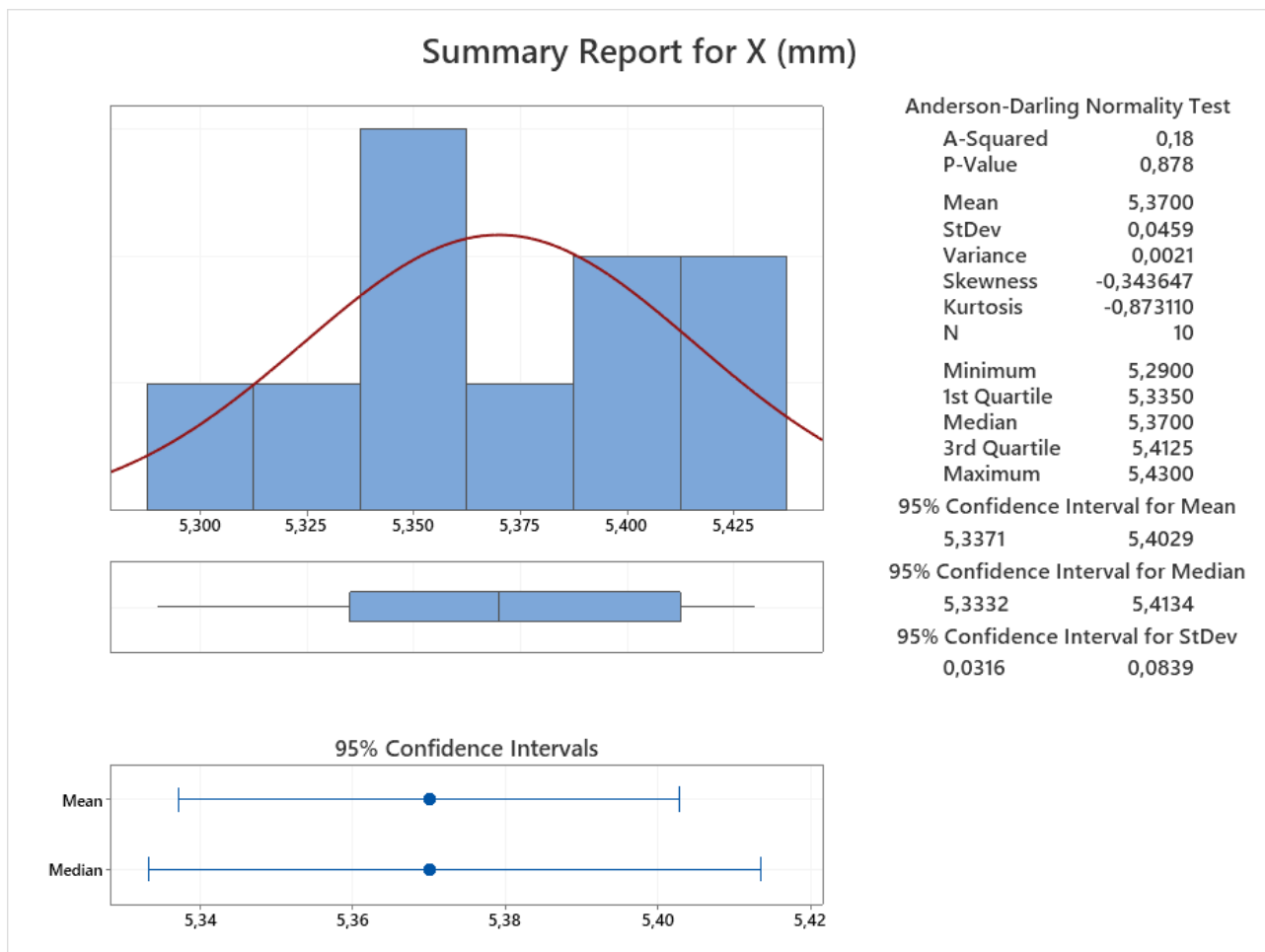
Postup v Minitabu:

Stat > Basic Statistics > Graphical Summary





Grafický výstup:



Alternativní postup:

Stat > Basic Statistics > 1-Sample t

One-Sample t for the Mean

C1 X (mm)
C2 V (ml)
C3 x
C4 y

One or more samples, each in a column

'X (mm)'

☐ Perform hypothesis test
Hypothesized mean:

Select Options... Graphs...

Help OK Cancel



Tabulkový výstup:

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
10	5,3700	0,0459	0,0145	(5,3371; 5,4029)

μ : population mean of X (mm)

Stat > Basic Statistics > 1 Variance

Tabulkový výstup:

Method

σ : standard deviation of X (mm)

The Bonett method is valid for any continuous distribution.

The chi-square method is valid only for the normal distribution.

Descriptive Statistics

		95% CI for σ using		95% CI for σ using
N	StDev	Variance	Bonett	Chi-Square
10	0,0459	0,00211	(0,0318; 0,0826)	(0,0316; 0,0839)

Pozor, pokud zadáváte číselné charakteristiky (Summarized data) je třeba jako směrodatnou odchylku (Standard deviation) směrodatnou odchylku σ !



Řešený příklad 7.3

Sledováním nákladů a ceny stejného výrobku u 10 výrobců byl získán dvourozměrný statistický soubor s koeficientem korelace $r = 0,82482$ (viz řešený příklad 1.3). Určete bodový odhad a intervalový odhad se spolehlivostí 0,99 koeficientu korelace ρ základního souboru.

Ř e š e n í:

Bodový odhad koeficientu korelace nákladů a ceny je $\rho = 0,82482$.

Po dosazení je $w = \frac{1}{2} \left(\ln \frac{1 + 0,82482}{1 - 0,82482} + \frac{0,82482}{10 - 1} \right) \approx 1,21753$.

Z tabulky **T1** je $u_{0,995} = 2,576$, takže

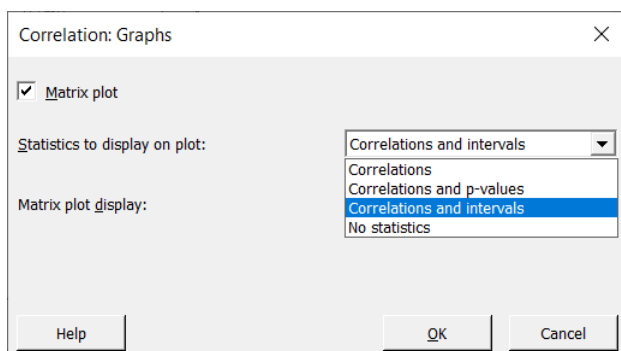
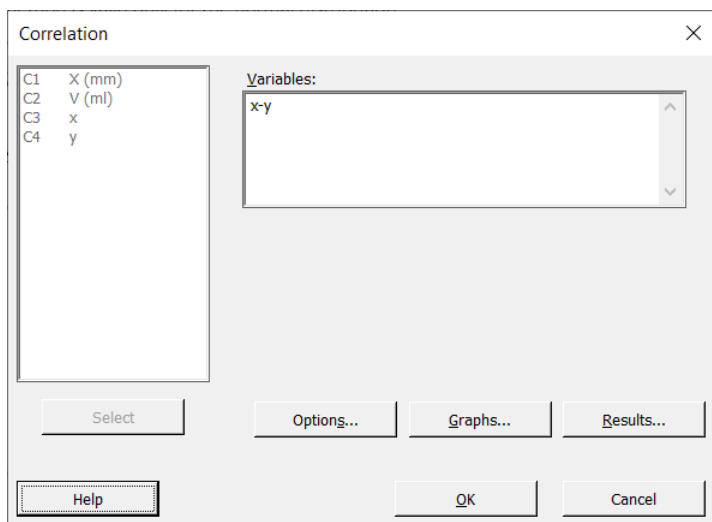
$$z_1 = 1,21753 - \frac{2,576}{\sqrt{10 - 3}} \approx 0,24397, \quad z_2 = 1,21753 + \frac{2,576}{\sqrt{10 - 3}} \approx 2,19110$$

a intervalový odhad koeficientu korelace nákladů a ceny ρ se spolehlivostí 0,99 je

$$\rho \in \langle \tanh 0,24397; \tanh 2,19110 \rangle \approx \langle 0,239242; 0,975313 \rangle.$$

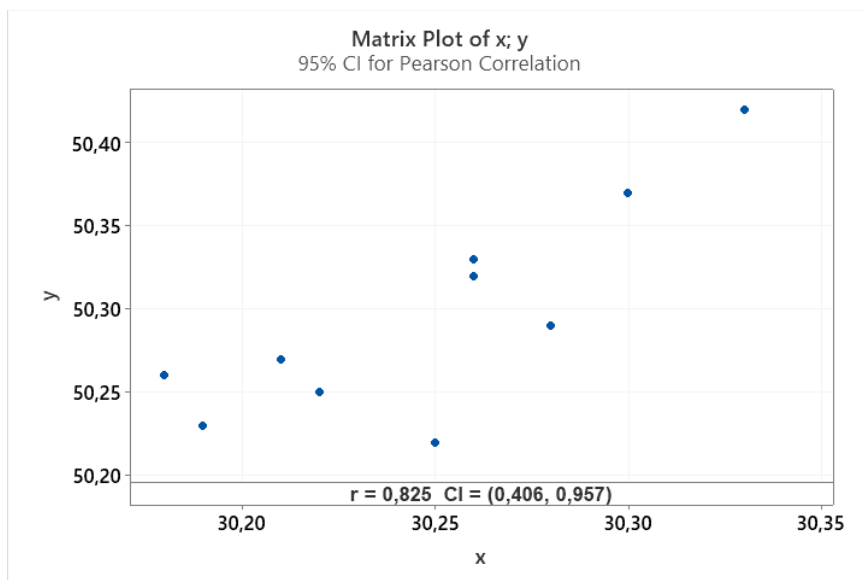
Postup v Minitabu:

Stat > Basic Statistics > Correlation





Grafický výstup:



Odhady parametru binomického rozdělení

Řešený příklad 7.4

Při průzkumu zájmu o nový výrobek odpovědělo ze 400 dotázaných zákazníků supermarketu STAMET kladně na otázku, zda si nový výrobek koupí, 80 zákazníků. Určete bodový a intervalový odhad podílu zákazníků p ze základního souboru všech zákazníků supermarketu STAMET.

Ř e š e n í:

Protože $x = 80$ a $n = 400$, je bodový odhad $p = \frac{80}{400} = 0,2$, tedy 20 % všech zákazníků supermarketu STAMET si chce koupit nový výrobek.

Z tabulky T1 pro spolehlivost 0,95 je $u_{0,975} = 1,960$, takže intervalový odhad podílu zákazníků p se spolehlivostí 0,95 je

$$p \in \left\langle \frac{80}{400} - 1,96 \sqrt{\frac{\frac{80}{400} \left(1 - \frac{80}{400}\right)}{400}}; \frac{80}{400} + 1,96 \sqrt{\frac{\frac{80}{400} \left(1 - \frac{80}{400}\right)}{400}} \right\rangle = \langle 0,1608; 0,2392 \rangle.$$

Pro spolehlivost 0,99 obdržíme analogickým způsobem intervalový odhad

$$p \in \langle 0,1485; 0,2515 \rangle.$$

Se spolehlivostí 0,95, resp. 0,99, si nový výrobek koupí přibližně 16 až 24 %, resp. 15 až 25 %, všech zákazníků supermarketu STAMET. Pokud má STAMET celkem 10 000 zákazníků, lze víceméně



očekávat, že prodá cca 2 000 nových výrobků. Z intervalového odhadu můžeme pak se spolehlivostí 0,95 usuzovat, že STAMET prodá přibližně $10\,000 \cdot 0,16 = 1\,600$ až $10\,000 \cdot 0,24 = 2\,400$ nových výrobků.

Postup v Minitabu:

Stat > Basic Statistics > 1 Proportion

The image shows two overlapping Minitab dialog boxes. The main box is 'One-Sample Proportion' with 'Summarized data' selected. It contains input fields for 'Number of events' (80) and 'Number of trials' (400). There is an unchecked checkbox for 'Perform hypothesis test' and an empty field for 'Hypothesized proportion'. At the bottom are buttons for 'Select', 'Options...', 'Help', 'OK', and 'Cancel'. The 'Options...' box is open, showing 'Confidence level' as 95,0, 'Alternative hypothesis' as 'Proportion ≠ hypothesized proportion', and 'Method' as 'Exact'. The 'Exact' method is highlighted in the list. It also has 'Help', 'OK', and 'Cancel' buttons.

Pokud chceme dostat stejné výsledky jako při ručním výpočtu zaškrtneme aproximaci normálním rozdělením (Normal approximation), ta ovšem není tak přesná, jako standardní výpočet přes binomické rozdělení.

Tabulkový výstup:

Method

p: event proportion

Exact method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% CI for p
400	80	0,200000	(0,161895; 0,242610)

Descriptive Statistics

N	Event	Sample p	99% CI for p
400	80	0,200000	(0,151087; 0,256239)



8 Testování statistických hypotéz

Při testování statistických hypotéz na PC pomocí statistického software se místo kritického oboru \bar{W}_α obvykle používá následující tzv. *P-hodnota*. Jestliže např. testujeme hypotézu $H: \mu = \mu_0$ proti dvoustranné alternativní hypotéze $\bar{H}: \mu \neq \mu_0$, pak pro pozorovanou hodnotu t testového kritéria T je *P-hodnota* je číslo $1 - P(-t \leq T \leq t)$. Výše uvedené konvenci rozhodnutí o daných hypotézách pomocí kritického oboru, resp. oboru nezamítnutí, odpovídá následující adekvátní postup. Jestliže $P < \alpha$, pak *zamítáme* hypotézu H a současně *nezamítáme* hypotézu \bar{H} na hladině významnosti α . Jestliže naopak $P \geq \alpha$, pak *nezamítáme* hypotézu H a současně *zamítáme* hypotézu \bar{H} na hladině významnosti α .

Testy hypotéz o parametrech normálního rozdělení

Řešený příklad 8.2

Měřením délky 10 válečků byly získány empirické charakteristiky $\bar{x} = 5,37$ mm a $s^2 = 0,0019$ mm² (viz řešený příklad 1.1). Na hladině významnosti 0,05 testujeme hypotézu, že střední naměřená délka válečku je 5,40 mm, tedy $H: \mu = 5,40$.

Ř e š e n í:

Pozorovaná hodnota testového kritéria je

$$t = \frac{5,37 - 5,40}{\sqrt{0,0019}} \sqrt{10 - 1} = -2,0647.$$

Pro $10 - 1 = 9$ stupňů volnosti je $t_{0,975} = 2,262$ z tabulky **T2**, takže $\bar{W}_{0,05} = \langle -2,262; 2,262 \rangle$. Protože $t \in \bar{W}_{0,05}$, hypotézu *nezamítáme*. Pro testování této hypotézy bylo možno použít také intervalový odhad se spolehlivostí 0,95 z příkladu 2.2. Protože tento odhad obsahuje hypotetickou hodnotu 5,40, *nezamítáme* danou hypotézu na hladině významnosti $1 - 0,95 = 0,05$.

Postup v Minitabu:

Stat > Basic Statistics > 1-Sample t



Výstup z Minitabu:

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
10	5,3700	0,0459	0,0145	(5,3371; 5,4029)

μ : population mean of X (mm)

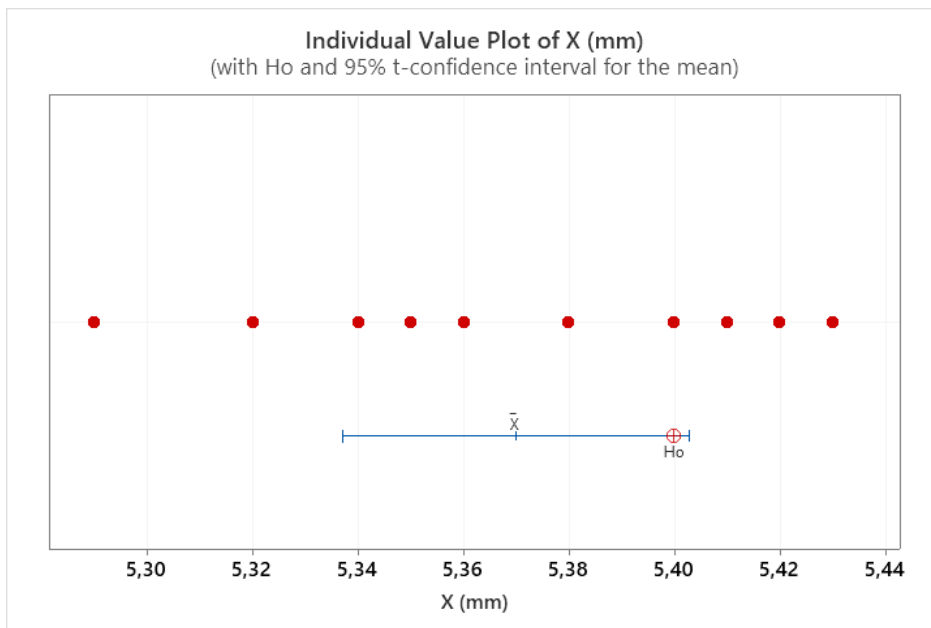
Test

Null hypothesis $H_0: \mu = 5,4$

Alternative hypothesis $H_1: \mu \neq 5,4$

T-Value	P-Value
-2,06	0,069

Protože $p = 0,069 > 0,05$, hypotézu $H: \mu = 5,40$ nezamítáme na hladině významnosti 0,05.



Řešený příklad 8.3

Na hladině významnosti 0,05 testujte hypotézu, že rozptyl naměřené délky válečku z příkladu 8.2 je $0,0025 \text{ mm}^2$, tedy $H: \sigma^2 = 0,0025$.

Ř e š e n í:

Pozorovaná hodnota testového kritéria je

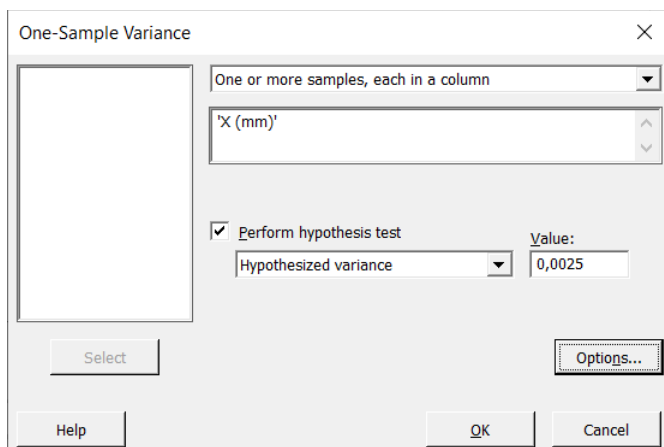
$$t = \frac{10 \cdot 0,0019}{0,0025} = 7,6.$$

Pro $10 - 1 = 9$ stupňů volnosti je $\chi^2_{0,025} = 2,700$ a $\chi^2_{0,975} = 19,023$ z tabulky **T3**, takže

$\bar{W}_{0,05} = \langle 2,700; 19,023 \rangle$. Protože $t \in \bar{W}_{0,05}$, hypotézu nezamítáme.

Postup v Minitabu:

Stat > Basic Statistics > 1 Variance





Výstup z Minitabu:

Test

Null hypothesis $H_0: \sigma^2 = 0,0025$

Alternative hypothesis $H_1: \sigma^2 \neq 0,0025$

Test			
Method	Statistic	DF	P-Value
Bonett	—	—	0,739
Chi-Square	7,60	9	0,850

Protože $p = 0,850 > 0,05$, hypotézu $H: \sigma^2 = 0,0025$ nezamítáme na hladině významnosti 0,05.

Řešený příklad 8.4

Sledováním nákladů X a ceny Y stejného výrobku u deseti výrobců byl získán dvourozměrný statistický soubor s koeficientem korelace $r = 0,82482$ (viz řešený příklad 1.3). Na hladině významnosti 0,01 testujte hypotézu, že veličiny X a Y jsou nekorelované (vzhledem k normálnímu rozdělení nezávislé), tedy $H: \rho = 0$.

Ř e š e n í:

Pozorovaná hodnota testového kritéria je

$$t = \left(\ln \frac{1 + 0,82482}{1 - 0,82482} - \ln \frac{1 + 0}{1 - 0} - \frac{0}{10 - 1} \right) \frac{\sqrt{10 - 3}}{2} \approx 3,1001.$$

Pro danou hladinu významnosti je $u_{0,995} = 2,576$ z tabulky **T1**, takže $\bar{W}_{0,01} = \langle -2,576; 2,576 \rangle$. Protože $t \notin \bar{W}_{0,01}$, hypotézu zamítáme a považujeme X, Y za závislé.

Postup v Minitabu je zřejmý: Stat > Basic Statistics > Correlation

Řešený příklad 8.5

Měřením teploty dvěma přístroji byly během osmi dnů získány dvojice $(x_i, y_i) = (51,8; 49,5), (54,9; 53,3), (52,2; 50,6), (53,3; 52,0), (51,6; 46,8), (54,1; 50,5), (54,2; 52,1), (53,3; 53,0)$ ($^{\circ}\text{C}$). Na hladině významnosti 1% testujte hypotézu, že rozdíl středních hodnot je nevýznamný, tedy $H: \mu(X) = \mu(Y)$.

Ř e š e n í:

Pro $d_i = x_i - y_i, i = 1, \dots, 8$, dostaneme $\bar{d} = 2,2^{\circ}\text{C}$ a $s(d) = 1,3172^{\circ}\text{C}$. Pozorovaná hodnota testového kritéria je

$$t = \frac{2,2}{1,3172} \sqrt{8 - 1} \approx 4,4190.$$



Pro $8 - 1 = 7$ stupňů volnosti je $t_{0,995} = 3,499$ z tabulky **T2**, takže $\bar{W}_{0,01} = \langle -3,499; 3,499 \rangle$. Protože $t \notin \bar{W}_{0,01}$, hypotézu zamítáme na hladině významnosti 1 % a považujeme rozdíl naměřených hodnot za statisticky významný.

Postup v Minitabu je zřejmý: Stat > Basic Statistics > Paired t

Řešený příklad 8.6

Zkouškami pevnosti drátů vyrobených dvěma různými technologiemi byly získány dva statistické soubory s charakteristikami $n_1 = 33$, $\bar{x} = 5,4637$ kN, $s^2(x) = 0,3302$ kN², $n_2 = 28$, $\bar{y} = 6,1179$ kN, $s^2(y) = 0,4522$ kN². Na hladině významnosti 0,05 testujte hypotézu, že rozdílné technologie nemají vliv na střední pevnost drátu (za předpokladu stejných rozptylů

$\sigma^2(X)$ a $\sigma^2(Y)$), tedy $H: \mu(X) - \mu(Y) = 0$.

Ř e š e n í:

Pozorovaná hodnota testového kritéria je

$$t = \frac{5,4637 - 6,1179 - 0}{\sqrt{33 \cdot 0,3302 + 28 \cdot 0,4522}} \sqrt{\frac{33 \cdot 28 (33 + 28 - 2)}{33 + 28}} \approx 4,030.$$

Pro $33 + 28 - 2 = 59$ stupňů volnosti je $t_{0,975} = 2,001$ interpolací z tabulky **T2**, takže $\bar{W}_{0,05} = \langle -2,001; 2,001 \rangle$. Protože $t \notin \bar{W}_{0,05}$, hypotézu zamítáme. Rozdílné technologie mají vliv na střední pevnost drátu.

Postup v Minitabu je zřejmý: Stat > Basic Statistics > 2-Sample t (zaškrtneme Assume equal variance)

Řešený příklad 8.7

Při vyšetřování životnosti výrobků v různých systémech extrémních provozních podmínek byly získány dva statistické soubory s charakteristikami $n_1 = 21$, $\bar{x} = 3,581$, $s^2(x) = 0,114$, $n_2 = 23$, $\bar{y} = 3,974$, $s^2(y) = 0,041$ (životnost výrobků je v hodinách). Za předpokladu různých rozptylů $\sigma^2(X)$ a $\sigma^2(Y)$ testujte na hladině významnosti 0,05, že první systém extrémních provozních podmínek zvyšuje oproti druhému systému extrémních provozních podmínek střední životnost výrobku o 0,5 hod., tedy hypotézu $H: \mu(X) - \mu(Y) = -0,5$.

Ř e š e n í:

Pozorovaná hodnota testového kritéria je



$$t = \frac{3,581 - 3,974 - (-0,5)}{\sqrt{\frac{0,114}{21-1} + \frac{0,041}{23-1}}} \approx 1,2303.$$

Z tabulky **T2** pro $1 - \alpha/2 = 0,975$ je $t(x) = 2,086$ pro $21 - 1 = 20$ stupňů volnosti a $t(y) = 2,074$ pro $23 - 1 = 22$ stupňů volnosti, takže

$$\bar{t}_{0,975} = \frac{\frac{0,114}{21-1} 2,086 + \frac{0,041}{23-1} 2,074}{\frac{0,114}{21-1} + \frac{0,041}{23-1}} \approx 2,083.$$

a $\bar{W}_{0,05} = \langle -2,083; 2,083 \rangle$. Protože $t \in \bar{W}_{0,05}$, hypotézu o zvýšení střední životnosti o 0,5 hod. nezamítáme.

Postup v Minitabu je zřejmý: Stat > Basic Statistics > 2-Sample t (nezaškrtneme Assume equal variance)

Řešený příklad 8.9

Na hladině významnosti 0,05 ověřte předpoklad o různých rozptylech v řešeném příkladu 8.7, tedy že $\sigma^2(X) \neq \sigma^2(Y)$, kde $s^2(x) = 0,114$, $n_1 = 21$, $s^2(y) = 0,041$, $n_2 = 23$.

Ř e š e n í:

Testujeme naopak hypotézu $H: \sigma^2(X) = \sigma^2(Y)$. Pozorovaná hodnota testového kritéria je

$$t = \frac{\max\left(\frac{21 \cdot 0,114}{21-1}; \frac{23 \cdot 0,041}{23-1}\right)}{\min\left(\frac{21 \cdot 0,114}{21-1}; \frac{23 \cdot 0,041}{23-1}\right)} \approx \frac{\max(0,11970; 0,04286)}{\min(0,11970; 0,04286)} \approx 2,7928.$$

Z tabulky **T4** je pro $k_1 = 21 - 1 = 20$ a $k_2 = 23 - 1 = 22$ stupňů volnosti $F_{0,975} = 2,389$, takže $\bar{W}_{0,05} = \langle 1; 2,389 \rangle$. Protože $t \notin \bar{W}_{0,05}$, hypotézu zamítáme a předpoklad o různých rozptylech v příkladu 8.7 považujeme za správný.

Postup v Minitabu je zřejmý: Stat > Basic Statistics > 2 Variances



Testy hypotéz o parametru binomického rozdělení

Řešený příklad 8.9

Podle expertního předpokladu bude mít zájem o nový výrobek 20 % zákazníků. Ze 400 dotázaných zákazníků projevilo zájem 62 zákazníků. Na hladině významnosti 0,05 testujeme hypotézu o reálnosti předpokladu, tedy $H: p = 0,2$.

Ř e š e n í:

Rozsah obou výběru je dostatečně velký a pro $x = 62$ a $n = 400$ je pozorovaná hodnota testového

$$\text{kritéria } t = \frac{\frac{62}{400} - 0,2}{\sqrt{\frac{0,2(1-0,2)}{400}}} = \frac{-0,045}{0,02} = -2,25.$$

Z tabulky **T1** je $u_{0,975} = 1,960$. Protože $t = -2,25 \notin \bar{W}_{0,05} = \langle -1,960; 1,960 \rangle$, hypotézu o předpokladu 20 % zájmu zamítáme na hladině významnosti 0,05. Skutečný zájem bude pravděpodobně menší. Na hladině významnosti 0,01 však hypotézu nezamítáme, neboť $u_{0,995} = 2,576$.

Postup v Minitabu: Stat > Basic Statistics > 1 Proportion

Výstup z Minitabu:

Test

Null hypothesis $H_0: p = 0,2$

Alternative hypothesis $H_1: p \neq 0,2$

P-Value

0,021



Protože $p = 0,021 < 0,05$, hypotézu o předpokladu 20 % zájmu zamítáme na hladině významnosti 0,05. Ovšem $p = 0,021 > 0,01$, hypotézu o předpokladu 20 % zájmu nezamítáme na hladině významnosti 0,01.

Řešený příklad 3.10

Obchodní inspekce provedla 250 kontrolních nákupů potravinářského zboží a 200 kontrolních nákupů průmyslového zboží. Zjistila přitom nedostatky u 108 nákupů potravinářského zboží a u 73 nákupů průmyslového zboží. Na hladině významnosti 0,05 testujme, zda kvalita nákupů je stejná u obou druhů zboží, tedy hypotézu $H: p_1 = p_2$, kde p_1, p_2 jsou teoretické podíly (pravděpodobnosti) nákupů s nedostatky u daných druhů zboží.

Ř e š e n í:

Rozsahy obou výběrů jsou dostatečně velké a pro $x = 108$, $n_1 = 250$, $y = 73$, $n_2 = 200$ je

$$\bar{f} = \frac{108 + 73}{250 + 200} = 0,40222,$$

takže pozorovaná hodnota testového kritéria je

$$t = \frac{\frac{108}{250} - \frac{73}{200}}{\sqrt{0,40222(1 - 0,40222)}} \sqrt{\frac{250 \cdot 200}{250 + 200}} \approx \frac{0,067 \cdot 10,5409}{0,49035} \approx 1,4403.$$

Z tabulky **T1** je $u_{0,975} = 1,960$. Protože $t = 1,4403 \in \bar{W}_{0,05} = <-1,960; 1,960>$, hypotézu o rovnosti podílů nákupů s nedostatky nezamítáme na hladině významnosti 0,05 a považujeme prodej obou druhů zboží za stejně nekvalitní.

Postup v Minitabu je zřejmý: Stat > Basic Statistics > 2 Proportions

Testy hypotéz o rozdělení

Vzhledem k tomu, že testy o parametrech rozdělení závisejí na tvaru pozorovaných rozdělení, je zapotřebí testovat, zda pozorovaná náhodná veličina (náhodný vektor) má předpokládané rozdělení. Nejčastěji se užívají následující **testy hypotéz o rozdělení**. Testy hypotéz o rozdělení se také nazývají **testy dobré shody**.

Poznámka: Test chí-kvadrát (Pearsonův test) v Minitabu není.

Postup v Minitabu (Test normality):

Stat > Basic Statistics > Normality Test



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MSMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Normality Test

Variable: "V (ml)"

Percentile Lines

☒ None

☐ At Y values:

☐ At data values:

Tests for Normality

☒ Anderson-Darling

☐ Ryan-Joiner (Similar to Shapiro-Wilk)

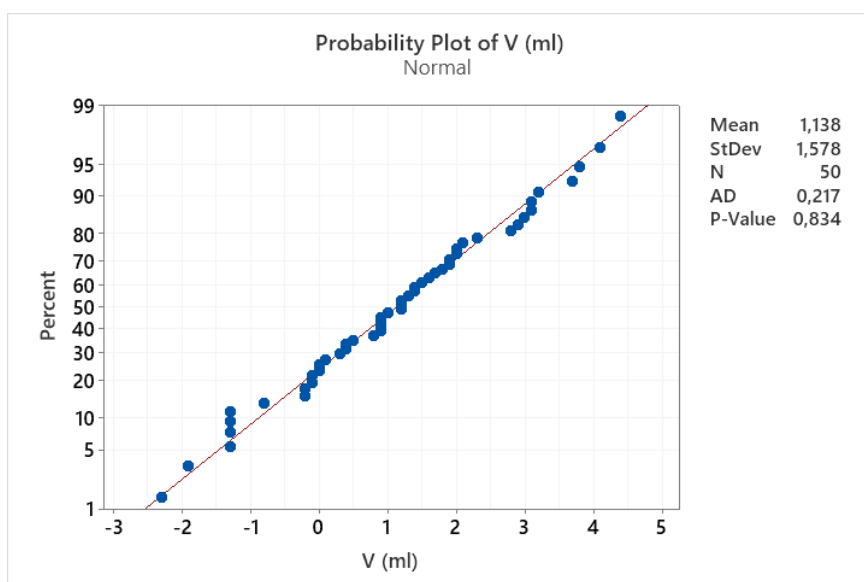
☐ Kolmogorov-Smirnov

Select

Title:

Help OK Cancel

Grafický výstup:



Protože $p = 0,834 > 0,05$, hypotézu o výběru z normálního rozdělení nezamítáme na hladině významnosti 0,05.



9 Regresní analýza

Řešený příklad 9.1

U osmi náhodně vybraných firem poskytujících konzultace v oblasti jakosti výroby byly v roce 1993 zjištěny počty zaměstnanců x a roční obraty y (mil. Kč):

x_i	3	5	5	8	9	11	12	15
y_i	0,8	1,2	1,5	1,9	1,8	2,4	2,5	3,1

Vyjádřete závislost ročního obrátu firmy na počtu zaměstnanců ve tvaru $y = \beta_1 + \beta_2 x$, vypočtěte intervalový odhad β_2 se spolehlivostí 0,95, testujte na hladině významnosti 0,05 hypotézu $H: \beta_1 = 0,2$, určete bodový a intervalový odhad $y(10)$ se spolehlivostí 0,95. Pomocí grafu a koeficientu korelace r posuďte vhodnost regresní funkce. Předpokládejte, že roční obrat má podmíněné normální rozdělení s konstantním rozptylem vzhledem k počtu zaměstnanců.

Ř e š e n í:

V následující tabulce jsou pomocné výpočty:

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	3	0,8	9	2,4	0,64
2	5	1,2	25	6,0	1,44
3	5	1,5	25	7,5	2,25
4	8	1,9	64	15,2	3,61
5	9	1,8	81	16,2	3,24
6	11	2,4	121	26,4	5,76
7	12	2,5	144	30,0	6,25
8	15	3,1	225	46,5	9,61
Σ	68	15,2	694	150,2	32,80

Vlastní výpočty provedeme v následujících krocích.

1) Jde o regresní přímku, takže s využitím výše uvedených vzorců obdržíme pro $n = 8$ z tabulky matici $\mathbf{H} = \begin{pmatrix} 8 & 68 \\ 68 & 694 \end{pmatrix}$, jejíž determinant je $\det \mathbf{H} = 8 \cdot 694 - 68^2 = 928$, takže bodový odhad β_2 je

$$b_2 = \frac{8 \cdot 150,2 - 68 \cdot 15,2}{928} = 0,1810344 \approx 0,181.$$



Dále je $\bar{x} = 68/8 = 8,5$, $\bar{y} = 15,2/8 = 1,9$, takže bodový odhad β_1 je

$$b_1 = 1,9 - 0,1810344 \cdot 8,5 = 0,3612068 \approx 0,361.$$

Potom bodový odhad regresní funkce je $y = 0,361 + 0,181x$.

2) Minimální hodnota reziduálního součtu čtverců je

$$S_{\min}^* = 32,80 - 0,3612068 \cdot 15,2 - 0,1810344 \cdot 150,2 \approx 0,1182758$$

a bodový odhad rozptylu σ^2 , resp. směrodatné odchylky σ , je

$$s^2 = 0,1182758 / (8 - 2) = 0,0197126, \text{ resp. } s = \sqrt{0,0197126} \approx 0,1404017.$$

3) Diagonální prvky matice \mathbf{H}^{-1} jsou

$$h^{11} = 694/928 \approx 0,7478448, \quad h^{22} = 8/928 \approx 0,00862069.$$

Z tabulky **T2** je pro $8 - 2 = 6$ stupňů volnosti $t_{0,975} = 2,447$. Intervalový odhad regresního koeficientu β_2 je

$$\beta_2 \in < 0,1810344 - 2,447 \cdot 0,1404017 \sqrt{0,00862069};$$

$$0,1810344 + 2,447 \cdot 0,1404017 \sqrt{0,00862069} > = < 0,1491353; 0,2129334 > \approx < 0,149; 0,213 >.$$

Bodový odhad přírůstku ročního obrátu odpovídajícího zvýšení počtu zaměstnanců firmy o jednoho je tedy 181 000 Kč a intervalový odhad tohoto přírůstku se spolehlivostí 0,95 je 149 000 Kč až 213 000 Kč.

4) Pozorovaná hodnota testového kritéria pro $H: \beta_1 = 0,2$ je

$$t = \frac{0,3612068 - 0,2}{0,1404017 \sqrt{0,7478448}} \approx 1,3277.$$

Pro alternativní hypotézu $\bar{H}: \beta_1 \neq 0,2$ je $\bar{W}_{0,05} = < -2,447; 2,447 >$. Vzhledem k tomu, že $t \in \bar{W}_{0,05}$, hypotézu $\beta_1 = 0,2$ na hladině významnosti 0,05 nezamítáme. Na dané hladině významnosti vlastně nezamítáme hypotézu, že firma bez zaměstnanců (pracují jen majitelé), neboť $y(0) = \beta_1$, bude mít roční obrát okolo 200 000 Kč.

5) Bodový odhad střední i individuální hodnoty ročního obrátu firmy pro 10 zaměstnanců je

$$y(10) = 0,3612068 + 0,1810344 \cdot 10 = 2,1715508 \approx 2,172.$$

U dané firmy lze tedy očekávat roční obrát okolo 2 172 000 Kč. Protože

$$h^* = \frac{1}{8} + \frac{8(10 - 8,5)^2}{928} = 0,1443965,$$

je intervalový odhad se spolehlivostí 0,95 střední hodnoty ročního obrátu firmy s 10 zaměstnanci



$$y(10) \in < 2,1715508 - 2,447 \cdot 0,1404017 \sqrt{0,1443965} ;$$

$$2,1715508 + 2,447 \cdot 0,1404017 \sqrt{0,1443965} > = < 2,0409985; 2,3021031 > \approx$$

$$\approx < 2,040; 2,302 >.$$

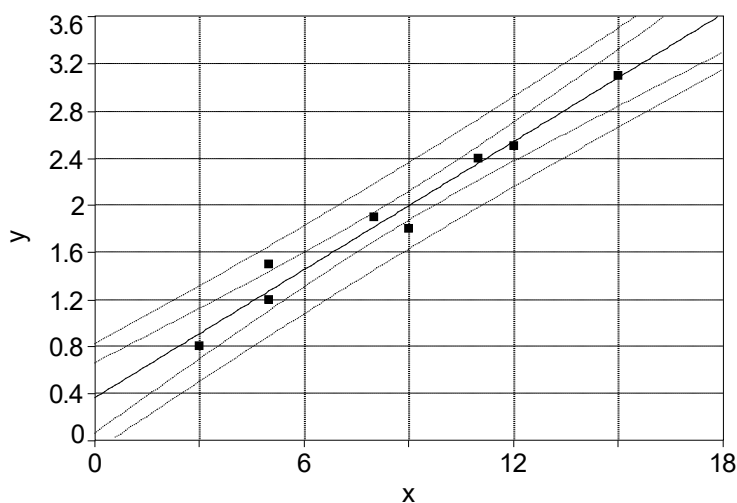
Se spolehlivostí 0,95 lze očekávat, že střední hodnota ročního obrátu takové firmy bude od 2 040 000 Kč do 2 302 000 Kč. Jestliže použijeme ve výpočtu $1 + h^*$ místo h^* , dostaneme intervalový odhad se spolehlivostí 0,95 individuální hodnoty ročního obrátu firmy s 10 zaměstnanci

$$y(10) \in < 2,1715508 - 2,447 \cdot 0,1404017 \sqrt{1,1443965} ;$$

$$2,1715508 + 2,447 \cdot 0,1404017 \sqrt{1,1443965} > = < 1,804; 2,539 >.$$

Se spolehlivostí 0,95 lze očekávat, že roční obrat (individuální hodnota ročního obrátu) takové firmy bude od 1 804 000 Kč do 2 539 000 Kč.

Závislost obrátu na počtu zaměstnanců



Obr. 4.2

6) Koeficient korelace (výpočet pomocí vzorce z kap. 1) je $r = 0,984798$, takže index determinace je $r^2 \approx 0,969827$. Z grafu na obr. 4.2 a velikosti koeficientu korelace vidíme, že zvolený tvar regresní funkce vcelku dobře vystihuje danou závislost. Podle často používané konvence lze říci, získaná regresní funkce vyjadřuje celkem $r^2 100 \% \approx 96,98 \%$ změn (variability) pozorovaného obrátu firmy.

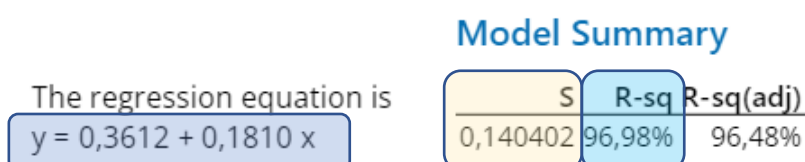


Postup v Minitabu:

Nejprve zjednodušený postup pro přímku.

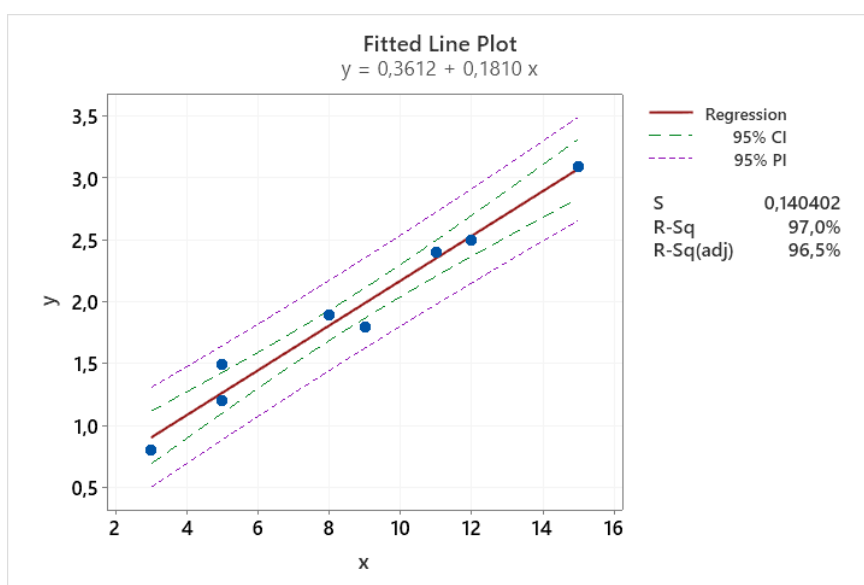
Stat > Regression > Fitted Line Plot

Výstup:



Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3,80172	3,80172	192,86	0,000
Error	6	0,11828	0,01971		
Total	7	3,92000			



Protože $p = 0,000 < 0,05$,
hypotézu, že model jako celek je nevýznamný zamítáme na hladině významnosti 0,05.



Podrobný model:

Stat > Regression > Regression > Fit Regression Model

Regression Equation

$$y = 0,361 + 0,1810 x$$

Coefficients

Term	Coef	SE Coef	95% CI	T-		VIF
				Value	P-Value	
Constant	0,361	0,121	(0,064; 0,658)	2,97	0,025	
x	0,1810	0,0130	(0,1491; 0,2129)	13,89	0,000	1,00



Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)	AICc	BIC
0,140402	96,98%	96,48%	0,194199	95,05%	0,99	-4,77

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	1	3,80172	96,98%	3,80172	3,80172	192,86	0,000
x	1	3,80172	96,98%	3,80172	3,80172	192,86	0,000
Error	6	0,11828	3,02%	0,11828	0,01971		
Lack-of-Fit	5	0,07328	1,87%	0,07328	0,01466	0,33	0,860
Pure Error	1	0,04500	1,15%	0,04500	0,04500		
Error							
Total	7	3,92000	100,00%				

Predikce:

Stat > Regression > Regression > Predict

Predict

Response: y

Enter individual values

x

10

Select

Options...

Results...

Storage...

View Model...

Help

OK

Cancel

Prediction

Fit	SE Fit	95% CI	95% PI
2,17155	0,0533520	(2,04100; 2,30210)	(1,80403; 2,53907)