

STATISTIKA

1. ZÁKLADNÍ POJMY

1.1 Náhodný pokus a náhodný jev

NÁHODNÝ POKUS – proces realizace souboru podmínek kde výsledek nemůžeme předem ovlivnit.

NÁHODNÝ JEJ – výsledek náhodného pokusu.

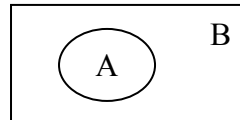
JISTÝ JEJ – jev, který za daných podmínek musí nastat.

NEMOŽNÝ JEJ – jev, který při realizaci téhož komplexu podmínek není možný.

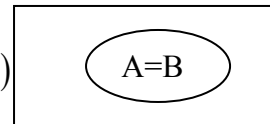
1.2 Vztahy mezi náhodnými jevy

Náhodné jevy značíme : $A, B, C, \dots A_1, B_1, C_1 \dots$

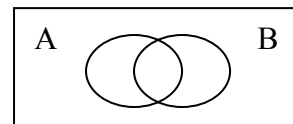
1.2.1 $A \subset B$ tzn.: nastane $A \Rightarrow$ nastane B
(A je částí B)



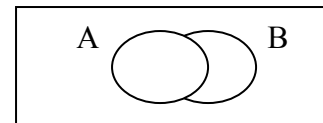
1.2.2 $A = B$ tzn.: $(A \Rightarrow B \wedge B \Rightarrow A)$; $(A \subset B \wedge B \subset A)$
(A a B si jsou rovny)



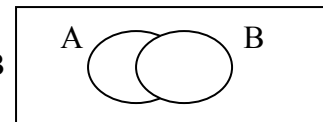
1.2.3 $A \cap B$ tzn.: nastane A a zároveň nastane i B
(průnik jevů)



1.2.4 $A \cup B$ tzn.: nastane alespoň jeden z jevů A nebo B
(sjednocení jevů)

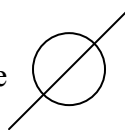


1.2.5 $A - B$ tzn.: nastane jev A a zároveň nenastane jev B
(rozdíl jevů)

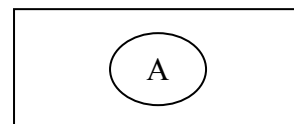


1.2.6 Jev Jistý – označujeme S

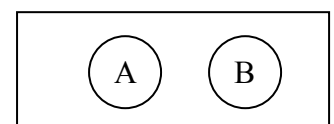
1.2.7 Jev Nemožný – označujeme $\emptyset$



1.2.8 Doplňěk jevu $\bar{A}$ – je opačný jevu A



1.2.9 Disjunktivní jevy (neslučitelné jevy) tzn.: $A \cap B = \emptyset$
Zobecnění : $\forall i, j : A_i \cap A_j = \emptyset$



Jevy A_1, A_2 jejichž sjednocení je jevem jistým – tzn.: $A_1 \cup A_2 \cup A_3 \dots = S$
toto nazýváme ULPNOU SKUPINOU JEVU a v případě jejich

neuskutečnitelnosti toto nazýváme UPLNOU SKUPINOU NESLUCITELNYCH JEVU.

Elementární jevy

a) Tvoří úplnou skupinu (úplný systém) neslučitelných jevů. Značíme **E**

$$\text{Tzn.: } E_1 \cup E_2 \cup E_3 \dots = S \wedge \forall ij : E_i \cap E_j = \emptyset$$

b) V dané situaci nejsou rozkládány na jevy podrobnější. Všechny elementární jevy tvoří tzv. základní soubor S.

$$A \subset B \wedge B \subset A \Leftrightarrow A = B$$

$$A \cup A = A$$

$$A \subset (A \cup B)$$

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

$$A \cap A = A$$

Platí že: $A \cap S = A$

$$A \cup \emptyset = A$$

$$A \cup \bar{A} = S$$

$$A \cap \bar{A} = \emptyset$$

$$\overline{\overline{A}} = A$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

1.3 Pravděpodobnost náhodného jevu

- pravděpodobnost náhodného jevu A je číslo P(A), které můžeme interpretovat jako míru možnosti nastoupení náhodného jevu.

1.3.1 Axiomatická definice

- pravděpodobnost je funkce, která každému náhodnému jevu přiřazuje reálné číslo přičemž splňuje 3 axiomy.

1.Axiom $P(A) \geq 0$ tzn.: nezáporné číslo

2.Axiom Pravděpodobnost sjednocení početné množiny neslučitelných jevů A1, A2, A3 ... náleží S je rovna součtu pravděpodobností těchto jevů.
 $P(A_1 \cup A_2 \cup A_3) = P(A_1) \oplus P(A_2) \oplus P(A_3)$

3.Axiom Pravděpodobnost jistého jevu (S) je rovna 1. $P(S)=1$

$$A \subset B \Rightarrow P(A) \leq P(B)$$

$$A = B \Rightarrow P(A) = P(B)$$

Z toho vyplývá: $P(\bar{A}) = 1 - P(A)$

$$P(\emptyset) = 0$$

$$0 < P(A) \leq 1$$

1.3.2 Klasická definice

$$P(A) = \frac{M_A}{N}$$

M_A ... počet jevů příznivých jevu A

N ... počet všech možných jevů

1.3.2.1 Statistická definice

-založena na pozorování a četnosti výskytu jevu A, a opakovaných nezávislých pokusech. Hodnotu pravděpodobnosti jevu A v dostatečně velké sérii pokusů.

$$P(A) \approx \frac{M}{N}$$

M...počet výskytů jevu A, N...počet nez. Pokusů

1.4 Geometrická definice

-uzavřená oblast S

$$P(A) = \frac{\mu(A)}{\mu(S)}$$

-reálný jev A, $A \subset S$

$\mu(A)$...míra.jevu..A

$\mu(S)$...míra.oblasti..S

1.5 Podmíněná pravděpodobnost

$P(A/B)$... pravděpodobnost jevu A vzhledem k jevu B (tzn.: jev B již nastal)

Def.: $P(A/B) = \frac{P(A \cap B)}{P(B)}$ pro $P(B) > 0$

Odborně: $P(B/A) = \frac{P(A \cap B)}{P(A)}$ z toho plyne $P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A/B) * P(B)}{P(A)}$

Bayesův vzorec: $P(A \cap B) = P(A) * P(B/A) = P(B) * P(A/B)$

Obecně: $P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) * P(A_2/A_1) * P(A_3/A_1 \cap A_2) \dots$

Příklad: Dvě mince

Jev A –padne na obou RUB

Jev B –padne na první RUB

Všechny možné jevy: RR, RL, LR, LL

$$P(A) = \frac{1}{4}$$

$$P(B) = \frac{1}{2}$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

Nezávislost jevů

Dva jevy jsou nezávislé, jestliže pravděpodobnost jednoho z nich nezávisí na tom zda druhý nastal nebo nenastal.

Matematická definice: Jev A je nezávislý na jevu B $\Leftrightarrow P(A/B) = P(A)$

Nezávislost dvou jevů je vždy oboustranná.

Pro nezávislé jevy platí: $P(A \cap B) = P(A) * P(B)$

1.6 Věta o sčítání pravděpodobností

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Pro neslučitelné: $P(A \cup B) = P(A) + P(B)$

1.7 Věta o úplné pravděpodobnosti

$H_1, H_2, H_3 \dots H_n$ tvoří úplný systém neslučitelných jevů $P\left(\bigcup_{i=1}^n H_i\right) = \sum_{i=1}^n P(H_i) = 1$

Jev A může nastat jen tehdy, nastane-li právě jeden z jevů H_i , pak platí:

$$P(A) = \sum_{i=1}^n P(H_i) * P(A/H_i)$$

Jevy H_i za nichž může nastat jev A se někdy nazývají hypotézy jevu A.

1.8 Bayerova věta

Nechť je dán jev A, který se může uskutečnit právě za jednu z n podmínek o nichž učiníme hypotézy $H_1, H_2, H_3 \dots H_n$. Byl proveden pokus jehož výsledek bylo

nastoupení jevu A pak:
$$P(H_i/A) = \frac{P(H_i) * P(\bar{A}/H_i)}{P(A)} = \frac{P(H_i) * P(A/H_i)}{\sum_{j=1}^n P(H_j) * P(A/H_j)}$$

Protože $H_1, H_2, H_3 \dots H_n$ vyčerpávají všechny předpoklady musí platit: $\sum_{i=1}^n P(H_i/A) = 1$

Pravděpodobnost zpoždění rychlíku způsobeného technickou závadou je 0,2. Určete pravděpodobnost toho, že budou zpožděny alespoň tři vlaky z osmnácti sledovaných.

Řešení: - nezávislé Bernaliho schéma $P_a(x) = \binom{a}{x} * P(x)^x * \bar{P}(x)^{a-x}$

a – počet testovaných

x – počet zkoumaných

Nezpoždí se žádný: $P_{18}(0) = \binom{18}{0} * 0,2^0 * (1-0,2)^{18}$

Zpozdí se jeden: $P_{18}(1) = \binom{18}{1} * 0,2^1 * (1-0,2)^{17}$

Zpozdí se dva: $P_{18}(2) = \binom{18}{2} * 0,2^2 * (1-0,2)^{16}$

Alespoň tři (doplňek součtu všech do 1): $P = 1 - P_{18}(0) - P_{18}(1) - P_{18}(2) = \underline{\underline{0,728}}$

Třikrát vystřelíme na cíl. Pravděpodobnost zásahu při každém výstřelu je 0,7. Určete:

- Rozdělení pravděpodobností počtu zásahů při třech nezávislých výstřelech
- Distribuční funkci a sestavte její graf

Řešení:

a) – nezávislé Bernaliho schéma $P_a(x) = \binom{a}{x} * P(x)^x * \bar{P}(x)^{a-x}$

Netrefím žádný $P_3(0) = \binom{3}{0} * 0,7^0 * (1-0,7)^3 = 0,027$

Trefím jeden $P_3(1) = \binom{3}{1} * 0,7^1 * (1-0,7)^2 = 0,189$

Trefím dva $P_3(2) = \binom{3}{2} * 0,7^2 * (1-0,7)^1 = 0,441$

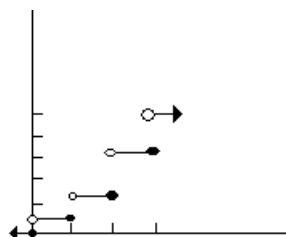
Trefím všechny tři $P_3(3) = \binom{3}{3} * 0,7^3 * (1-0,7)^0 = 0,343$

Počet zásahů	0	1	2	3
P(úspěšnosti)	0,027	0,189	0,441	0,343

b) Zde musíme vědět, že distribuční (komulativní) pravděpodobnost se sčítá, takže začneme od 0, další členy nabývají velikost plus předchozí člen. Rozdělení na ose je od $-\infty$, postupně podle střel takže do 0, pak od 0 do 1, ..., až od 3 do ∞ .

$$F(X) = \begin{cases} 0 & x \in (-\infty; 0) \\ 0,027 & x \in (0; 1) \\ 0,216 & x \in (1; 2) \\ 0,657 & x \in (2; 3) \\ 1 & x \in (3; \infty) \end{cases}$$

graf:



Náhodná veličina X má v intervalu $(-\infty; \infty)$ hustotu pravděpodobnosti definovanou vztahem:

$$f(x) = \begin{cases} 0 & x \notin \langle 0; 1 \rangle \\ b(x^2 + 2x + 1) & x \in \langle 0; 1 \rangle \end{cases}$$

Stanovte konstantu b a distribuční funkci náhodné veličiny X .

Řešení: Nejprve musíme stanovit hodnotu b , kterou zjistíme ze vzorečku $b(x^2 + 2x + 1)$, ale nesmíme zapomenout, že $x \in \langle 0; 1 \rangle$, takže to musíme vypočítat pomocí integrálu od 0 do 1.

$$\int_0^1 b(x^2 + 2x + 1) = 1$$

$$b \left[\frac{x^3}{3} + \frac{2x^2}{2} + x \right]_0^1 = 1$$

$$b \left(\frac{1}{3} + 1 + 1 \right) = 1$$

$$b = \frac{3}{7}$$

Druhá část je zjistit distribuční funkci náhodné veličiny X . Známe vzoreček:

$$F(x) = \int_{-\infty}^x (u^2 + 2u + 1) du, \text{ který zde využijeme. Nesmíme ale zapomenout rozložit integrál na}$$

integrály, to znamená že můžeme rozdělit hranice $\int_{-\infty}^x = \int_{-\infty}^0 + \int_0^x$. A ještě víme, že $\int_{-\infty}^0 = 0$.

$$\begin{aligned} F(x) &= \int_{-\infty}^x (u^2 + 2u + 1) du = \int_0^x b(u^2 + 2u + 1) du = \int_0^x \frac{3}{7} (u^2 + 2u + 1) du = \frac{3}{7} \left[\frac{u^3}{3} + u^2 + u \right]_0^x = \\ &= \frac{3}{7} * \left(\frac{x^3}{3} + x^2 + x \right) \end{aligned}$$

Poslední krok je zapsat distribuční funkci:

$$F(X) = \begin{cases} 0 & x \in (-\infty; 0) \\ \frac{3}{7} * \left(\frac{x^3}{3} + x^2 + x \right) & x \in (0; 1) \\ 1 & x \in (1; \infty) \end{cases}$$

Rozdělení pravděpodobnosti náhodné veličiny X je dáno tabulkou. Najděte střední hodnotu a disperzi náhodné veličiny $Y = 5 - 2x$.

x_i	-1	0	1	2	3
$p(x_i)$	0,1	0,3	0,2	0,25	0,15

Řešení: První část úkolu je nalezení střední hodnoty E_x . Na tu máme vzoreček $E_x = \sum_x P(x)$.

$$E_x = (x_1 * p(x_1)) + (x_2 * p(x_2)) + (x_3 * p(x_3)) + (x_4 * p(x_4)) + (x_5 * p(x_5)) =$$

$$= (-1 * 0,1) + (0 * 0,3) + (1 * 0,2) + (2 * 0,25) + (3 * 0,15) = \underline{\underline{1,05}}$$

Když už známe E_x , tak můžeme vypočítat E_y ze vzorečku $Y = 5 - 2x$. Vypadá to takhle:

$$Y = 5 - 2x \Leftrightarrow E_y = E(5 - 2x) = E(5) - E(2x) = 5 - 2E(x) = 5 - 2 * (1,05) = \underline{\underline{2,9}}$$

$E(x)$ se pro nás zde rovná s E_x , proto upravíme do fáze, kdy můžeme dosadit. Nesmíme zapomenout, že když máme **$E(5)$ (střední hodnotu 5tky) je 5.**

Naším druhým úkolem je vypočítat disperzi náhodné veličiny $Y = 5 - 2x$. K tomu nám pomůže vzoreček $D(x) = \sum_{x_i} (x_i - E(x))^2 * P(x_i)$ neboli **ROZPTYL**.

$$DX = (x_1 - E_x)^2 * P(x_1) + (x_2 - E_x)^2 * P(x_2) + (x_3 - E_x)^2 * P(x_3) + (x_4 - E_x)^2 * P(x_4) +$$

$$+ (x_5 - E_x)^2 * P(x_5) = (-1 - 1,05)^2 * 0,1 + (0 - 1,05)^2 * 0,3 + (1 - 1,05)^2 * 0,2 + (2 - 1,05)^2 * 0,25 +$$

$$+ (3 - 1,05)^2 * 0,15 = \underline{\underline{1,547}}$$

Poslední věc, kterou uděláme je, že z vlastního rozptylu $D(kx) = k^2 * D(x)$. Víme že **$D(x)$ (neboli DX) je 1,547** a máme vztah $Y = 5 - 2x$, z kterého můžeme odstranit 5 a zbude nám $Y = -2x$. Ted' už stačí pouze dosadit:

$Y = 5 - 2x = -2x$ a dosadíme do $D(kx) = k^2 * D(x)$. -2 je pro nás k , $D(x)$ je x :

$$D(kx) = k^2 * D(x) = (-2)^2 * 1,547 = \underline{\underline{6,19}}$$

Náhodná veličina X má konstantní hodnotu pravděpodobností v intervalu $(0, a)$, to znamená, že její hustota pravděpodobnosti má tvar

$$f(x) = \begin{cases} 1/a & 0 < x < a \\ 0 & x \leq 0 \vee x \geq a \end{cases}$$

S použitím vlastností střední hodnoty a rozptylu určete:

$$E(2X + 3), E(3X^2 - 2X + 1), D(2X + 3), D(X^2 + 1).$$

Řešení: Zde použijeme druhý důležitý vzoreček, který potřebujeme a to: $EX = \int_{-\infty}^{\infty} xf(x)dx$.

Nejprve si vypočítáme střední hodnotu:

$$EX = \int_{-\infty}^{\infty} xf(x)dx = \int_0^a x \frac{1}{a} dx = \frac{1}{a} \left[\frac{x^2}{2} \right]_0^a = \frac{a}{2}$$

Pro výpočet využijeme rozptyl Dx , který nám říká, že když $(EX)^2$ odečteme od EX^2 (neboli $DX = E(x^2) - (EX)^2$), tak ho získáme. EX víme, že je $\frac{a}{2}$. Takže nám stačí vypočítat $E(x^2)$.

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^a x^2 \frac{1}{a} dx = \frac{1}{a} \left[\frac{x^3}{3} \right]_0^a = \frac{a^2}{3}$$

Dosadíme do vzorečku:

$$DX = E(x^2) - (EX)^2 = \frac{a^2}{2} - \left(\frac{a}{2}\right)^2 = \frac{a^2}{2} - \frac{a^2}{4} = \left(\frac{2-1}{4}\right)a^2 = \frac{a^2}{4}$$

Dvourozměrná náhodná veličina

-tedy náhodný vektor (X, Y)

-zákon rozdělení pravděpodobnosti je dán ve formě **sdrúžené distribuční funkce** $F(x, y)$

$F(x, y) = P(X \leq x, Y \leq y)$ – tyto dvě nerovnosti platí zároveň

-sdrúžená distribuční funkce je opět funkcí neklesající, spojitou zleva, a to vzhledem ke každé z náhodných veličin. $F(-\infty; y) = F(x; -\infty) = F(-\infty; -\infty) = 0$ a $F(\infty; \infty) = 1$ – vyplývá

z nemožnosti jevů $X < -\infty, Y < -\infty$ a z jistoty jevu $X < \infty, Y < \infty$

.-platí:

$$P(x_1 \leq X < x_2; y_1 \leq Y < y_2) = P(X < x_2; Y < y_2) - P(X < x_2, Y < y_1) - P(X < x_1; Y < y_2) + P(X < x_1, Y < y_1) \quad \text{*****}$$

-Uvažujeme-li pouze pravděpodobnost, že náhodná veličina X nabude hodnoty menší než x bez jakékoliv podmínky, pro veličinu Y (hodnota této veličiny může být jakákoliv) potom tento vztah $P(X < x, Y < \infty) = F(x, \infty) = F_X(x)$

- Jde jednorozměrnou distribuční funkci $F_X(x)$ nazveme **marginální (okrajovou) distribuční funkcí náhodné veličiny X** . Obdobně pro náhodnou veličinu Y potom platí

$$P(X < \infty, Y < y) = F(\infty, y) = F_Y(y) \text{ – marginální distribuční funkce náhodné veličiny } Y$$

-Pro diskrétní náhodné veličiny X, Y :

-Sdrúžené pravděpodobnosti $P(X=x, Y=y) = P(x, y)$

-Splňují vztahy:

1. nezáporné

$$2. \sum_x \sum_y P(x, y) = 1$$

$$3. P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \sum_{x=x_1}^{x_2} \sum_{y=y_1}^{y_2} P(x, y)$$

marginální (okrajové) pravděpodobnosti

$$-P_X(x) = \sum_y P(x, y) \text{ – součet marginálních pravděpodobností}$$

$$-P_Y(y) = \sum_x P(x, y) \text{ – Součet sdrúžených pravděpodobností hodnoty } Y$$

-sdrúžené $P(x, y)$ a marginální pravděpodobnosti $P_X(x)$ a $P_Y(y)$ zapisujeme do tabulky sdrúžených a marginálních pravděpodobností

x/y	y_1	y_2	y_3	y_s	$\sum_y$
x_1	$P(x_1, y_1)$	$P(x_1, y_2)$	$P(x_1, y_3)$	$P(x_1, y_s)$	$P_X(x_1)$
x_2	$P(x_2, y_1)$	$P(x_2, y_2)$	$P(x_2, y_3)$	$P(x_2, y_s)$	$P_X(x_2)$
x_3	$P(x_3, y_1)$	$P(x_3, y_2)$	$P(x_3, y_3)$	$P(x_3, y_s)$	$P_X(x_3)$
x_R	$P(x_R, y_1)$	$P(x_R, y_2)$	$P(x_R, y_3)$	$P(x_R, y_s)$	$P_X(x_s)$
$\sum_x$	$P_Y(y_1)$	$P_Y(y_2)$	$P_Y(y_3)$	$P_Y(y_s)$	1

Sdružená distribuční funkce

$$F(x, y) = \sum_{t < x} \sum_{u < y} P(t, u) - \text{pro diskrétní } X, Y$$

$$F(x, y) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(t, u) dt \right] du - \text{pro spojitě}$$

Je dvojrozměrná sdružená hustota pravděpodobnosti náhodných veličin (X, Y) , která splňuje:

1. $f(x, y) \geq 0$ všechna x, y náleží $\mathbb{R}$ – jsou nezáporná
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$, $P(x_1 \leq X < x_2, y_1 \leq Y < y_2) = \int_{y_1}^{y_2} \left[\int_{x_1}^{x_2} f(x, y) dx \right] dy$ -platí pro otevřené i uzavřené intervaly.

Marginální (okrajové) hustoty jsou $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$ $f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$

-ze známé distribuční (sdružené) funkce $F(x, y)$ stanovíme hustotu pravděpodobnosti

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

Př: Na výrobcích měříme délku s přesností $\pm 0,5$ mm a šířku 0,2 mm. Označme náhodnou veličinu X chybu při měření délky a náhodnou veličinu Y chybu při měření šířky. Sdružená hustota pravděpodobností $f(x, y)$ uvnitř mezí chyb je rovnoměrně rozložena tedy $f(x, y)$ x náleží $(-0,5; 0,5)$; y náleží $(-0,2; 0,2)$, $f(x, y) = 0$ jinde

- a) určete k
- b) marginální hustoty pravděpodobnosti
- c) určete sdruženou parciální derivaci a marginální distribuční funkci
- d) pravděpodobnost, že délku změříme s maximální chybou $\pm 0,1$ mm (délka) a zároveň šířku s maximální chybou $\pm 0,1$ mm.

$$a) \int_{-0,2}^{0,2} \left[\int_{-0,5}^{0,5} k dx \right] dy = 0,4k = 1 = \underline{\underline{k = 2,5}}$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_{-0,2}^{0,2} 2,5 dy = [2,5y]_{-0,2}^{0,2} = \underline{\underline{1}}$$

$$b) f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_{-0,5}^{0,5} 2,5 dx = \underline{\underline{2,5}}$$

$$F(x, y) = \int_{-0,2}^y \left[\int_{-0,5}^x 2,5 dt \right] dy = 2,5xy + 0,5x + 1,25y + 0,25 = (x + 0,5) * (2,5y + 0,5)$$

$$c) F_X(x) = F(x; 0,2) = \int_{-0,5}^x \left[\int_{-0,2}^{0,2} f(t, y) dy \right] dt = x + 0,5$$

$$F_Y(y) = F(0,5; y) = \int_{-0,2}^y \left[\int_{-0,5}^{0,5} f(x, u) dx \right] du = 2,5y + 2,5$$

$$P(-0,1 < X < 0,1; -0,1 < Y < 0,1) = F(0,1; 0,1) - F(0,1; -0,1) - F(-0,1; 0,1) + F(-0,1; -0,1) =$$

d) $= 0,45 - 0,15 - 0,30 + 0,10 = \underline{\underline{0,10}}$

Nezávislost náhodných veličin

- u jevů $P(A|B) = P(A) \Leftrightarrow$ nezávislé
- definice podmíněné pravděpodobnosti $P(A|B) = \frac{P(A \cap B)}{P(B)} \quad P(B) \neq 0$
- obdobně dvě náhodné veličiny X a Y jsou nezávislé, jestliže zákon rozdělení jedné veličiny nezávisí na tom jaké hodnoty nabyla druhá veličina neboli jestliže podmíněný zákon rozdělení této veličiny se neliší od zákona nepodmíněného (marginálního)
- podmíněná distribuční funkce $F(x|y)$

$$\text{a) diskrétní případ } F(x|y) = \frac{\sum_{t \leq x} P(t, y)}{P_y(y)} \quad P_y(y) \neq 0$$

$$\text{b) ve spojitém případě } F(x|y) = \frac{\int_{-\infty}^x f(t, y) dt}{f_y(y)} \quad f_y(y) \neq 0$$

- **y je předem dáno** – tedy je to distribuční funkce jednorozměrná
- $f(x|y) = \frac{f(x, y)}{f_y(y)} \quad f_y(y) \neq 0$
- Záměnou veličin vznikají podmíněné zákony rozdělení veličiny Y pro dané $X=x$
- Nezávislost veličin: jestliže náhodné veličiny X, Y mají sdruženou distribuční funkci $F(x, y)$, pak náhodné veličiny X, Y jsou nezávislé právě tehdy a jen tehdy jestliže $F(x, y) = F_X(x) * F_Y(y)$
- Popíšeme-li sdružené rozdělení pravděpodobností $P(x, y)$ pak jsou nezávislé právě tehdy $P(x, y) = P_X(x) * P_Y(y)$
- Popisujeme-li sdružené rozdělení pravděpodobností sdruženou hustotou $f(x, y)$ nezávislé $\Leftrightarrow f(x, y) = f_X(x) * f_Y(y)$
- Jsou-li náhodné veličiny X, Y nezávislé pak jejich podmíněná rozdělení **jsou rovna marginálním rozdělení**

$$F(x|y) = F_X(x)$$

- Například (pro všechna y) $\forall y: P(x|y) = P_X(x)$
 $f(x|y) = f_X(x)$

Charakteristiky náhodných veličin

a) polohy (EX, medián, kvantit)

EX – předchozí látka

- 1. počáteční moment

$$\text{kvantil} - F(x_p) = \int_{-\infty}^{\infty} f(x) dx = p \quad P(X < x_p) = p \quad 100 \cdot p \% \text{ kvantil}$$

Medián – střední hodnota – 50% kvantil = $F(x_p) = 0,5$

Modus – kde maximum $f(x)$

b) Variability - DX (vlastnosti předchozí látka) - 2. centrální moment

c) Šikmost – můžeme mít zešikmení doprava či doleva – 3. centrální moment

$$- A_3(x) = \frac{E([X - EX]^3)}{(\sqrt{DX})^3}$$

d) Špičatost – 4. centrální moment

$$- A_4 = \frac{E([X - EX]^4)}{(\sqrt{DX})^4}$$

- směrodatná odchylka $\delta = \sqrt{DX}$

- (exces, kurtosis)

Charakteristiky dvourozměrné náhodné veličiny

- jako v jednorozměrných (polohy, variability, šikmosti – a to pro každé marginální rozdělení)

- dále máme charakteristiky podmíněné (Regresní funkce, Skedastická funkce)

- charakteristika vzájemného vztahu – kovariace (míra těsnosti)

- kovariace patří mezi součinnová momenty a definujeme ji jako střední hodnotu součinu odchylek obou veličin od jejich středních hodnot

$$- \text{cov}(X, Y) = E[(X - EX) * (Y - EY)] \text{ neboli } \text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - EX) * (y - EY) f(x, y) dx dy$$

$$- \text{cov}(X, Y) = E(XY) - EX * EY \quad \text{neboli } \text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x * y * f(x, y) dx dy - EX * EY$$

Na kovarianci je založen Koeficient Korelace

Koeficient Korelace

$$- \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{DX} * \sqrt{DY}} \Leftrightarrow \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)} * \sqrt{\text{var}(Y)}} \Leftrightarrow \rho(X, Y) = \frac{\text{cov}(X, Y)}{\rho(X) \rho(Y)}$$

- koeficient korelace je bez rozměrnou charakteristikou měřící těsnost vztahu mezi dvěma veličinami a nabývá hodnot v intervalu $\langle -1; 1 \rangle \Leftrightarrow -1 \leq \rho(X, Y) \leq 1$

- veličiny se nazývají nekorelované

- ale pozor! Z toho, že náhodné veličiny jsou nekorelované, obecně neplyne, že jsou nezávislé. $X, Y \Rightarrow \rho(X, Y) = 0$; ρ je mírou lineární nezávislosti

Vlastnosti cov

1. $\text{cov}(a_1, Y) = \text{cov}(X, a_2) = \text{cov}(a_1, a_2) = 0$
2. $\text{cov}(a_1 + b_1 X, a_2 + b_2 Y) = b_1 b_2 \text{cov}(X, Y)$
3. $\text{cov}(X, X) = DX$
4. $\text{cov}(X, Y) = \text{cov}(Y, X)$
5. $\text{cov}(X, Y) = E(XY) - EX * EY$
6. $\text{cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{cov}(X_i, Y_j)$

Příklad:

$$f(x, y) = x + y \quad 0 < x < 1, 0 < y < 1$$

$$= 0 \quad \text{jinak}$$

$$\text{cov}=? \quad \rho=?$$

$$EX = EY = \frac{7}{12} \quad DX = DY = \frac{11}{144}$$

$$\left(EX = \int_0^1 x f_x(x) dx; EY = \int_0^1 y f_y(y) dy \right) \quad DX = E(X^2) - EX^2 \quad DY = E(Y^2) - EY^2$$

$$E(XY) = \int_0^1 \int_0^1 x * y f(x, y) dx dy = \int_0^1 \int_0^1 x * y (x + y) dx dy = \frac{1}{3}$$

$$\text{cov}(X, Y) = E(XY) - EX * EY = -\frac{1}{144}$$

$$\Rightarrow \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{DX} * \sqrt{DY}} = -\frac{1}{11} \text{ - velmi slabá nepřímá lineární závislost (obě veličiny jsou}$$

velmi slabě korelované)

Rozdělení pravděpodobností náhodných veličin

- a) Diskrétních náhodných veličin - diskrétní rozdělení nám vznikají
- Alternativní rozdělení $A(\pi)$
 - Alternativní rozdělení nastane s pravděp. π a nenastane pravdě. $1-\pi$.

A nastane $\rightarrow 1$ $P(X = 1) = \pi$
A nastane $\rightarrow 0$ $P(X = 0) = 1 - \pi$

X_i	0	1
$\Pi(x_i)$	$1-\pi$	π

- praxi pro zkoumání jevu, které se realizují pouze dvěmi variantami
- (výskyt $\frac{\text{dobrého}}{\text{vadného}}$ výrobku při technické hodnotě, $\frac{\text{úspěch}}{\text{neúspěch}}$ tel. Signálu na 1. pokus.

Příklad: Úspěch spojení na 1. pokus
 $\pi = 0,23$ (pravděpodobnost)
 $1 = \text{úspěch}, 0 = \text{neúspěch}$

$$A(0,23), EX=0,23, DX=0,23*0,77=0,1771$$

Binomické rozdělení $Bi(n;p)$

- opakované Benoulliho pokusy (nezávislost)
- sledujeme rozdělení, že jev A nastane právě x-krát $\forall x = 0,1,2,\dots,n \rightarrow$ binomické rozdělení

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad p - \text{pravděpodobnost úspěchu jevu A}$$

$$EX = np$$

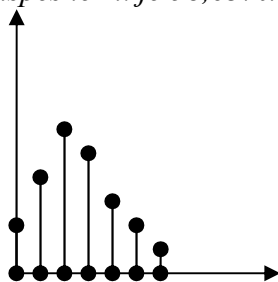
$$DX = np(1-p)$$

- Binomické rozdělení vzniká vlastně složením u nezávislých veličin s alternativním rozdělením (tj. počet špatných výrobků zjištěných ve výběru n výrobků; účinnost přípravku na n subjektech; připojení k serveru nezávislých pokusech)
- *Příklad: Snažíme se zalogovat na počítač deseti nezávislých pokusech. Pravděpodobnost úspěšného zalogování na první pokus 0,23. Určete pravděpodobnost, že se nám podaří zalogovat nejvýše dvakrát a nakresl pravděpodobnostní funkci rozdělení.*

$$bi(10;0,23)$$

$$P(x \leq 2) = \sum_0^2 \binom{10}{x} * 0,23 * 0,77^{10-x} = \underline{\underline{0,5863}} - \text{pravděpodobnost, že se nalogují}$$

úspěšně 2x je 58,63%.



- čísla velmi blízko nule (od 7 do 10)

Poissonovo rozdělení – PO(λ)

- jestliže u binomického rozdělení je n dostatečně velké a p je dostatečně malé

$$\rightarrow (p \leq 0,1; n \geq 30) \text{ pak } \binom{n}{x} p^x (1-p)^{n-x} \approx \frac{\lambda^x}{x!} \quad \begin{array}{l} \lambda = np \\ x = 0, 1, \dots, n \end{array}$$

$$EX = \lambda \quad DX = \lambda$$

λ – intenzita počtu (průměrný očekávaný počet výskytu sledovaného jevu v za jednotku času = např. příjezdy aut k benzínové stanici, příchody k obsluze, příchody zákazníků k pokladně)

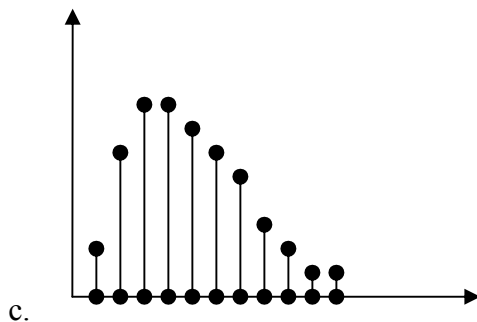
- př: bylo zjištěno při výrobě polyn. vlákna dojde na zprádacích strojích za hodinu průměrně k ucpání čtyř trysek, které je nutno vyměnit a vyčistit. Tento počet je náhodná veličina X s poissonovým rozdělením pravděpodobností s parametrem $\lambda=4[\text{hod}^{-1}]$. Určete pravděpodobnost, že bude třeba vyměnit za :

- Právě dvě trysky
- Nejvýše dvě trysky
- Nakreslete graf rozdělení pravděpodobností

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} = \frac{4^x}{x!} e^{-4}$$

$$\text{a. } p(2) = \frac{4^2}{2!} e^{-4} = \underline{\underline{0,1465}}$$

$$\text{b. } p(x \leq 3) = F(3) = \frac{4^0}{0!} e^{-4} + \frac{4^1}{1!} e^{-4} + \frac{4^2}{2!} e^{-4} = \underline{\underline{0,2381}}$$



Spojité rozdělení

Normální rozdělení

Hustota $f(x) = \frac{1}{\delta\sqrt{2\pi}} * e^{-\frac{(x-\mu)^2}{2\delta^2}} \quad -\infty < x < \infty$
 $-\infty < \mu < \infty, \delta > 0$

$EX = \mu \quad DX = \delta^2$

$N(\mu, \delta^2)$ Pozor!: Kubanová/Linda $N(\mu; \delta)$

$F(x) = \int_{-\infty}^x f(v)dv = \int_{-\infty}^x \frac{1}{\delta\sqrt{2\pi}} * e^{-\frac{(v-\mu)^2}{2\delta^2}} * dv$

$Z = \frac{x-\mu}{\delta} \rightarrow N(0,1)$

Koupit: Kritické hodnoty bla bla bla, Kubanová

Exponenciální rozdělení – $Exp(\alpha)$

- z Poissonova rozdělení jestliže sledujeme jev mezi dvěma časovými událostmi

$f(x) = \alpha * e^{-\alpha x} \quad F(x) = 1 - e^{-\alpha x}$

$EX = \frac{1}{\alpha} \quad DX = \left(\frac{1}{\alpha}\right)^2 \quad x \geq 0 \quad \alpha > 0$

- často je α průměrný počet výskytu sledovaného jevu za časovou jednotku
- obecně je to doba životnosti nebo doba čekání na nějakou událost

Rovnoměrné rozdělení - $R(a; b)$

$f(x) = \frac{1}{(b-a)} \quad x \in (a, b)$

$EX = \frac{a+b}{2} \quad DX = \frac{(b-a)^2}{12}$

- pravděpodobnost je rovnoměrně rozdělena na celém intervalu
- například odečítání měření na lineární stupnici

Tebyševova nerovnost

- pro jakoukoliv náhodnou veličinu X se střední hodnota EX a variabilitou DX a $\forall \varepsilon > 0$ je

$$P(|X - EX| < \varepsilon) \geq 1 - \frac{DX}{\varepsilon^2} \Leftrightarrow \left(P(|X - \mu| \geq \varepsilon) \leq \frac{\delta^2}{\varepsilon^2} \right)$$

- pro jakékoliv $k \geq 1$ nejméně $\left(1 - \left(\frac{1}{k^2}\right)\right) * 100\%$ leží v intervalu

$(\mu - k * \delta; \mu + k * \delta) \Rightarrow 50\% \text{ je v } \pm 1,414\delta$
 $\text{Nejméně } 75\% \text{ je v } \pm 2\delta$
 $\text{Nejméně } 88,88\% \text{ je v } \pm 3\delta$

Zákon velkých čísel

-jestliže $X_1, X_2, \dots, X_n$ je posloupnost podvojně nezávislých náhodných veličin mající střední hodnoty $E(X_1), E(X_2), \dots, E(X_n)$ a shora ohraničené rozptyly

$D(X_1) \leq c, D(X_2) \leq c, \dots, D(X_n) \leq c$ kde c je nějaké konečné číslo, potom pro libovolné

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \varepsilon\right) = 1 \text{ (tedy průměr výsledku dostatečně velkého počtu pokusů}$$

se bude libovolně málo lišit od průměru středních hodnot)

- praktickým důsledkem je možnost odhadu teoretického průměru na základě průměru z dostatečně velkého počtu pokusů (dostatečného počtu pozorování)

Centrální limitní věty

- hovoří o asymptotickém rozdělení

- zabývá se normálním zákonem rozdělení jako zákonem limitním

- a) **Věta Lindeberg-Lévyho** – tvrdí, že součet (a tedy i průměr) vzájemně nezávislých náhodných veličin s konečnými středními hodnotami a rozptyly má pro dost velká n přibližně normální rozdělení

- $X_i (i = 1, 2, \dots, n)$ mají totéž rozdělení se středními hodnotami μ a rozptylem σ^2 , platí tyto limitní vztahy pro $X = X_1 + X_2 + \dots + X_n$

$$P\left(u_1 < \frac{X - n\mu}{\sigma\sqrt{n}} < u_2\right) \rightarrow \int_{u_1}^{u_2} \phi(u) du \dots (tj. \phi(u_2) - \phi(u_1))$$

ϕ - hustota $N(0,1)$ Φ - distribuční funkce normovaného normálního rozdělení

- b) **Věta Moivre-Laplacova** – vyjadřuje konvergenci binomického rozdělení normálnímu rozdělení.

- náhodná veličina X má $Bi(n, p) \Rightarrow \lim_{n \rightarrow \infty} P\left(\frac{X - np}{\sqrt{np(1-p)}} < x\right) = \Phi(x)$ - distribuční

funkce $N(0,1)$

- $\int_{-\infty}^{\infty} \phi(u) du$ - hustota $N(0,1)$

Deskriptivní (popisná statistika)

Příklady: Str. 29 Př.: 1,3,5

Spotřeba Fabie v litrech/100km (30 uzlů)

6,8	8,0	6,4	8,2	5,9	6,8	7,7	5,6	8,3	7,8
7,2	6,5	5,5	6,6	7,4	6,6	7,2	8,6	7,2	7,3
6,9	7,5	6,4	7,0	6,7	5,6	7,2	8,1	6,7	6,5

Min = 5,5

Max = 8,6

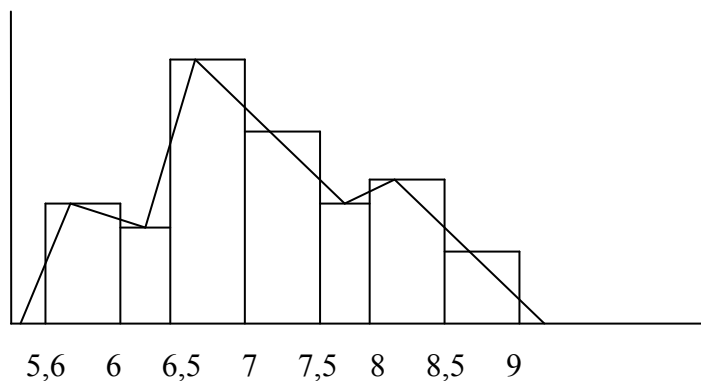
Max-Min=3,1 – výběrové variační rozpětí

Str.22

$$k = 1 + 3,22 \log n \doteq 5,9$$

$$k \doteq \text{int}(2\sqrt{n}) \doteq 10$$

Zvolíme 7 tříd:	Zj	kolikrát je zastoupena	Mj	$ff = \frac{Mj}{n}$
1. <5,5;6,0)	5,75	////	4	0,1333
2. <6,0;6,5)	6,25	//	2	0,0667
3. <6,5;7,0)	6,75	////////	9	0,3000
4. <7,0;7,5)	7,25	////////	7	0,2333
5. <7,5;8,0)	7,75	///	3	0,1000
6. <8,0;8,5)	8,25	///	4	0,1333
7. <8,5;9,0)	8,75	/	1	0,0333



- polygon četností (relativních četností když na ose Y ff) – bodový graf
- histogram četností (relativních četností) – jsou ty obdelníkový graf

1. Aritmetický průměr: $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{30} (6,8 + 8,0 + \dots) \doteq 7,010$ - (výběrový průměr)

Výběrové charakteristiky

a) Polohy

- výběrový průměr $\bar{X} = \sum_{i=1}^n X_i / n$ (aritmetický průměr)
- výběrový modus (kde největší třídň četnost výběru)
- výběrový medián – prostřed v neklesající posloupnosti hodnot výběru (když sudý počet pak průměr z prostředních hodnot)

b) Variability

- výběrový rozptyl $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \rightarrow S = \sqrt{S^2}$ - výběrová směrodatná odchylka
- výběrový rozptyl $S^{*2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ (někde S^{*2}) $\rightarrow S^* = \sqrt{S^{*2}}$ - výběrová směrodatná odchylka nevychýlená
- výběrové variační rozpětí $R_{VAR} = X_{MAX} - X_{MIN}$
- Náhodný výběr – n rozměrnou náhodnou veličinu (vektor) $X = (X_1, X_2, \dots, X_n)$.
 $X_1, X_2, \dots, X_n$ jsou vzájemně nezávislé a všechny mají stejné rozdělení (tj. totéž v distribuční funkci $F(x)$) budeme nazývat náhodným výběrem v rozsahu n z tohoto rozdělení (to je z rozdělení mající distribuční funkci $F(x)$).
- na základě náhodného výběru se snažíme odhadnout rozdělení celého souboru, z kterého výběru vychází
- odhadnu EX aritmetickým průměrem $\bar{x}$ výběru
- odhadnu DX výběrovým rozptylem s^2 výběru
- **Bodový odhad** – když odhaduji statistiky a parametry jedním číslem.
- Odhad metodou momentu -pokud odhaduji pomocí momentů hovoříme
- Odhad maximální věrohodnosti – pokud odhaduji pomocí funkce věrohodnosti
- při bodovém odhadu nevíme nic o chybě toho odhadu, pokud vím (nebo si z výběru zjistím) o jaké jde rozdělení. Mohu tuto informaci využít ke stanovení chyby odhadu $\rightarrow$ (vzniká intervalový odhad (tzv. intervaly spolehlivosti))

Intervaly spolehlivosti pro střední hodnotu normálního rozdělení

- a) **znám δ celého normálního rozdělení (NR)** – využití věty 1.5.2 str. 44, tj. že

náhodná veličina! $Z = \frac{\bar{X} - \mu}{\delta / \sqrt{n}}$ má $N(0,1)$ (když X je z $N(\mu, \delta)$ nebo když $n \geq 30$)

$$P(Z \in (-z_\alpha; z_\alpha)) = 1 - \alpha$$

$$P(-z_\alpha < Z < z_\alpha) = 1 - \alpha$$

$$P\left(-z_\alpha < \frac{\bar{X} - \mu}{\delta / \sqrt{n}} < z_\alpha\right) = 1 - \alpha$$

$$P(-z_\alpha * \delta / \sqrt{n} < \bar{X} - \mu < z_\alpha * \delta / \sqrt{n}) = 1 - \alpha$$

$$P(-\bar{X} - z_\alpha * \delta / \sqrt{n} < -\mu < z_\alpha * \delta / \sqrt{n} - \bar{X}) = 1 - \alpha$$

$$P(\bar{X} + z_\alpha * \delta / \sqrt{n} > \mu > -z_\alpha * \delta / \sqrt{n} + \bar{X}) = 1 - \alpha$$

$$P(\mu \in \langle \bar{X} - z_\alpha * \delta / \sqrt{n}; \bar{X} + z_\alpha * \delta / \sqrt{n} \rangle) = 1 - \alpha$$

-kde typický z_α je $(1 - \alpha) * 100\%$ kvantil $N(0,1)$ oboustranný interval spolehlivosti $(1 - \alpha)$ pro střední hodnotu μ – je takzvaná koeficient spolehlivosti (konfidence) – se volí $1 - \alpha = 0,95$

- b) **neznámá sigma** – věta 1.5.4 str.44, tj. že $T = \frac{\bar{X} - \mu}{S} * \sqrt{n-1}$ má Studentovo

rozdělení s $n-1$ stupni volnosti (když X z $N(\mu, \delta)$ nebo $n \geq 30$)

- obdoba jako v a) $\rightarrow P(\mu \in \langle \bar{X} - t_\alpha S / \sqrt{n-1}; \bar{X} + t_\alpha S / \sqrt{n-1} \rangle) = 1 - \alpha$

Interval spolehlivosti pro rozptyl normálního rozdělení

- využijeme věty 1.5.3, tj. že $\chi = \frac{n * S^2}{\delta^2}$ má χ^2 rozdělený s $n-1$ stupni volnosti, když výběr z NR (celého normálního rozdělení) $\rightarrow$ interval spolehlivosti pro rozptyl

$$I_{1-\alpha} = \left\langle \frac{n * S^2}{\chi_2}; \frac{n * S^2}{\chi_1} \right\rangle, \text{ kde } \chi_2 \text{ je } (1 - \alpha / 2) * 100\% \text{ní je kvantil } \chi^2 \text{ s } n-1 \text{ stupně volnosti}$$

Jednostranné intervaly spolehlivosti

Ad a) pro μ , když znám δ

$I_{1-\alpha} = \left\langle \bar{X} - z_{2\alpha} \delta / \sqrt{n}; +\infty \right\rangle$ - levostranný $(1 - \alpha)$ interval spolehlivosti

$I_{1-\alpha} = \left\langle -\infty; \bar{X} + z_{2\alpha} \delta / \sqrt{n} \right\rangle$ - pravostranný $(1 - \alpha) \times 100\%$ ní interval spolehlivosti pro střední hodnotu

Příklady:

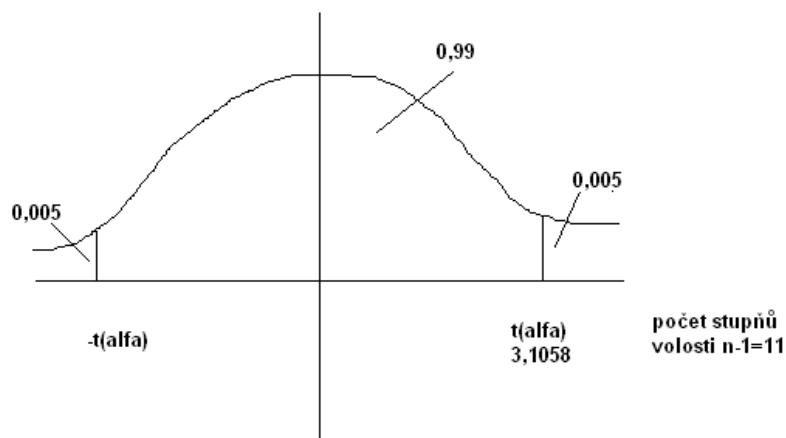
Př. 2.2

Cena vína	40	58	52	45	60	53	55	57	39	42	58	46
-----------	----	----	----	----	----	----	----	----	----	----	----	----

Předp. NR pravdě.

99% interval spolehlivosti pro střední hodnotu NR ceny vín

$n=12$, $\bar{x} = 50,4167$; $s=7,2968$;



$$P(T > t_{\alpha, n}) = \alpha \Rightarrow \alpha = 0,01 \Rightarrow \underline{\underline{t_{0,1,11} = 3,1058}}$$

$$I_{0,99} = \left\langle \bar{x} - \frac{s}{\sqrt{n-1}} t_{\alpha, n-1}; \bar{x} + \frac{s}{\sqrt{n-1}} t_{\alpha, n-1} \right\rangle = \langle 43,5837; 57,32.. \rangle$$

Pevnost – 10 testů (na 10-ti lanech)

[t]	3,2	3,3	3,4	3,0	3,1	3,3	3,6	3,3	3,4	3,2
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

$n = 10$

$\bar{x} = 3,28$

1) $\mu = 3,2$, $\delta = 0,2$

$\alpha = 0,05$

$\alpha = 0,01$

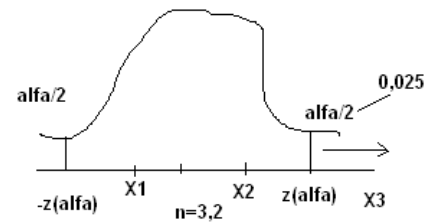
$$Z = \frac{\bar{X} - \mu}{\delta / \sqrt{n}} \text{ má } N(0,1)$$

$H_0 : \mu = 3,2$

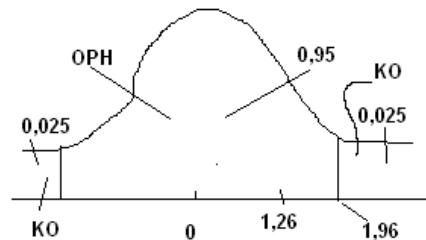
$H_1 : \mu \neq 3,2$

α - hladina významnosti

Realizace: $z = \frac{3,28 - 3,2}{0,2} * \sqrt{10} = \underline{\underline{1,26}}$



OPH - oblast příp. hodnot



Testování hypotéz

- statistická hypotéza je tvrzení o vlastnostech základního souboru, případně více základních souborů, o jehož pravdivosti se chceme přesvědčit, přičemž předem nevíme jestli je pravdivé nebo ne. Vztahuje se buď k tvaru nebo parametrům rozdělení pravděpodobností základního souboru a můžeme ji ověřit testováním.
- **Nulová hypotéza (H0)**
 - o Její platnost ověřujeme
- **Alternativní hypotéza (H1)**
 - o je to co bude platit, když neplatí H0
- říkáme, že testujeme H0 proti H1
- platnosti H0 rozhodujeme na základě zvolené funkce náhodného výběru ($X_1, X_2, \dots$), tato funkce se nazývá testovací kritérium

Kritická oblast (KO)

- je to podmnožina množiny hodnot testovacího kritéria jejichž pravděpodobnost alfa je za předpokladu platnosti hypotézy, tak malá, že náhodný jev „hodnota testovacího kritéria padne do kritické oblasti“ pokládáme za jev nemožný

Oblast přípustných hodnot (OPH)

- je množina hodnot testovaného kritéria, které nepatří do KO

Kritická hranice (KH)

- odděluje kritickou oblast od oblasti přípustných hodnot

Hladina významnosti testu - α

- pravděpodobnost kritické oblasti

Postup pro všechny testy

- 1) vezmu jednu realizaci náhodného výběru ($x_1, x_2, \dots, x_n$)
 - 2) stanovím test
 - 3) zvolím α (z pravidla 0,05)
 - 4) na základě ($x_1, x_2, \dots, x_n$) vypočítám realizaci náhodné vhodné veličiny (kritéria)
 - 5) na základě, kam padne realizace testovací veličiny rozhodnu zda zamítnu H_0 či ne (když padne do KO – zamítnu H_0 ; když nepadne do KO – nezamítnu H_0)
- POZOR! Nezamítnutí H_0 , že platí – může platit ale taky nemusí

Chyba prvního druhu – zamítnu H_0 , když H_0 platí ($= \alpha$)

Chyba druhého druhu – přijmu H_0 , když H_0 neplatí (β)

$(1 - \beta)$ - síla testu

p – hodnota – udává nejnižší možnou hodnotu pro zamítnutí H_0 pro danou realizaci náhodného výběru

Dělení

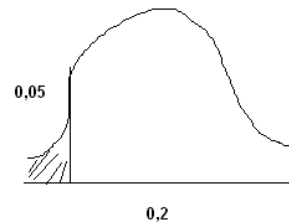
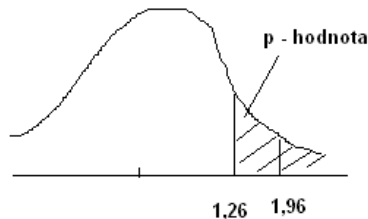
-parametrické testy (testy významnosti) – se týkají parametrů základního souboru

-neparametrické testy (ostatní testy + testy shody) – dvoustranné, jednostranné testy

jednostranné

$H_0: \mu \geq 3,2$

$H_1: \mu < 3,2$



3.2 – Testy významnosti

3.2.1 Jednovýběrový test významnosti pro střední hodnotu normálního rozdělení

- též lze použít pro velké výběry

$x_1, x_2, \dots, x_n$

$H_0: EX = k$

$H_1: EX \neq k$

$(H_0: \mu = k$

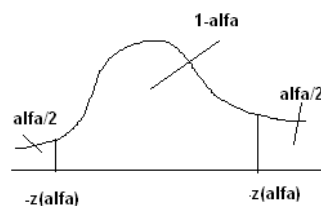
$H_1: \mu \neq k)$

a) známe parametr δ

Testovací kritérium

$$Z = \frac{\bar{X} - k}{\delta} * \sqrt{n} \text{ - má } N(0,1)$$

$$z_\alpha = \Phi^{-1}\left(\frac{2 - \alpha}{2}\right)$$



b) neznáme δ

$$T = \frac{\bar{X} - k}{S} * \sqrt{n-1} \text{ - studentovo rozdělení s } n-1 \text{ stupni}$$

volnosti α

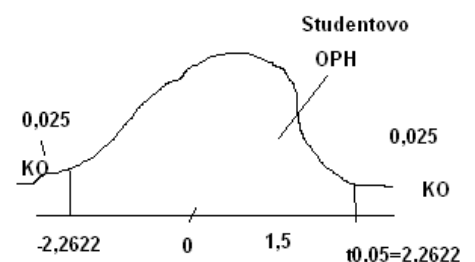
$\Rightarrow$ Vypočítám s z výběru $\Rightarrow s=0,16$, $\alpha=0,05 \Rightarrow$ realizace $T \Rightarrow$

$$t = \frac{\bar{x} - k}{s} * \sqrt{n-1} = \frac{3,28 - 3,2}{0,16} * \sqrt{10-1} = 1,5$$

$-2,2622 \leq 1,5 \leq 2,2622 \Rightarrow t \in OPH$ tedy realizace nepadla do KO

$\Rightarrow$ nezamítám H_0 , a není tedy důvod (na základě našeho výběru) zamítnout hypotézu, že $EX=3,2$ (tedy může být 3,2)

$(I_{0,95} = \langle 3,159; 3,401 \rangle)$



má pravdu (může mít pravdu), že $\delta = 0,2$?

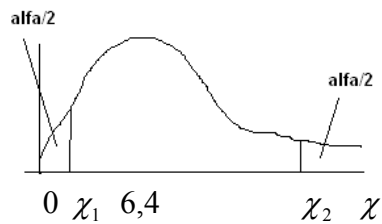
3.2.2 Jednovýběrový test významnosti pro rozptyl normál. rozdělení

$(X_1, X_2, \dots, X_n)$ je z $\underline{NR}(\mu, \delta)$

$$H_0: DX = k^2 \quad H_1: DX \neq k^2$$

Testovací kritérium: dle věty 1.5.3 str. 44

$$\chi = \frac{nS^2}{k^2} \text{ má } \chi - \text{kvadrát rozdělení s } n-1 \text{ stupni volnosti}$$



$$\chi = 0,05 \Rightarrow$$

$$\chi_{0,025;9} = 19,0228 = \chi_2$$

$$\chi_{0,025;9} = 2,70039 = \chi_1$$

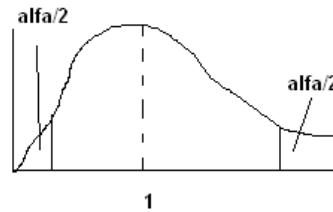
$\chi = \frac{10 * s^2}{0,2^2} = \frac{10 * 0,16^2}{0,2^2} = 6,4 \Rightarrow 2,70059 \leq 6,4 \leq 19,0228 \Rightarrow \text{nezamítáme } H_0 \text{ a tedy dodavatel může mít pravdu.}$

Dvouvýběrový test pro rozptyl

$(X_1, X_2, \dots, X_n) \sim N(\mu_1, \sigma_1^2)$

$(Y_1, Y_2, \dots, Y_n) \sim N(\mu_2, \sigma_2^2)$

$H_0: DX = DY \quad H_1: DX \neq DY$



Testovací kritérium

$F = \frac{\bar{S}_1^2}{\bar{S}_2^2}$ - F-rozdělení s $n_1 - 1$ u $n_2 - 1$ stupni volnosti, hranice $F_{\alpha}, n_1 - 1, n_2 - 1$

V praxi: $F = \frac{\max\{\bar{S}_1^2, \bar{S}_2^2\}^2}{\min\{\bar{S}_1^2, \bar{S}_2^2\}}$

Příklad:

X	101	100,5	100	100	99,7	100,2	100,3	99,6
Y	100	100	99,8	99,9	100	100,3	100,7	100

$n_1 = 8$

$H_0: DX = DY \quad H_1: DX \neq DY$

$n_2 = 8$

$\bar{S}_1^2 = 0,20268$

$\bar{S}_2^2 = 0,08125$

$f = \frac{\max\{\bar{S}_1^2, \bar{S}_2^2\}^2}{\min\{\bar{S}_1^2, \bar{S}_2^2\}} = \frac{0,20268}{0,08125} = 2,495$

Pro f tabulka na straně 18

$\alpha = 0,10$

Hranice: $F_{0,005;8-1;8-1} = 8,885$

$2,495 \leq 8,85 \Rightarrow$ nepadlo do KO $\Rightarrow H_0$ nezamítáme

Dvouvýběrový test pro rovnost středních hodnot

$$H_0: EX = EY \quad H_1: EX \neq EY$$

a) známe δ_1 a δ_2

$$\text{Testovací kritérium: } Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}}} - \text{ má } N(0,1)$$

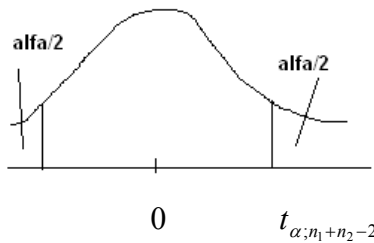
Hledáme obdobně jako Z u jednovýběrových

b) neznáme δ_1 a δ_2 , ale předpokládáme, že $\delta_1 = \delta_2$

$$H_0: EX = EY \quad H_1: EX \neq EY$$

$$\text{Testovací kritérium: } T = \frac{\bar{X} - \bar{Y}}{\sqrt{n_1 S_1^2 + n_2 S_2^2}} * \sqrt{\frac{n_1 * n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} - \text{ Studentovo rozdělení s}$$

$n_1 + n_2 - 2$ stupně volnosti



c) neznáme δ_1 a δ_2 předpokládáme, že $\delta_1 \neq \delta_2$ (zjistím F – testem na rozptyly)

$$H_0: EX = EY \quad H_1: EX \neq EY$$

$$\text{Testovací kritérium: } T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}}}$$

$$KH = \frac{\frac{S_2^2}{n_1 - 1} t_{\alpha; n_1 - 1} + \frac{S_1^2}{n_2 - 1} t_{\alpha; n_2 - 1}}{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}}$$

Příklad:

Středa	726	600	750	700	730	550	595	715		
Pátek	1002	1050	950	823	1250	1410	684	825	923	995

$$\alpha = 0,05$$

Zde odlišné střední hodnoty? (předpoklad NR- normalita)

$$H_0: EX = EY \quad H_1: EX \neq EY$$

$$n_1 = 8 \quad \bar{x} = 670,85 \quad S_1^2 = 5122,69 \quad \bar{S}_1^2 = 5854,5$$

$$n_2 = 10 \quad \bar{x} = 991,2 \quad S_2^2 = 40258,36 \quad \bar{S}_2^2 = 44732,62$$

$$1) ? \delta_1 = \delta_2$$

$$H_0: DX = DY \quad H_1: DX \neq DY$$

$$f = \frac{\max\{\bar{S}_1^2; \bar{S}_2^2\}}{\min\{S_1^2; S_2^2\}} = \frac{44732,62}{5854,5} = 7,641$$

$$F_{0,025;10-1;8-1} = F_{0,025;9;7} = 4,823$$

7,641 > 4,823 $\Rightarrow$ realizace padla do KO $\Rightarrow$ H0 zamítáme a dále předpokládáme, že rozptyly jsou různé

$$\Rightarrow t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}}} = \frac{6,7075 - 991,2}{\sqrt{\frac{5122,69}{7} + \frac{40259,36}{9}}} = -4,4417$$

$$t_{0,025;n_1-1} = 2,3646 \quad -n_1 = 7$$

$$t_{0,025;n_2-1} = 2,2622 \quad -n_2 = 9$$

$$KH = \frac{\frac{S_1^2}{n_1 - 1} t_{\alpha;n_1-1} + \frac{S_2^2}{n_2 - 1} t_{\alpha;n_2-1}}{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}} = \frac{\frac{5122,69}{7} * 2,3646 + \frac{40259,36}{9} * 2,2622}{\frac{5122,69}{7} + \frac{40259,36}{9}} = \underline{\underline{2,2766}}$$

$|-4,44217| > 2,2776 \Rightarrow$ realizace padla do KO $\Rightarrow$ H0 zamítáme $\Rightarrow$ střední hodnota průjezdu ve středu a v pátek jsou stejné

Párový T-test (test pro párové hodnoty závislých výběrů)

$$(X_1; Y_1)(X_2; Y_2) \dots (X_n; Y_n) \quad N(\mu_1; \mu_2; \sigma_1; \sigma_2)$$

Hypotéza $EX - EY = k$ převádíme na hypotézu $ED=k$, kde ED je střední hodnota příslušných rozdílů $H_0: ED = k \quad H_1: ED \neq k$

Testovací kritérium:

$$T = \frac{\bar{D} - k}{S_d} * \sqrt{n-1}, \quad \bar{D} = \frac{1}{n} \sum_{i=1}^n D_i, \quad S_d = \sqrt{\frac{1}{n} \sum_{i=1}^n (D_i - \bar{D})^2}$$

KH: studentovo, $n-1$ st. volnosti, n_2-1

Č	1	2	3	4	5	6	7	8
Před	3,19	2,94	2,80	3,00	3,23	3,29	2,95	2,87
Po	3,25	2,95	2,97	3,13	3,24	3,24	3,07	2,77
Di	-0,06	-0,1	-0,017	-0,13	-0,01	0,05	-0,12	0,1

$$\bar{d} = -0,04375$$

$$S_d = 0,08745$$

1. ? Zda se úpravou změní tvrdost $k=0$

$$H_0: ED = 0 \quad H_1: ED \neq 0$$

$$T = \frac{\bar{D} - k}{S_d} * \sqrt{n-1} = \frac{-0,04375}{0,08745} * \sqrt{8-1} = -1,3236$$

$$t_{0,01;7} = 3,495$$

$|-1,3236| \leq 3,4995 \Rightarrow H_0$ nezamítáme a tedy nelze tvrdit, že proces má vliv na tvrdost

χ^2 -testy

1. Test shody

2. Test nezávislosti (z kontingenční tabulky)

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

O_i – pozorované četnosti
 E_i – očekávané (teoretické) četnosti

Příklad:

$$? NR(\mu = 7; \sigma = 0,8) \quad N(0,1) \quad Z = \left(\frac{x - \mu}{\sigma} \right) \left(-1,875 = \frac{5,5 - 7}{0,8} \right) \quad 0,1056 = \Phi(-1,25) - \Phi(-1,875)$$

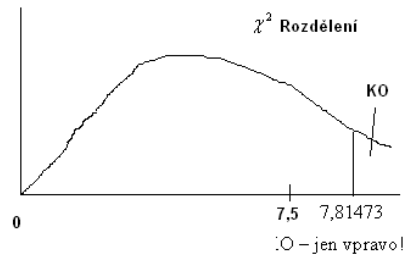
			Epi	-Enp	$\chi^2 = (O_i - E_i)^2$
1. < 5,5; 6,0)	4	< -1,875; -1,25)	0,1056	3 , 168	$\frac{(6-8,028)^2}{8,028} = 0,5123$
2. < 6,0; 6,5)	2	< -1,25; -0,625)	0,1620	4 , 86	
3. < 6,5; 7,0)	9	< -0,625; 0)	0,2324	6 , 972	$\frac{(9-6,975)^2}{6,972} = 0,59$
4. < 7,0; 7,5)	7	< 0; 0,625)	0,2324	6 , 972	
5. < 7,5; 8,0)	3	< 0,625; 1,25)	0,1620	4 , 86	$\frac{(7-6,972)^2}{6,972} = 0,00$
6. < 8,0; 8,5)	4	< 1,25; 1,875)	0,0749	2 , 247	
7. < 8,5; 9,0)	1	< 1,875; 2,25)	0,0307	0 , 921	

$$\overline{\quad} = 30$$

$$\overline{\quad} = 1,00000$$

-tento příklad není dokončen!!!

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$



z volání -	Č	Z	B	M
	40%	20%	20%	20%
O _i	30	30	20	20
E _i	40	20	20	20

$$\chi^2 = \frac{(30 - 40)^2}{40} + \frac{(30 - 20)^2}{20} + \frac{(20 - 20)^2}{20} + \frac{(20 - 20)^2}{20} = \frac{100}{40} + \frac{100}{20} = 2,5 + 5 = 7,5$$

$$\alpha = 0,05$$

Stupně volnosti: $r-k-1=3$ (víme, že $r=4$, $k=0$) $\chi_{0,05;3} = 7,81473$ - najdeme v tabulkách

$7,5 < 7,8147 \Rightarrow H_0$ nezamítám, vedoucí může mít pravdu

Test nezávislosti (z kontingenční tabulky)

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Příklad 3.4.7

Dokončené vzdělání	Změna životní úrovně			
	Zlepší	Nezlepší	Zhorší	
Základní	48	82	70	200
Střední	42	63	35	140
Vysoké	34	31	35	100
	124	176	140	440

$\frac{124 - 200}{440}$	$\frac{176 - 200}{440}$	$\frac{140 - 200}{440}$
$\frac{124 - 140}{440}$	$\frac{176 - 140}{440}$	$\frac{176 - 140}{440}$
$\frac{124 - 100}{440}$	$\frac{176 - 100}{440}$	$\frac{176 - 100}{440}$

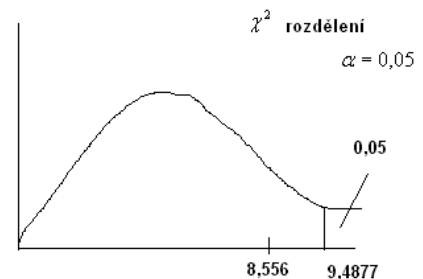
56,36	80	63,63
39,45	56	44,54
38,18	40	31,81

$$\chi^2 = \frac{(40 - 56,36)^2}{56,36} + \frac{(82 - 80)^2}{80} + \frac{(70 - 63,63)^2}{63,63} + \frac{(42 - 34,45)^2}{34,45} + \dots = \underline{\underline{8,556}}$$

Počet stupňů volnosti: $(z-1)*(s-1) = (3-1)*(3-1) = \underline{\underline{4}}$

$$\chi_{0,05;4}^2 = 9,4877$$

$8,556 < 9,4877 \Rightarrow H_0$ nezamítám



Neparametrické testy

$X_1, X_2, \dots, X_n$, spojité, mají F dostav. Funkci

$\tilde{X}$ -medián (mají stejný)

$H_0 : \tilde{X} = c \quad H_1 : \tilde{X} \neq c$

Od všech hodnot odečteme $c \rightarrow$ posunu medián do 0 $\Rightarrow$?medián=0

Příklad 3.4.1

Oktanové číslo testováno

12 vzorků

?medián je 98 na $\alpha = 0,05$

X_i	98,2	96,8	98	96,3	99,8	96,9	95,2	95,6	96,1	97,7	98,1	98,7
-------	------	------	----	------	------	------	------	------	------	------	------	------

$H_0 : \tilde{X} = 98 \quad H_1 : \tilde{X} \neq 98$

$X_i - 98$	0,2	-1,2	0	-1,7	1,8	-1,1	-2,8	-2,4	-1,9	-0,3	0,1	0,7
------------	-----	------	---	------	-----	------	------	------	------	------	-----	-----

Když 0 tak vypustím o 0 sdružíme počet o 1

$Y=4$

$N=12-1=11 \quad KO \leq K1 \text{ nebo } \geq K2$

Tabulky: hrani u $k1:1; K2=10$

$K1 < 4 < K2 \Rightarrow H_0$ nezamítá, medián může být jedna

Wilcoxonový testy

Jednovýběrový Wilcoxonový test

$H_0: F_{(x)} = 1 - F_{(-x)}$ (ne. Distr. Funkce je symetrická)

H_1 : není symetrická (aplikace těž pásový -> jednovýběrový)

	1	2	3	4	5	6	7	8	9	10
Staré	10,7	12,2	11,8	11,9	11,5	10,9	11,3	12,1	12,0	11,8
Nové	10,8	11,9	11,4	11,6	11,6	11,0	11,2	12,0	12,1	11,6
Di	-0,1	+0,3	+0,4	+0,3	-0,1	-0,1	+0,1	+0,1	-0,1	+0,2
	3,5	8,5	10	8,5	3,5	3,5	3,5	3,5	3,5	7

$$\frac{1+2+3+4+5+6}{6} = 3,5 \quad \text{Pořadí z absolutních hodnot } 6 * |0,1| \Rightarrow \text{průměr z pořadí}$$

$$1,2,3,4,5,6=3,5$$

$$S^+ = \sum_{x_1 \geq 0} y^+ = 8,5 + 10 + 8,5 + 3,5 + 3,5 + 7 = 41$$

$$S^- = \sum_{x_2 \leq 0} y^- = 3,5 + 3,5 + 3,5 + 3,5 = 14$$

$$\{S^+, S^-\} = 14$$

$14 > 8 \Rightarrow$ nepadla do $K_0 \Rightarrow H_0$ nezamítáme $\Rightarrow$ nemělo vliv

Dvouvýběrový Wilcoxonův test

$$(X_1, X_2, \dots, X_n)$$

$$(Y_1, Y_2, \dots, Y_m)$$

H_0 : Distribuční funkce jsou stejné

H_1 : Distribuční funkce jsou rozdílné

Příklad: ceny benzínu testovány u stanic v Praze a Brně

Praha	23,50 (12,5)	22,50 (7,5)	22,10 (4 => 3,5)	22,30 (6)	23,10 (10,5)	23,50 (12,5)	23,10 (10,5)	23,60 (23,6)
Brno	22,00 (2)	21,70 (1)	22,10 (3 => 3,5)	22,50 (7,5)	22,60 (9)	22,20 (5)		

?jsou ceny v Praze a Brně stejné (na hladině vyzn. 0,05) – čísla v závorka je pořadí, když jsou stejné, dá se půlka a (7,5 je průměr mezi 7 a 8)

Počet údajů Praha $m=8$

Počet údajů Brno $n=6$

$$t_1 = 12,5 + 7,5 + 3,5 + 6 + 10,5 + 12,5 + 10,5 + 14 = 77$$

$$t_2 = 1 + 2 + 3,5 + 5 + 7,5 + 9 = 28$$

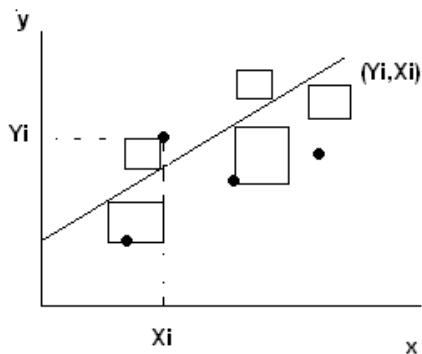
$$u_1 = m * n + \frac{m(n+1)}{2} = 48 + 72 / 2 - 77 = 7$$

$$u_2 = m * n + \frac{m(n+1)}{1} = \dots \text{nevím}$$

Regresní analýza

Regresní funkce - $E(Y|x)$

- střední hodnota podmíněného hodnocení veličiny Y závisí na volbě podmínky (v hodnotě veličiny x) a je tedy její funkcí. Tato funkce se nazývá funkcí regresní veličiny Y vzhledem k X
- $E(Y|X) = \int_{-\infty}^{\infty} yf(y|x)dy$ - jen funkce x
- Základem regresní analýzy je na základě náhodného výběru odhadnout regresní funkci nebo testovat hypotézy o regresní funkci nebo o parametru této funkce
- Parametry regresní funkce nazýváme regresní parametry nebo též regresní koeficienty
- Odhad parametrů mám pak umožní předpovědět veličinu Y pro nějakou danou hodnotu x
- Základním modelem regresní analýzy je model, kde proměnné $x_1, x_2, \dots, x_k$ jsou pevnými (nenáhodnými) proměnnými a náhodné veličiny Y_i mají i rozptyl a jsou nezávislé
- Jednoduchým modelem lineární regrese budeme nazývat model $Y_i = \alpha + \beta x_i + \varepsilon_i$
- ε_i - jsou nezávislé náhodné veličiny pro, které platí střední hodnota všechna $E(\varepsilon_i) = 0$, disperse $D(\varepsilon_i) = \sigma^2$
- ε_i - se nazývá náhodná položka (zahrnuje působení náhodných vlivů)
- Přímka : $y = \alpha + \beta x$ - se nazývá regresní přímka β - její směrnice
- Neznámé parametry α, β, σ^2
- Budeme po řadě značit A, B, S^2
- Bodové odhady α, β získáme metodou nejmenších čtverců



-součet čtverců reziduí (reziduální součet čtverců) ESS (error sum of squares)

$$-? \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \min$$

$$\begin{aligned} ESS &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - (A + BX_i))^2 = \sum_{i=1}^n (Y_i - A - BX_i)^2 = \sum_{i=1}^n (X_i^2)^2 = \\ &= \sum_{i=1}^n (Y_i^2 - 2AY_i - 2BX_iY_i + A^2 + 2ABX_i + B^2X_i^2) \\ \frac{\partial ESS}{\partial A} &= (-2 \sum Y_i) + 2nA + 2B \sum X_i \\ \frac{\partial ESS}{\partial B} &= (-2 \sum X_iY_i) + 2A \sum X_i + 2B \sum X_i^2 \end{aligned}$$

$$\frac{\partial ESS}{\partial A} = 0 \wedge \frac{\partial ESS}{\partial B} = 0 - \text{soustava normálních rovnic}$$

Řešením normálních rovnic :

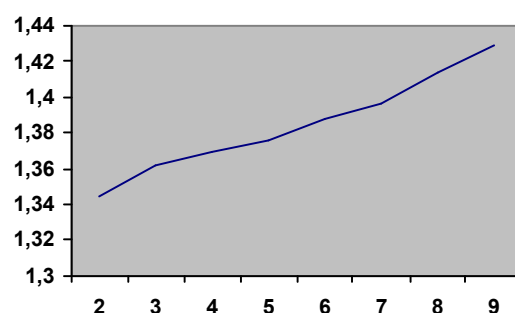
$$B = \frac{n \left(\sum_{i=1}^n X_i Y_i \right) - \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right)}{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2}$$

$$A = \bar{Y} - B\bar{X} \quad \left(A = \frac{1}{n} \sum_{i=1}^n Y_i - B \frac{1}{n} \sum_{i=1}^n X_i \right) \Rightarrow (\bar{Y}, \bar{X}) \in \text{na regresní přímce!!}$$

- dá se ukázat, že ESS je minimum (dle druhých parciálních derivací větších než nula)
- A,B paramet. α, β
- Nevychýlená (tj. $E(B) = \beta$ a $E(A) = \alpha$)

Příklad: 4.2.1. str. 113

Vzorek	Koncentrace v %	Index lomu
1	0	1,3329
2	10	1,3440
3	25	1,3612
4	33	1,3693
5	40	1,3761
6	50	1,3881
7	60	1,3970
8	80	1,4142
9	100	1,4291



$$y(x) = \alpha + \beta x$$

$$1) b = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2} = 0,000973 \quad \underline{\underline{\hat{y} = 1,3361 + 0,000973x}}$$

$$a = \frac{1}{n} \sum_{i=1}^n Y_i - \frac{b}{n} \sum_{i=1}^n X_i = 1,3361$$

2) Pro koncentraci 76,2%

$$y = 1,3361 + 0,000973 * 76,2 = \underline{\underline{1,4102426}}$$

3) určete „vyrovnanou“ hodnotu indexu lomu pro koncentraci 60%

$$60\% \quad y = 1,3361 + 0,000973 * 60 = \underline{\underline{1,39448}}$$

Jak charakterizovat variabilitu Y?

$$S_y = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad \text{- charakterizuje celkový rozptyl – celkový součet čtverců odchylek}$$

Nás nejvíce zajímá součet čtverců odchylek od regresní přímky jako charakteristika rozptylu kolem regresní přímky

Za předpokladu ε_i mají navíc (k nezávislosti a stejnému rozptylu a $E(\varepsilon_i) = 0$) normální rozdělení je maximálně věrohodným odhadem parametru σ^2 statistika

$$\hat{\sigma}^2 = S^2_{rez} = \frac{1}{n-2} \sum_{i=1}^n (Y_i - A - Bx_i)^2 - \text{tzv. reziduální rozptyl}$$

Kde $\sum_{i=1}^n (Y_i - A - Bx_i)^2$ - je reziduální součet čtverců Se - vysvětluje část celkové variability,

která je způsobena náhodnými odchylkami

$$\left(Se = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - A - Bx_i)^2 \right)$$

Zbytek – tzv. vysvětlitelný regresní součet čtverců odchylek St

$$St = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^n (A + Bx_i - \bar{Y})^2 \quad Sy = St + Se$$

$$Sy = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Podíl vysvětlené části rozptylu celkovému rozptylu vyjadřuje index determinace $I^2 = \frac{St}{Sy}$

(pozor $1 - I^2$ = koeficient determinace a udává jaká část není vysvětlena regresním modelem“)

Index determinace určuje jakou část variability lze vysvětlit daným modelem, nabývá hodnot $\langle 0, 1 \rangle$

Např. $I^2 = 0,9$ znamená, že 90% variability lze vysvětlit regresním modelem a zbylých 10% je vlivem náhodného kolísání

Test B

$$H_0: \beta = \beta_0 \quad H_1: \beta = \beta_1$$

$$T = \frac{\beta - \beta_1}{S_{mez}} \sqrt{\sum (x_i - \bar{x})^2} \quad \text{- studentovo s n-2 stupni volnosti}$$

$$\alpha: H_0 = \alpha = \alpha_0 \quad H_1: \alpha \neq \alpha_0$$

$$H_0: \alpha + \beta x = y_0 \quad H_1: \alpha + \beta x \neq y_0$$

Linearizace regresních funkcí

- je to transformací -> převedu na lineární funkci

$$\rightarrow \ln y = \ln a + b * \ln x$$

- např. $y = ax^b$ zlogaritmováním $y' = \ln y, \dots, x'_i = \ln x_i$

$$\underline{\underline{y' = a' + bx'}}$$

Korelační analýza

-analýza těsnosti vazby mezi veličinami

$$- X, Y \quad \rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sqrt{DX} \sqrt{DY}}$$

Výběrový koeficient korelace

- nechť je dán dvojrozměrný náhodný výběr $((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$

$$- R_{X,Y} = \frac{\overline{\text{cov}(X, Y)}}{S_X S_Y} \quad \overline{\text{cov}(X, Y)} = \frac{1}{n} (X_i - \bar{X})(Y_i - \bar{Y}) \quad \text{- výběrová kovariance}$$

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

$$- S_Y = \sqrt{\frac{1}{n} (Y_i - \bar{Y})^2}$$

Vlastnosti $R_{X,Y}$:

1. $R_{X,Y} \in \langle -1, 1 \rangle$
2. $R_{X,Y} = R_{Y,X}$
3. $R_a X + b, cY + d = R_{X,Y} \quad \text{pro } ac > 0$
4. $R_a X + b, cY + d = R_{X,Y} \quad \text{pro } ac < 0$

ρ test

- $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ ze zákl. souboru (X, Y)

- které mají $N(\mu_1, \mu_2; \delta_1, \delta_2)$!

- $H_0: \rho = 0$

$H_1: \rho \neq 0$

Nekontrované – není lineární závislost kontrované – je lineární závislost

Za předpokladu $NR \wedge \rho = 0$

$$T = \frac{R}{\sqrt{1-R^2}} * \sqrt{n-2} \quad \text{má studentovo rozdělení s n-2 stupně volnosti}$$

Když nevím zda X,Y mají NR, nebo chci obecně jinou záv. Než lineární (monotonní funkce)
➔ aplikujeme Spearmanův test (vychází z pořadí hodnot x_i, y_i porovnání)

R_s - spearmanův korelační koeficient

$$R_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i - Q_i)^2$$

R_i – pořadí jednoho výběru, Q_i – pořadí druhého výběru

H_0 : není korelace H_1 : je korelace

RS=.....

V tabulkách kritické hranice pro různá α

Příklad 5.4.3

X_i (pořadí v NHL)	1	2	3	4	5	6	7	8
Y_i (příjem)	3,75	2,5	1,9	3,3	2,15	2,0	3,0	2,88
$(R_i - Q_i)^2$	8	4	1	7	3	2	6	5

$n=8$

$$R_s = 1 - \frac{6}{8(64 - 1)} * (49 + 4 + 4 + 9 + 4 + 16 + 1 + 9) = -0,143$$

$$\alpha = 0,05 \quad R_{0,05} = 0,905$$

$|-0,143| < 0,6905 \Rightarrow$ nezamítáme $H_0 \Rightarrow$ není korelace mezi pořadím a příjmem

Příklad 5.4.1

$\alpha = 0,05$?korelace mezi X,Y

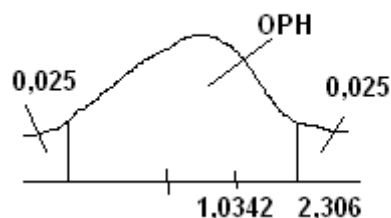
Za předpokladu, že X,Y mají NR (normální rozdělení)

X_i	94	98	127	88	95	111	75	102	82
Y_i	2,1	1,9	3,5	1,5	3,2	1,6	1,9	2,5	1,9

$$H_0: \rho = 0 \quad H_1: \rho \neq 0$$

$$T = \frac{R}{\sqrt{1-R^2}} * \sqrt{n-2} \quad \text{má studentovo rozdělení s } n-2 \text{ stupně volnosti}$$

$$r = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sqrt{\sum X_i^2 - n \bar{X}^2} * \sqrt{\sum Y_i^2 - n \bar{Y}^2}} \doteq 0,3425 \quad r = \frac{\frac{1}{n} \sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n} \sum (X_i - \bar{X})^2} * \sqrt{\frac{1}{n} \sum (Y_i - \bar{Y})^2}}$$



$|1,031| < 2,306 \Rightarrow \text{nezamítáme} \Rightarrow \text{veličiny jsou nekorelované (tj. není lineární závislost mezi X,Y)}$

Druhy způsob:

X_i	94	98	127	88	95	111	75	102	82	85
Y_i	2,1	1,9	3,5	1,5	3,2	1,6	1,9	2,5	4	9
	5	7	10	4	6	9	8	2	2	3
$(R_i - Q_i)^2$	1	9	0	9	16	25	36	0	4	36

$$R_s = 1 - \frac{6}{10(100-1)} * 133 \doteq 0,19394$$

$$R_{0,05|10} = 0,6364 \quad |0,19394| < 0,6364 \Rightarrow H_0 \text{ nezamítáme} \Rightarrow \text{není korelace}$$

Metody odhadu:

- ➔ momentů
- ➔ max. věrohodnosti
- ➔ jiné (např. Baysova metoda)

ϕ - skutečná hodnota parametru

$\hat{\phi}$ - jeho odhad

1. odhad je konzistentní

- $\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P(|\hat{\phi} - \phi| < \varepsilon) = 1$ (tj. s rostoucím n výběru konverguje odhad ke skutečné hodnotě podle pravděpodobnosti)

2. odhad je nevychýlený

- $E(\hat{\phi}) = \phi$

3. odhad je defektivní

- $D(\hat{\phi}) = \min$

Metoda max. věrohodnosti

Funkce věrohodnosti

$$L(x_1, x_2, \dots, x_n; \phi) = \begin{cases} \prod_{i=1}^n p(x_i; \phi) & - x_1, x_2, \dots, x_n - \text{realizace,} \\ \prod_{i=1}^n f(x_i; \phi) & \pi - \text{zkratka pro násobení } \left(\prod_{i=1}^n = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n \right) \end{cases}$$

Maximalizujeme $\underline{L}$ - tedy $\frac{dL}{d\phi} = 0$

$$N(\mu, \delta) \quad f(x) = \frac{1}{\delta \sqrt{2\pi}} * e^{-\frac{(x-\mu)^2}{2\delta^2}} \quad x \in (-\infty, \infty)$$

$$L(x_1, x_2, \dots, x_n; \mu; \delta) = \prod_{i=1}^n \frac{1}{\delta \sqrt{2\pi}} * e^{-\frac{(x_i-\mu)^2}{2\delta^2}}$$

$$\ln L(x_1, x_2, \dots, x_n; \mu; \delta) = \sum_{i=1}^n \left[\ln \frac{1}{\delta \sqrt{2\pi}} - \ln \delta - \frac{(x_i - \mu)^2}{2\delta^2} \ln e \right] - \ln e = 1$$

$$\frac{\partial L}{\partial \mu} = 0 \quad \frac{\partial L}{\partial \delta} = 0$$

-vypočítají se tyto dvě parciální derivace