

$$* P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \leftarrow \begin{array}{l} \text{P jevu A} \\ \text{pokud nastane B} \end{array}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad \leftarrow \text{pokud nastane A}$$

## Pravděpodobnost

Klasická definice pravděpodobnosti  $P(A) = \frac{m}{n}$   $m \leq n$

Věta o násobení pravděpodobností  $P(A \cap B) = P(A) \cdot P(B)$   $\leftarrow$  pro nezávislé jevy (mohou nastat současně)

Věta o sčítání pravděpodobností  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$   $\leftarrow$  současně  
 'nebo'  $\leftarrow$  Pro neslučitelné  $P(A \cap B) = 0$

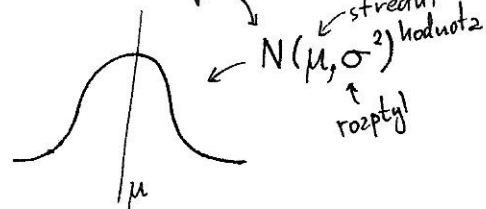
\* Podmíněná pravděpodobnost  $P(A \cap B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B)$

Úplná pravděpodobnost  $P(A) = \sum P(B_i) \cdot P(A/B_i)$   
 jev má teprve nastat

Bayesův vzorec  $\leftarrow$  inverzní pravděpodobnost  $P(B_i/A) = \frac{P(B_i) \cdot P(A/B_i)}{\sum P(B_i) \cdot P(A/B_i)}$   $\leftarrow$  úplná pravděpodobnost  
 pravděpodobnost, že za A může B

Distribuční funkce  $P(X < x) = F(x)$

Normování náhodné veličiny  $U = \frac{X - \mu}{\sigma}$   $\leftarrow$  veličina (třeba  $\bar{x}$ )  
 $\leftarrow$  střední hodnota  $\leftarrow$  směrodatná odchylka



Doplnění k normálnímu rozdělení  $F(-u) = 1 - F(u)$   $P(a < X < b) = F(b) - F(a)$   $\leftarrow$  horní mez - dolní mez  
 $F(\text{záporné}) = 1 - F(\text{kladné})$

výběr s vrácením  $\rightarrow$  Binomické rozdělení  $P(X=k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$   $\leftarrow$   $n$  nad  $k$   $= nCk$   
 pravděpodobnost, že jev nastane k-krát z n pokusů při pravděpodobnosti p  $\leftarrow$  n ... počet pokusů  
 $\leftarrow$  k ... kolikrát chci úspěch  $\leftarrow$  pravděpodobnost úspěchu

Poissonovo rozdělení  $P(X=k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$   $\leftarrow$  k ... kolikrát chci  $\leftarrow$  za časovou (stejnou) jednotku  
 hodně pokusů, malá pravděpodobnost, že nějaký čas (ideál  $n > 30, p \leq 0,1$ )  $\leftarrow$  lambda ... kolikrát v průměru

výběr bez vrácení  $\rightarrow$  Hypergeometrické rozdělení  $P(X=k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$   $\leftarrow$  N ... celkem  
 pravděpodobnost je závislá na předchozích  $\leftarrow$  n ... počet kontrolovaných prvků  
 pokusech (losování míček)  $\leftarrow$  M ... zajímá nás  
 $\leftarrow$  k ... kolik chceme z M

Rovnoměrné rozdělení  $f(x) = \frac{1}{b-a}$   $F(x) = \frac{x-a}{b-a}$   
 konstantní hustota pravděpodobnosti

Exponenciální rozdělení  $f(x) = \lambda e^{-\lambda x}$   $F(x) = 1 - e^{-\lambda x}$   
 dva po sobě jdoucí výskyt jevu  
 $\sim$  rozdělení doby čekání

## Základní statistické charakteristiky

Aritmetický průměr

Forma prostá  $\bar{x} = \frac{\sum x_i}{n}$

$(\sum xy) \div (\sum y)$   
 $\leftarrow$  Forma vážená  $\bar{x} = \frac{\sum x_i n_i}{\sum n_i}$

## Rozptyl

Forma prostá  $s_0^2 = \frac{\sum (x_i - \bar{x})^2}{n}$  *← počet pozorování*  $s_0^2 = \frac{\sum x_i^2 - \bar{x} \sum x_i}{n}$  *← výpočtová forma*  $s_0^2 = \frac{\sum x_i^2}{n} - (\bar{x})^2$  *← nejlepší pro ruční výpočet*

Forma vážená  $s_0^2 = \frac{\sum (x_i - \bar{x})^2 \cdot n_i}{n}$  *← váha,  $n \leq n_i$*   $s_0^2 = \frac{\sum x_i^2 n_i - \bar{x} \sum x_i n_i}{n}$   $s_0^2 = \frac{\sum x_i^2 n_i}{n} - \left( \frac{\sum x_i n_i}{n} \right)^2$  *← spočítat části, pak jen dosadit*

Směrodatná odchylka  
stejně jednotky

$$s_0 = \sqrt{s_0^2} \quad \sigma = \sqrt{\sigma^2}$$

Variační koeficient  $v = \frac{s_0}{\bar{x}} \cdot 100$  [%]  
pro porovnávání variability  
veličin s různými jednotkami  
(převede je na procenta)

	VS	Bodový odhad	ZS
rozptyl	$s_0^2$	$s^2$	$\sigma^2$
směr. odchylka	$s_0$	$s$	$\sigma$
průměr	$\bar{x}$	$\bar{x}$	$\mu$
relativní četnost	$f_i$	$f_i$	$p$

## Teorie odhadu

Bodový odhad rozptylu  
jedno číslo - odhad parametru ZS

výběr s vrácením  $s^2 = s_0^2 \cdot \frac{n}{n-1}$  *← Besselova oprava*  
*↑ neznáme N* *↑ počet vzorků*

výběr bez vrácení  $s^2 = s_0^2 \cdot \frac{n}{n-1} \cdot \frac{N-1}{N}$  *← Besselova oprava*  
*↑ konečnostní násobitel*

Intervalový odhad průměru

$P(\bar{x} - \Delta < \mu < \bar{x} + \Delta) = 1 - \alpha$  *← obvykle 0,05 při 95%  
→  $\langle \bar{x} - \Delta; \bar{x} + \Delta \rangle$*

a) Známe-li rozptyl ZS: (nebo směrodatnou odchylku ZS)

$u_\alpha$  - kritická hodnota normálního rozdělení

výběr s vrácením  $\Delta = u_\alpha \cdot \sqrt{\frac{\sigma^2}{n}}$  *← delta, chyba*  
*↑ neznáme N* *↑ z tabulek*  
určení rozsahu výběru  $n = \frac{u_\alpha^2 \cdot \sigma^2}{\Delta^2}$

výběr bez vrácení  $\Delta = u_\alpha \cdot \sqrt{\frac{\sigma^2}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$  *← známe N*

$n = \frac{u_\alpha^2 \cdot \sigma^2 \cdot N}{\Delta^2 \cdot (N-1) + u_\alpha^2 \cdot \sigma^2}$  *[i] N = celkový počet  
n = vybraný počet*

určení spolehlivosti odhadu  $(u_\alpha) = \sqrt{\frac{n \cdot \Delta^2}{\sigma^2}}$  *jeu kritická hodnota*  
viz. dole

$u_\alpha = \sqrt{\frac{n \cdot (N-1) \cdot \Delta^2}{\sigma^2 \cdot (N-n)}}$

(ani směrodatnou odchylku ZS)

b) Neznáme-li rozptyl ZS: pokud  $n > 100$ , ber hodnoty  $u_\alpha$  z normálního rozdělení

$t_\alpha$  - kritická hodnota Studentova t-rozdělení pro  $(n-1)$  stupeň volnosti  $[f] = n-1$

výběr s vrácením  $\Delta = t_{\alpha(n-1)} \cdot \sqrt{\frac{s^2}{n}}$  *← delta, chyba*  
*↑ po Besselově opravě*  
*↑ z tabulek*

výběr bez vrácení  $\Delta = t_{\alpha(n-1)} \cdot \sqrt{\frac{s^2}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

$\langle \bar{x} - \Delta; \bar{x} + \Delta \rangle$   
*↑ interval*  
určení rozsahu výběru

$n = \frac{t_\alpha^2 \cdot s^2}{\Delta^2}$  *← chyba<sup>2</sup>*

$n = \frac{t_\alpha^2 \cdot s^2 \cdot N}{\Delta^2 \cdot (N-1) + t_\alpha^2 \cdot s^2}$

*[i] Pokud známe N,  
použít vzorce vpravo*

$2F(t_\alpha) - 1$  *jeu kritická hodnota*

určení spolehlivosti odhadu  $(t_\alpha) = \sqrt{\frac{n \cdot \Delta^2}{s^2}}$  *← chyba*  
*↑ po Besselově opravě*

$t_\alpha = \sqrt{\frac{n \cdot (N-1) \cdot \Delta^2}{s^2 \cdot (N-n)}}$

$P(-t_\alpha \leq \mu \leq t_\alpha) = F(t_\alpha) - F(-t_\alpha) = F(t_\alpha) - (1 - F(t_\alpha)) = F(t_\alpha) - 1 + F(t_\alpha) = 2 \cdot F(t_\alpha) - 1$

$F(t_\alpha)$  z tabulek [distribuční funkce normálního rozdělení]

↑ zaokrouhlit na 2 desetinná místa! [řádek - desetiny, sloupec - setiny]  
př.: 0.1 0.02 → 0.12

[i]  $f_i$ : relativní četnost =  $\frac{m}{n}$   
 • kolikrát se něco vyskytlo v %

### Intervalový odhad rozptylu

$\chi^2_\alpha$  – kritická hodnota  $\chi^2$  – rozdělení pro  $(n-1)$  stupeň volnosti  
 ↗ chí kvadrát

oboustranný interval

$$P\left(\frac{(n-1) \cdot s^2}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1) \cdot s^2}{\chi^2_{1-\alpha/2}(n-1)}\right) = 1 - \alpha$$

Pokud hledáme jen ~~na~~ jednu stranu intervalu  $u_\alpha$  se mění na  $u_{2\alpha}$

e.g.  
 $u_\alpha \Rightarrow u_{2\alpha}$  ;  $\alpha = 0.05$   
 $2\alpha = 0.1$

$f_i$   
 Intervalový odhad relativní četnosti  $P(f_i - \Delta < p < f_i + \Delta) = 1 - \alpha$

$u_\alpha$  – kritická hodnota normálního rozdělení

výběr s vracením  $\Delta = u_\alpha \cdot \sqrt{\frac{f_i \cdot (1-f_i)}{n}}$  <sup>2 tabulek</sup>

výběr bez vracení  $\Delta = u_\alpha \cdot \sqrt{\frac{f_i \cdot (1-f_i)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

určení rozsahu výběru  $n = \frac{u_\alpha^2 \cdot f_i(1-f_i)}{\Delta^2}$  <sup>chyba</sup>

$n = \frac{u_\alpha^2 \cdot f_i(1-f_i) \cdot N}{\Delta^2 \cdot (N-1) + u_\alpha^2 \cdot f_i(1-f_i)}$

určení spolehlivosti odhadu  $u_\alpha = \sqrt{\frac{n \cdot \Delta^2}{f_i(1-f_i)}}$  <sup>kritická hodnota</sup>  
 $2F(u_\alpha) - 1$ , viz minulá strana

OBVYKLE NENÍ VÝSLEDEK →  $u_\alpha = \sqrt{\frac{n \cdot (N-1) \cdot \Delta^2}{f_i(1-f_i) \cdot (N-n)}}$  <sup>to ono</sup>

### Testování statistických hypotéz

#### Jednovýběrové testy

jeden soubor, pokud není určeno jinak,  $\alpha$  stanovujeme sami (třeba 0,05; ale všude stejně!)

Test hypotézy o hodnotě průměru

známe-li rozptyl ZS:  $u = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$  <sup>průměr ZS to s čím porovnáváme (normy)</sup>

neznáme-li rozptyl ZS:  $t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{n}}}$  <sup>OBVYKLE neznáme Bessel!</sup>

kritická hodnota:  $u$  – normální rozdělení,  $t$  – Studentovo t-rozdělení pro  $(n-1)$  stupeň volnosti

15 vzorků →  $t_\alpha$  při  $f=14$

Test hypotézy o hodnotě rozptylu

$\chi^2 = \frac{(n-1) \cdot s^2}{\sigma_0^2}$  <sup>Bessel</sup>

kritická hodnota:  $\chi^2$ -rozdělení o  $(n-1)$  stupni volnosti

Test hypotézy o hodnotě relativní četnosti

$u = \frac{\frac{m}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$   <sup>$m = n_i$  = absolutní četnost  
 $n$  ... celkový počet  
 $p_0$  ... četnost se kterou to porovnáváme  
 $[27\% \Rightarrow p_0 = 0,27]$</sup>

kritická hodnota:  $u$  – normální rozdělení

procenta [m je f %]  
 $f_i = \frac{m}{n}$  ; je zadáno že první jakosti jsou ze 47 % ? pak  $m = n \cdot 0,47$   
 celkem kusů

← mělo by platit  
 $n > \frac{5}{p_0}$

[ZÁVISLÉ SOUBORY JSOU NA DALŠÍ STRANĚ!] testování stejných pacientů  
testování na stejných plošinách

## Dvouvýběrové testy

### Test hypotézy o shodě dvou rozptylů

$$F = \frac{s_1^2}{s_2^2} \leftarrow \text{vždy ten větší rozptyl}$$

$F < F_\alpha \rightarrow$  shodné rozptyly (a)

$F > F_\alpha \rightarrow$  nestejně rozptyly (b, c)

kritická hodnota:  $F_\alpha$  – rozdělení pro  $(m-1)$  a  $(n-1)$  stupně volnosti

$$F(m-1, n-1)$$

↖ v tabulce je ve sloupečku to s větším rozptylem první řádek 0,05  
druhý 0,01

### Test hypotézy o shodě dvou průměrů

[i] NEJDŘÍVE F TEST ↕

a) dvouvýběrový t – test při shodných rozptylech

$$t = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

průměr prvního      průměr druhého

$$s = \sqrt{\frac{1}{m+n-2} \cdot [(m-1) \cdot s_1^2 + (n-1) \cdot s_2^2]}$$

[m] počet prvního souboru      [n] počet druhého souboru

kritická hodnota: Studentovo t-rozdělení pro  $(m+n-2)$  stupně volnosti

b) t – test při nestejných rozptylech (Welchův test)

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

obě Bessel!

$t < t_{\alpha f} \rightarrow$  přijímáme

kritická hodnota: Studentovo t-rozdělení pro  $f$  stupňů volnosti, kde

$$f = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{\left(\frac{s_1^2}{m}\right)^2}{m-1} + \frac{\left(\frac{s_2^2}{n}\right)^2}{n-1}}$$

$$s_1^2 = s_{01}^2 \cdot \frac{m}{m-1}$$

← počet jednoho

$$s_2^2 = s_{01}^2 \cdot \frac{n}{n-1}$$

← počet druhého

Rozdíl b) a c):

liší se jen v ověření  $t$ ,

❖ Welch si spočítá stupně volnosti  $f$  a dogleď  $t_\alpha$  v tabulkách

❖ B-F si spočítá  $t_\alpha$  a jen porovná

Mám radši c) → na nic se nezapomene

c) t – test při nestejných rozptylech (Behrens – Fisherův test)

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

Bessel u obou

$t < t_\alpha^* \rightarrow$  přijímáme

kritická hodnota: přepočítaná hodnota Studentova t-rozdělení

$$t_\alpha^* = \frac{t_{\alpha(m-1)} \cdot \frac{s_1^2}{m} + t_{\alpha(n-1)} \cdot \frac{s_2^2}{n}}{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

tabulky

# ZA VÍSLÉ VÝBĚRY

d) t - test pro závislé výběry

$$t = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{n}}}$$

← průměr  
diferenciál  
← rozptyl  
diferenciál

$$\bar{d} = \frac{\sum d_i}{n} = \frac{\sum (x_i - y_i)}{n}$$

$$s_d^2 = \frac{1}{n-1} \sum (d_i - \bar{d})^2$$

kritická hodnota: Studentovo t-rozdělení pro (n - 1) stupeň volnosti

Test hypotézy o shodě dvou relativních četností

$$u = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\bar{p}(1-\bar{p}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\bar{p} = \frac{m_1 + m_2}{n_1 + n_2}$$

kritická hodnota: u - normální rozdělení

$$y = a_{yx} + b_{yx} \cdot x_i$$

$$x = a_{xy} + b_{xy} \cdot y_i$$

Soustava normálních rovnic - přímka

$$\begin{cases} na + b \sum x_i = \sum y_i \\ a \sum x_i + b \sum x_i^2 = \sum x_i y_i \end{cases}$$

to samé

Výpočtové vzorce

závislost:  
 $r < 0,4$  nízká  
 $0,4 \leq r \leq 0,7$  středně silná  
 $0,7 \leq r < 0,9$  vysoká (silná)  
 $0,9 \leq r$  velmi vysoká  
 $0 =$  nekorelované,  $|1| =$  vše na jedné přímce

Korelační koeficient

$$r = \frac{\text{cov}(xy)}{s_x \cdot s_y}$$

$$\text{cov}(xy) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2] [n \sum y_i^2 - (\sum y_i)^2]}}$$

Spearmanův koeficient

$$r_s = 1 - \frac{6 \cdot \sum d_i^2}{n(n^2 - 1)}$$

pořadí

"pořadová"  
korelace

Když je zadáno:

	1.	2.	3.
první	1	3	4
druhá	5	2	6
$d_i$	-4	1	-2

→  $\bar{d} = \frac{-4+1+(-2)}{3}$   
 ← první - druhý

vysvětlivky k d:

## K REGRESI:

(přímka)  
 a ... absolutní člen ... kdy to protne y  
 b ... regresní koeficient ... směrnice,  
 ukazatel změny  
 <-1;1> r ... korelační koeficient ...  
 měří těsnost závislosti,  
 o kolik se závislá proměnná změní při změně  
 nezávislé o 1  
 $(100)r^2$  ... koeficient determinace ...  
 z kolika % je vysvětlována proměnná  
 vysvětlena, vysvětlující  
 y ... závislá (vysvětlována)  
 x ... nezávislá (vysvětlující)  
 $r_s$  ... Spearmanův koeficient  
 korelační

## Regresní a korelační analýza

Li Dole o kalkulačce

$$a_{yx} = \bar{y} - b_{yx} \cdot \bar{x}$$

$$a_{xy} = \bar{x} - b_{xy} \cdot \bar{y}$$

$$r = \pm \sqrt{b_{yx} \cdot b_{xy}}$$

stejně znaménko jako  
běha

$$r = b_{yx} \cdot \frac{s_x}{s_y} = b_{xy} \cdot \frac{s_y}{s_x}$$

## CASIO fx-911ES PLUS

1. PŘECHOD DO STAT (obvykle STAT/SD)

MODE + 3: STAT

2. Data (ukončit zadávání pomocí AC)  
 MODE + STAT + 2: A+BX (ostatní mívají vložení dat  
 pod M+)

3. SHIFT + 1 (stat)  
 → 4: Var →  $n, \bar{x}, s_{o1} (s_x), \bar{y}, s_{o2} (s_y)$

→ 5: Reg → 1: A ... absolutní člen  
 2: B ... regresní koeficient  
 3: r ... korelační koeficient