Statistické hodnocení výsledků zkoušek

Petr Misák

misak.p@fce.vutbr.cz

Co je to statistika?

- ***„Statistika je jako bikiny. Odhalí téměř vše, ale to nejdůležitější nám zůstane skryto.“*** (autor neznámý)
- ***„Statistika nuda je, má však cenné údaje..“*** (Zdeněk Svěrák)
- ***„Statistika je nauka, která nám říká jak získat přesné informace z nepřesných čísel.“*** (Jan Hendl)
- ***„Nevěřím statistice, kterou jsem sám nezfalšoval.“*** (Podvržený výrok Winstona Churchilla rozšířil Joseph Goebbels.)
- ***„Statistiky už máme natolik sofistikované, že z nich lze doložit prakticky cokoliv.“*** (Jan Keller)
- ***„Statistické myšlení bude jednoho dne pro zdatného občana právě tak nezbytné, jako je schopnost číst a psát“.*** (H. G. Wells)

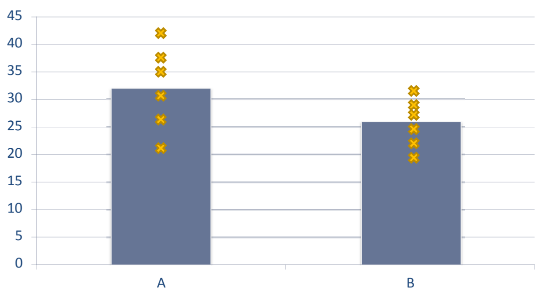
Úvod – statistické myšlení

- Jasné ***vymezení problému***, který má být řešen.
- Stanovení ***rozhodující veličiny*** – jakostní vlastnosti a způsobu jejího zjišťování.
- Zabezpečení ***stálých podmínek*** při jejím zjišťování.
- Uvědomění si, že výsledky měření vykazují jistou (často jen částečně odstranitelnou) ***variabilitu.***
- ***Vytváření podskupin*** homogenních výsledků, zahrnujících pouze náhodnou proměnlivost.
- Respektovat ***náhodné odebírání*** jednotek ***do náhodných výběrů***, tak aby každá jednotka v souboru měla stejnou pravděpodobnost, že může být vybrána do výběru.

Úvod – statistické myšlení

- Studium nejen ***celkové variability***, ale i variability ***uvnitř*** podskupin a variability ***mezi*** podskupinami (v čase).
- Provádění ***dostatečného počtu pozorování***.
- ***Vážení rizik*** chybných závěrů, činěných na základě neúplné informace z náhodných výběrů.
- ***Prezentování dat*** přehledně, ve zhuštěné formě číselně, nebo graficky.
- Charakterizování dat číselně, udáním ***polohy*** na číselné ose a míry ***proměnlivosti*** – variability.
- Uvědomění si nejen variability studované náhodné veličiny, ale i z ní odvozené ***variability vypočítaných statistik*** – výběrových charakteristik.

Úvod – statistické myšlení



Popisná statistika

- Informace obsažené ve velkém počtu dat se jeví lidskému pozorovateli jako nepřehledné.
- Úkolem popisné statistiky je tuto informaci ***zhustit do snadněji vnímatelné formy*** různých tabulek, grafů, číselných a jiných charakteristik.

Popisná statistika

- ***Hromadné jevy*** – jevy, které vznikají za určitých podmínek opakovaně u velkého počtu prvků (*statistických jednotek*)

  Příklad: sériová a hromadná výroba, výsledky laboratorních zkoušek, výsledky kontrol kvality, ekonomické výsledky, vlastnosti lidí.

- ***Statistické jednotky*** – elementární jednotky statistického pozorování

  Příklad: zaměstnanci v podniku, výrobky, poskytované služby, neshodné výrobky, stroje, zařízení, měřidla, lidé, zvířata, věci, události.

Popisná statistika

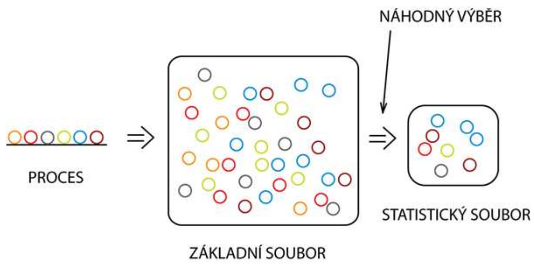
Statistický soubor – množina všech statistických jednotek, u nichž zkoumáme příslušné statistické znaky

Jednorozměrný statistický soubor – u každé statistické jednotky zjišťujeme pouze jeden statistický znak

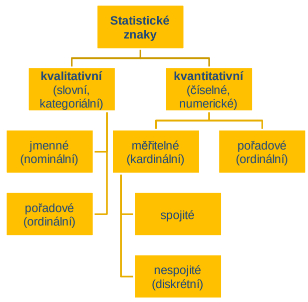
Vícerozměrný statistický soubor – u každé statistické jednotky zjišťujeme dva a více statistických znaků

Základní soubor – statistický soubor všech jednotek, který je předmětem sledování a o němž chceme provádět závěry

Popisná statistika



Popisná statistika



Statistické zkoumání

Statistické zkoumání lze zpravidla rozdělit do tří etap:

1. ***shromažďování dat*** (příprava a sběr)
2. ***zpracování dat***
3. ***rozbor dat*** (vyhodnocení)

Statistické zkoumání – shromažďování dat

1. ***zadání úkolu***
2. ***volba jednotky*** (zkušební místo, část konstrukce,…)
3. ***vymezení souboru*** (kterých jednotek se zkoumání týká)
4. ***určení statistického znaku*** (rozměr, objem, hmotnost, pevnost v tlaku, …)
5. ***způsob měření (hodnocení) znaku*** (kvantitativní, kvalitativní, spojité, …)
6. ***sběr dat*** (kdo a jakým způsobem data zjišťuje a eviduje)

Statistické zkoumání – zpracování dat

1. ***Výpočet***

   popisné statistiky, nástroje matematické statistiky

2. ***Grafické znázornění***

   Grafy dávají rychlou a přehlednou představu jednak o rozložení dat uvnitř souboru a jednak o trendech (časová řada).

Statistické zkoumání – popisná statistika

Třídění – jednorozměrný statistický soubor s kvantitativním znakem

- Uspořádáme data sledovaného kvantitativního znaku do rostoucí posloupnosti.
- Ke každé variantě znaku přiřadíme počty příslušných jednotek, které nazýváme četnosti.
- Hodnoty zaznamenáme do tzv. ***tabulky četností***.

Statistické zkoumání – popisná statistika

Varianta znaku x_i	*Četnost*		*Kumulativní četnost*	
	absolutní f_i	*relativní* f_i / n	*absolutní* F_i	*relativní* F_i / n
x_1	f_1	f_1 / n	$F_1 = f_1$	f_1 / n
x_2	f_2	f_2 / n	$F_2 = f_1 + f_2$	$f_1 / n + f_2 / n$
…	…	…	…	…
x_j	f_j	f_j / n	$F_j = \sum_{k=1}^{j} f_k = n$	$F_j = \sum_{k=1}^{j} \frac{f_k}{n} = 1$
Celkem	$\sum_{k=1}^{j} f_j = n$	$\sum_{k=1}^{j} \frac{f_j}{n} = 1$		

Popisná statistika – charakteristiky polohy

- Určují umístění souboru na číselné ose.
  výběr o rozsahu n: $x_1, x_2, \dots, x_n$

- ***Aritmetický (výběrový) průměr***
  - neroztříděný soubor

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  - roztříděný soubor

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} f_j x_j^*$$

Popisná statistika – charakteristiky polohy

- ***Medián*** – hodnota konkrétní prostřední jednotky statistického souboru
  uspořádání podle velikosti: $x_{(1)}, x_{(2)}, \dots, x_{(n)}$

$$\tilde{x} = \begin{cases} \text{prostřední hodnota; pro } n \text{ liché} \\ \text{průměr dvou prostředních hodnot; pro } n \text{ sudé} \end{cases}$$

- ***Modus***
  - hodnota v jejíž okolí se vyskytuje nejvíce hodnot
  - nejčetnější hodnota souboru

Popisná statistika – charakteristiky variability

- ***Rozptyl***

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2$$

- ***Výběrový rozptyl:***

$$s_0^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \left(\frac{1}{n-1}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2$$

- ***Směrodatná odchylka a výběrová směrodatná odchylka***

$$s = \sqrt{s^2} \qquad s_0 = \sqrt{s_0^2}$$

Popisná statistika – charakteristiky variability

- ***Variační koeficient***

$$V = \frac{s}{\bar{x}}$$

- Jde o relativní míru variability (uvádí se též v %).
- Má smysl pouze pro znak, který nabývá pouze kladných nebo záporných hodnot.

Popisná statistika – charakteristiky souměrnosti

- ***Koeficient šikmosti (asymetrie)***

$$A = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$$

- ukazuje, jak jsou hodnoty kolem aritmetického průměru rozloženy;
- symetrické rozložení má koeficient šikmosti roven nule.

Popisná statistika – charakteristiky souměrnosti

- ***Korelační koeficient***

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s(x)s(y)}$$

- ukazuje míru lineární závislosti dvou veličin
- $-1 \leq r \leq 1$
- $r \approx 0$ - sledované veličiny jsou ***nekorelované***
- $r \approx 1$ - sledované veličiny jsou ***korelované***

Výpočty v MS EXCEL

Výběrový průměr - **PRŮMĚR**(číslo1; číslo2; …)

Výběrový medián - **MEDIAN**(číslo1; číslo2; …)

Výběrový modus - **MODE**(číslo1; číslo2; …)

Směrodatná odchylka stat. souboru - **SMODCH**(číslo1; číslo2; …)

Výběrová směrodatná odchylka - **SMODCH.VÝBĚR**(číslo1; …)

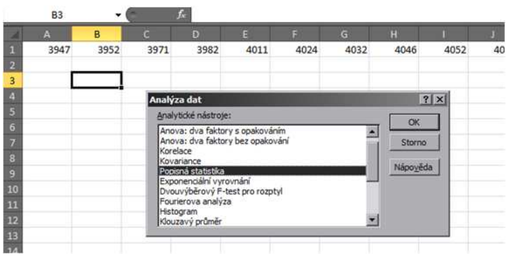
Výběrový rozptyl - **VAR.VÝBĚR**(číslo1; číslo2; …)

Maximální hodnota - **MAX**(číslo1; číslo2; …)

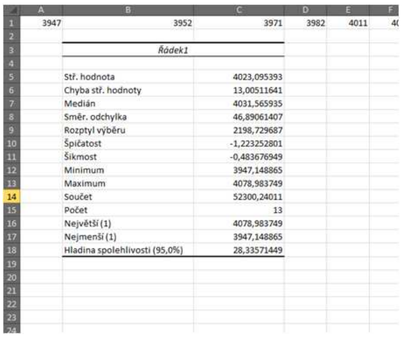
Minimální hodnota - **MIN**(číslo1; číslo2; …)

Počet hodnot - **POČET**(číslo1; číslo2; …)

Výpočty v MS EXCEL



Výpočty v MS EXCEL



	Řádek1
Stř. hodnota	4023,095393
Chyba stř. hodnoty	13,00511641
Medián	4031,565935
Směr. odchylka	46,89061407
Rozptyl výběru	2198,729687
Špičatost	-1,223252801
Šikmost	-0,483676949
Minimum	3947,148865
Maximum	4078,983749
Součet	52300,24011
Počet	13
Největší (1)	4078,983749
Nejmenší (1)	3947,148865
Hladina spolehlivosti (95,0%)	28,33571449

Statistické zkoumání – zpracování dat grafické znázornění

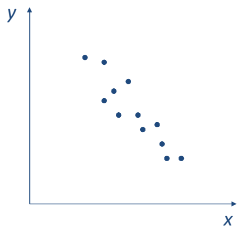
- Bodový graf
- Spojnicový graf
- Histogram
- Výsečový graf (koláč)
- Krabicový graf
- Glyf (radarový graf)

Grafické znázornění - bodový graf

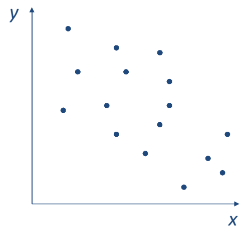
Slouží ke zjištění či ověření ***vzájemné závislosti*** mezi dvěma kvantitativními znaky

Nahrazuje výpočty korelačních koeficientů v případech, kdy chceme získat o případné závislosti pouze orientační informaci.

Grafické znázornění - bodový graf

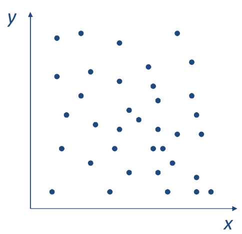


Silná záporná závislost

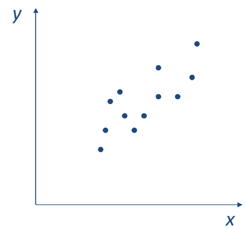


Slabá záporná závislost

Grafické znázornění - bodový graf



Nezávislost

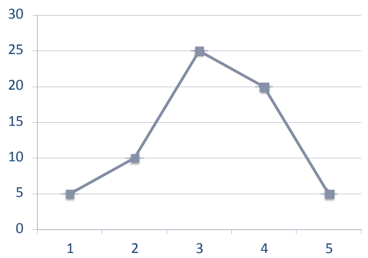


Silná kladná závislost

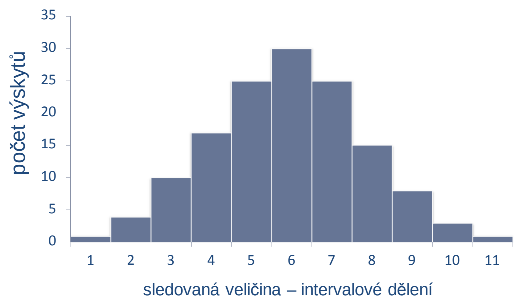
Grafické znázornění – spojnicový graf

slouží k prostému znázornění četností

polygon četností



Grafické znázornění dat - histogram



Grafické znázornění dat - histogram

Příklad:

Rychlost prostupu ultrazvukových vln

Číslo měření	1	2	3	4	5	6	7	8	9	10
Rychlost [m/s]	4079	4011	4117	4189	4184	4221	3947	4084	4177	4024
Číslo měření	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Rychlost [m/s]	4052	3982	3971	4032	4046	3952	4060	4072	4109	4074

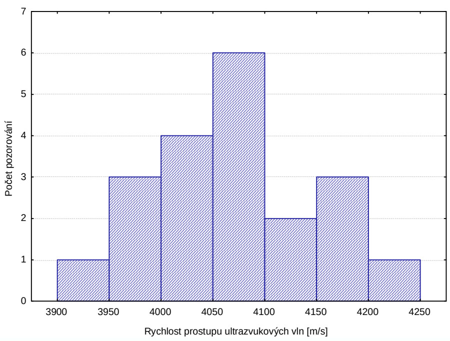
Grafické znázornění dat - histogram

1. Seřazení podle velikosti
2. Výpočet rozpětí

   $R = x_{max} - x_{min} = 274\ m/s$
3. Výpočet délky třídy

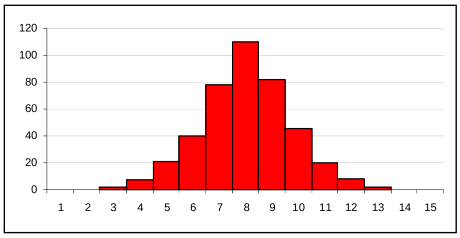
   $h = R/počet\_tříd = 274/7 = 39{,}14$
4. Sestavení tříd

Grafické znázornění dat - histogram



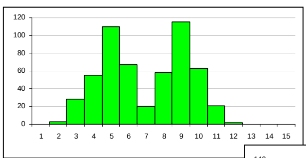
Grafické znázornění dat - histogram

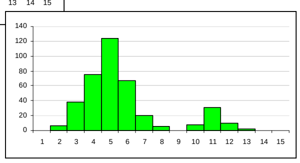
a) Symetrický histogram zvonovitého tvaru



Grafické znázornění dat - histogram

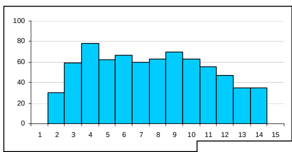
b) Dvojvrcholové histogramy

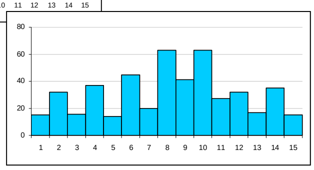




Grafické znázornění dat - histogram

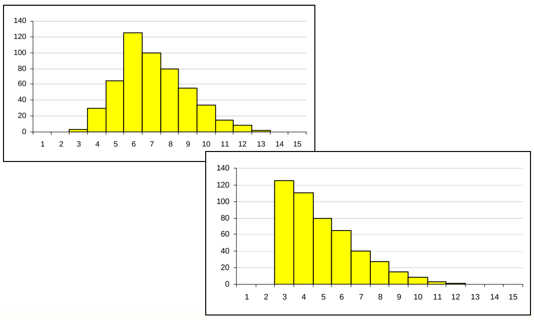
c) Histogramy plochého a hřebenovitého tvaru



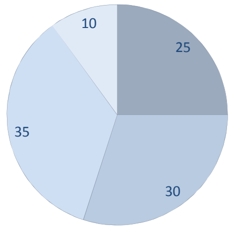


Grafické znázornění dat - histogram

d) Histogramy asymetrického tvaru

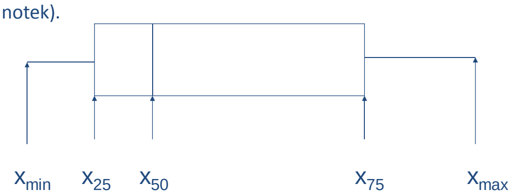


Grafické znázornění dat – výsečový graf

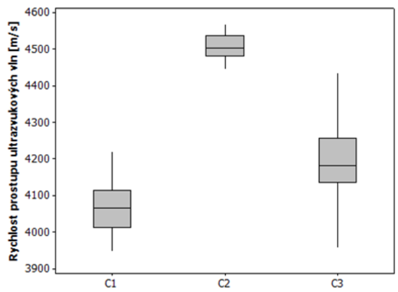


Grafické znázornění dat – krabicový graf

Slouží ke znázornění extrémních hodnot a kvartilů (kvartily dělí statistický soubor na 4 části přičemž každá část obsahuje 25% jednotek).

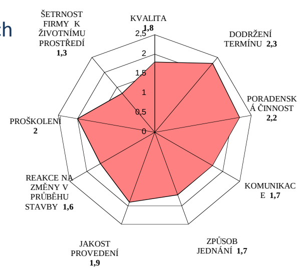


Grafické znázornění dat – krabicový graf



Grafické znázornění dat – glyf

Slouží k určení vzájemných souvislostí dvou či více objektů.



Matematická statistika

- sběr údajů, jejich popis a analýzu
- rozšíření platnosti závěrů z malého počtu vzorků na soubor, z něhož vzorky pocházejí
- zpracování a vyhodnocování informací o realitě, která není známá

Matematická statistika

Věrohodnost závěrů analýzy vyžaduje, aby:

- výrobní dávky byly vyrobeny za ***stejných podmínek,***
- ***podmínky pokusu*** byly specifikovány předem a byly dodržovány během celého pokusu,
- vzorky byly odebrány ***náhodně a byly reprezentativní*** pro soubor, z něhož jsou odebrány.

Matematická statistika

- ***Náhodný pokus –*** je takový pokus, který může dávat různé výsledky i při dodržení stejných podmínek
- ***Náhodný jev –*** je tvrzení o výsledku náhodného pokusu, o kterém lze po jeho uskutečnění jednoznačně rozhodnout, zda je či není pravdivé.
- ***Pravděpodobnost*** – míra nastoupení náhodného jevu

Matematická statistika

- Výsledky náhodného pokusu (realizace náhodné veličiny) tedy ani realizace náhodného jevu nelze s jistotou předpovědět.
- ***Náhodná veličina*** – *X* je reálná proměnná, která nabývá náhodně reálných číselných hodnot *x.*
  - *spojitá*
  - *diskrétní*

Matematická statistika

- Náhodná veličina je jednoznačně určena svou ***distribuční funkcí:***

$$F(x) = P(X < x)$$

- Distribuční funkce určuje tzv. ***rozdělení pravděpodobnosti*** náhodné veličiny
  - *spojitá* náhodná veličina – spojité rozdělení pravděpodobnosti
  - *diskrétní* náhodná veličina – diskrétní rozdělení pravděpodobnosti

Matematická statistika

Spojitá náhodná veličina:

hustota pravděpodobnosti $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$

Vlastnosti:

1. $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$
2. $f(x) = F`(x)$
3. $P(a \le X \le b) = P(a < X < b) = P(a \le X < b) = P(a < X \le b) =$
   $= \int_{a}^{b} f(x)\,\mathrm{d}x = F(b) - F(a)$
4. $P(X = c) = 0$

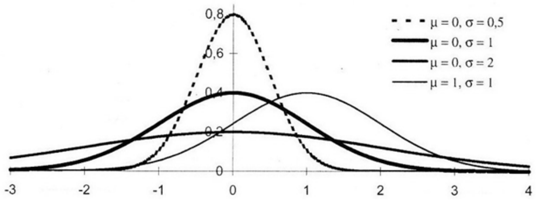
Matematická statistika

- Diskrétní rozdělení pravděpodobnosti
  - Binomické rozdělení – náhodný výběr s vracením
  - Hypergeometrické rozdělení – náhodný výběr bez vracení
  - Poissonovo rozdělení
- Spojité rozdělení pravděpodobnosti
  - Rovnoměrné rozdělení
  - Normální rozdělení

    funkční charakteristiky: střední hodnota μ

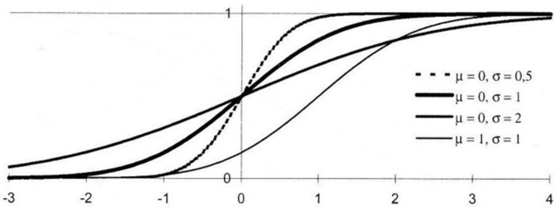
    směrodatná odchylka σ
  - Studentovo rozdělení (*t* – rozdělení)

Matematická statistika – hustota pravděpodobnosti



Matematická statistika – distribuční funkce

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$



Matematická statistika

- ***Kantil***

je hodnota, která rozděluje soubor hodnot určitého statistického znaku na dvě části, jedna obsahuje ty hodnoty, které jsou menší (nebo stejné) než tento kvantil, a druhá část naopak obsahuje hodnoty, které jsou větší (nebo stejné) než kvantil.

X – spojitá náhodná veličina s distribuční funkcí F(x)

její ***P-kvantil*** (*P*100%* kvantil) je číslo x_P , pro které platí:

$$P = F(x_P)$$

Matematická statistika

Používají se tyto kvantily:

- medián (prostřední kvantil): $x_{0,5}$
- dolní kvartil: $x_{0,25}$
- horní kvartil: $x_{0,75}$
- decily: $x_{0,1}, x_{0,2}, \dots$
- percentily: $x_{0,01}, x_{0,02}, \dots$

Matematická statistika

- ***Náhodný výběr***

statistický soubor $(x_1, \dots, x_n)$ získáme *n*-krát opakováním náhodného pokusu –> pozorování náhodné veličiny

= ***pozorovaná hodnota náhodného výběru*** $(X_1, \dots, X_n)$

⇒ Realizací náhodného výběru získáme obecně různé statistické soubory.

Statistika (výběrová charakteristika) = funkce náhodného výběru $T(X_1, \dots, X_n)$

Matematická statistika – odhady parametrů rozdělení

- Skutečnou hodnotu parametrů rozdělení pravděpodobnosti náhodné veličiny obvykle ***neznáme.***
- Odhadujeme ji pomocí statistického souboru
- ***Odhad***:
  - Nestranný
  - Stranný (vychýlený)
- ***Bodový odhad*** parametru je pozorovaná hodnota $t = T(x_1,\dots,x_n)$ na statistickém souboru $(x_1,\dots,x_n)$

Matematická statistika – odhady parametrů rozdělení

Bodové odhady:

- ***Střední hodnota*** – aritmetický průměr
- ***Směrodatná odchylka*** – výběrová směrodatná odchylka
- ***Rozptyl*** – druhá mocnina výběrové směrodatné odchylky

Matematická statistika – odhady parametrů rozdělení

Intervalový odhad

- (interval spolehlivosti, konfidenční interval) pro parametr ε se ***spolehlivostí*** 1 – α, kde $\alpha \in \langle 0;1 \rangle$ je interval <t_1; t_2>, kde hodnoty t_1 a t_2 jsou dané statistickým souborem
- Spolehlivost 1 – α volíme 0,95 nebo 0,99
- Intervalový odhad střední hodnoty normálního rozdělení:

$$\left\langle \bar{x} - t_{1-\alpha/2}\frac{s}{\sqrt{n-1}}; \bar{x} + t_{1-\alpha/2}\frac{s}{\sqrt{n-1}} \right\rangle$$

Matematická statistika
Testování statistických hypotéz

- Při sledování náhodných veličin jsme často nuceni ověřit určité předpoklady či domněnky o jejich vlastnostech pomocí jejich pozorovaných hodnot.
- ***Statistická hypotéza H*** je tvrzení o vlastnostech rozdělení pravděpodobnosti pozorované náhodné veličiny *X*.

Matematická statistika
Testování statistických hypotéz

Postup jímž ověřujeme danou hypotézu, se nazývá ***test statistické hypotézy.***

$H: \eta = \eta_0$ – ***nulová hypotéza***

$H_A: \eta \neq \eta_0$ – ***alternativní hypotéza*** – volíme dle požadavků úlohy

Hypotéza:

- Dvoustranná
- Jednostranná

Matematická statistika
Testování statistických hypotéz

Pro testování hypotézy $H: \eta = \eta_0$ proti nějaké zvolené alternativní hypotéze se konstruuje vhodné ***testovací kritérium*** $T(X_1,\dots,X_n)$.

Obor hodnot testovacího kritéria *T* se za předpokladu, že platí hypotéza *H,* rozdělí na dvě podmnožiny:

- *Kritický obor* W_α
- *Obor nezamítnutí* $\overline{W}_\alpha$

Hladina významnosti α – pravděpodobnost toho, že testovací kritérium nabude hodnotu z kritického oboru.

Matematická statistika
Testování statistických hypotéz

Rozhodnutí o hypotéze

Jestliže pozorovaná hodnota testovacího kritéria $t = T(x_1,\dots,x_n)$ na statistické souboru $(x_1,\dots,x_n)$ padne do kritického oboru, ***zamítáme hypotézu*** *H* současně ***nezamítáme alternativní hypotézu*** H_A.

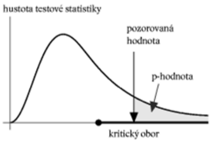
Chyby

Chyba prvního druhu – hypotéza *H* platí a my ji zamítáme. Pravděpodobnost této chyby je hladina významnosti α.

Chyba druhého druhu – Hypotéza *H* neplatí a my ji nezamítáme. Pravděpodobnost této chyby se nazývá síla testu.

Matematická statistika
Testování statistických hypotéz

- Obvyklým výstupem většiny softwarů, které umožňují testovaní statistických hypotéz, není přímo zamítnutí či nezamítnutí hypotézy, ale tzv. ***P - hodnota***.
- P - hodnota udává mezní hladinu významnosti, při které bychom danou hypotézu ještě zamítali. Hypotézu H zamítáme na hladině významnosti , jestliže P - hodnota je menší než α.



Matematická statistika
Testování statistických hypotéz

Studentův t-test

Hypotézy:

- zda normální rozdělení, z něhož pochází náhodný výběr, má ***určitou konkrétní střední hodnotu***, přičemž rozptyl je neznámý;
- Zda dvě normální rozdělení se stejným (třeba i neznámým) rozptylem, z nichž pocházejí dva nezávislé náhodné výběry mají ***stejné střední hodnoty*** (případně lišící se o určitou hodnotu).

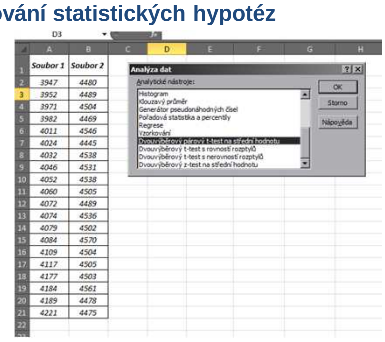
Matematická statistika
Testování statistických hypotéz

Studentův t-test

- ***Jednovýběrový*** – H_0: $\mu = \mu_0$
- ***Párový*** – testuje se rozdíl středních hodnot, stejný rozsah
- ***Dvouvýběrový*** – dva výběry, různý rozsah

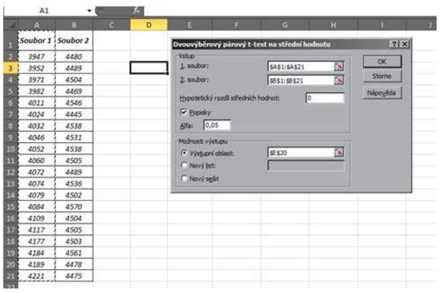
Matematická statistika
Testování statistických hypotéz

Studentův t-test



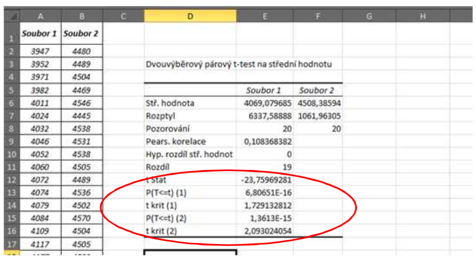
Matematická statistika
Testování statistických hypotéz

Studentův t-test



Matematická statistika
Testování statistických hypotéz

Studentův t-test



Soubor 1	Soubor 2
3947	4480
3952	4489
3971	4504
3982	4469
4011	4546
4024	4445
4032	4538
4046	4531
4052	4538
4060	4505
4072	4489
4074	4536
4079	4502
4084	4570
4109	4504
4117	4505

Dvouvýběrový párový t-test na střední hodnotu

	Soubor 1	Soubor 2
Stř. hodnota	4069,079685	4508,38594
Rozptyl	6337,58888	1061,96305
Pozorování	20	20
Pears. korelace	0,108368382	
Hyp. rozdíl stř. hodnot	0	
Rozdíl	19	
t Stat	-23,75969281	
P(T<=t) (1)	6,80651E-16	
t krit (1)	1,729132812	
P(T<=t) (2)	1,3613E-15	
t krit (2)	2,093024054	

Matematická statistika
Testování statistických hypotéz

ANOVA (Analysis Of Variance)

- Srovnáváme, zda rychlost prostupu ultrazvukových vln je ve třech různých částech ŽB konstrukce stejná.
- Obecně – srovnáváme 2 a více skupin
- Proč nesrovnat po dvojicích? -> roste šance, že uděláme chybu prvního druhu.

  => je výhodnější testovat pouze ***jednu hypotézu***

Matematická statistika
Testování statistických hypotéz

ANOVA (Analysis Of Variance)

Hypotéza:

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$.

- Předpoklad homogenity variance (a normality).
- H_A: není pravda, že jsou všechny střední hodnoty stejné (tedy alespoň jedna se liší od ostatních)

Matematická statistika
Testování statistických hypotéz

ANOVA (Analysis Of Variance)

- Nejjednodušší varianta: *Single Factor ANOVA*

  Model: $X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

náhodná variabilita N(0,σ²)

Společná střední hodnota

"posunutí" *i*-té skupiny proti společnému průměru

H_0 je tedy možné vyjádřit $\alpha_i = 0$ pro všechna i
(jinými slovy - posunutí mezi skupinami není, je tam jen náhodná variabilita)

Matematická statistika
Testování statistických hypotéz

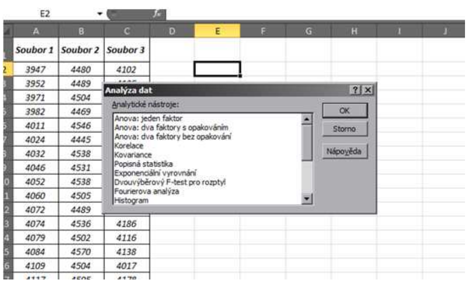
ANOVA (Analysis Of Variance)

příklad :

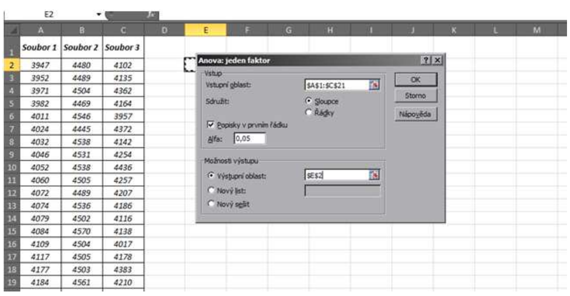
Porovnání tří částí konstrukce – rychlost šíření ultrazvukových vln

	A	B	C	D
1	Soubor 1	Soubor 2	Soubor 3	
2	3947	4480	4102	
3	3952	4489	4135	
4	3971	4504	4362	
5	3982	4469	4164	
6	4011	4546	3957	
7	4024	4445	4372	
8	4032	4538	4142	
9	4046	4531	4254	
10	4052	4538	4436	
11	4060	4505	4257	
12	4072	4489	4207	
13	4074	4536	4186	
14	4079	4502	4116	
15	4084	4570	4138	
16	4109	4504	4017	
17	4117	4505	4178	
18	4177	4503	4383	
19	4184	4561	4210	
20	4189	4478	4255	
21	4221	4475	4139	

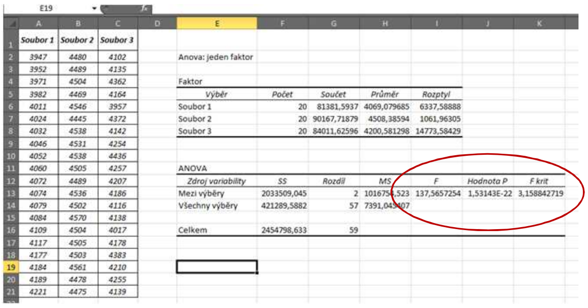
Matematická statistika
Testování statistických hypotéz



Matematická statistika
Testování statistických hypotéz



Matematická statistika
Testování statistických hypotéz



	Soubor 1	Soubor 2	Soubor 3
2	3947	4480	4102
3	3952	4489	4135
4	3971	4504	4362
5	3982	4469	4164
6	4011	4546	3957
7	4024	4445	4372
8	4032	4538	4142
9	4046	4531	4254
10	4052	4538	4436
11	4060	4505	4257
12	4072	4489	4207
13	4074	4536	4186
14	4079	4502	4116
15	4084	4570	4138
16	4109	4504	4017
17	4117	4505	4178
18	4177	4503	4383
19	4184	4561	4210
20	4189	4478	4255
21	4221	4475	4139

Anova: jeden faktor

Faktor

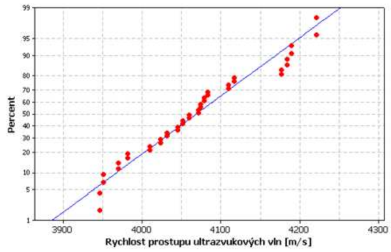
Výběr	Počet	Součet	Průměr	Rozptyl
Soubor 1	20	81381,5937	4069,079685	6337,58888
Soubor 2	20	90167,71879	4508,38594	1061,96305
Soubor 3	20	84011,62596	4200,581298	14773,58429

ANOVA

Zdroj variability	SS	Rozdíl	MS	F	Hodnota P	F krit
Mezi výběry	2033509,045	2	1016754,523	137,5657254	1,53143E-22	3,158842719
Všechny výběry	421289,5882	57	7391,045407			
Celkem	2454798,633	59				

Matematická statistika
Testy dobré shody

Testování a fitování rozdělení pravděpodobnosti



Matematická statistika
Testy dobré shody

Testování a fitování rozdělení pravděpodobnosti

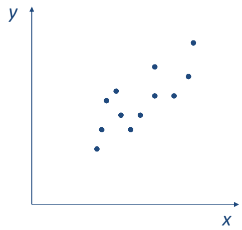
- Anderson – Darling
- Ryan – Joiner (Shapiro – Wilk)
- Kolmogorov - Smirnov

Regresní analýza

- Hledání a zkoumání závislostí proměnných, jejichž hodnoty získáme při realizaci experimentu.

  např.: vztah mezi nepřímou a přímou metodou zkoušení

- Proložení bodového diagramu

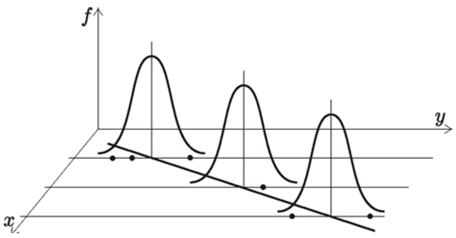


Regresní analýza

- Závislost mezi veličinami *X* a *Y* vyjadřuje regresní funkce:

  y = φ(**x**; β),

  kde β = (β_1, …, β_m) jsou regresní koeficienty



Regresní analýza

- Pro určení neznámých regresních koeficientů β = (β_1, …, β_m) minimalizujeme tzv. ***reziduální součet čtverců***:

$$S^* = \sum_{i=1}^{n} \left[y_i - \varphi(\mathbf{x}_i, \beta) \right]^2$$

=> ***Metoda nejmenších čtverců***

- Lineární regresní funkce:

$$y = \sum_{j=1}^{m} \beta_j f_j(\mathbf{x})$$

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA STAVEBNÍ
ÚSTAV STAVEBNÍHO ZKUŠEBNICTVÍ

Dotazy?

Děkuji za pozornost!

Petr Misák

misak.p@fce.vutbr.cz