ZÁKLADY STATISTIKY

aneb

PRŮVODCE

STATISTICKÝM ZPRACOVÁNÍM

KVANTITATIVNÍCH DAT

Přednášky np4001+nk4001

Doc. RNDr. Jiří Zháněl, Dr.

Garant předmětu

DOPORUČENÁ LITERATURA

Anděl, J. (1993). ***Statistické metody*. Praha: Matfyzpress.**

Cyhelský, L., Kahounová, J. & Hindls, R. (1996). ***Elementární statistická analýza*. Praha: Management Press.**

Gajda,V. & Zvolská, J. (1982). ***Úvod do statistických metod*. PF Ostrava. [Skriptum].**

Gibilisco, S. (2009). ***Statistika bez předchozích znalostí*. Brno: Computer Press.**

DOPORUČENÁ LITERATURA

Hendl, J. (2012). ***Přehled statistických metod zpracování dat. Analýza a metaanalýza dat.*** **Praha: Portál.**

Kovář, R. & Blahuš, P. (1989). ***Aplikace vybraných statistických metod v antropomotorice*****. Praha: SPN. [Skriptum].**

Meloun M. & Militký, J. (1994). Statistické zpracování experimentálních dat. Praha: Plus.

DOPORUČENÁ LITERATURA

Meloun M. & Militký, J. (1996). ***Statistické zpracování experimentálních dat. Sbírka úloh*****. Pardubice: Univerzita Pardubice.**

Seger, J. & Hindls, R. (1993). ***Statistické metody v ekonomii*****. Praha: H & H.**

Seger, J. & Hindls, R. (1995). ***Statistické metody v tržním hospodářství*****. Praha: Victoria Publishing.**

A mnoho dalších …

PROGRAM PŘEDNÁŠEK

ZÁKLADY STATISTIKY 1

1. ÚVOD

1.1 Historie statistiky, pojem a struktura statistiky, základní statistické pojmy

1.2 Teorie měření, měřící stupnice (škály), metodologické problémy měření

PROGRAM PŘEDNÁŠEK

2. DESKRIPTIVNÍ (POPISNÁ) STATISTIKA

2.1 Statistické třídění dat, zpracování a grafické znázornění

2.2 Míry polohy

2.3 Míry variability

2.4 Míry závislosti

2.4.1 Závislost pevná & volná (statistická, korelační)

2.4.2 Lineární korelace a lineární regrese

2.4.3 Součinová a pořadová korelace

PROGRAM PŘEDNÁŠEK

ZÁKLADY STATISTIKY 2

3. ANALYTICKÁ STATISTIKA

3.1 Opakování: výzkumné soubory

3.2 Opakování: hypotézy (nulová/alternativní)

3.3 Statistická a věcná významnost

3.4 Testování statistických hypotéz

1.1 HISTORIE STATISTIKY

(k samostudiu)

„Nur wer die Vergangenheit kennt, hat eine Zukunft“.

„Only he who knows the past has a future“.

Wilhelm von Humboldt

(1767-1835, německý učenec a státník, spoluzakladatel Humboldt-Universität zu Berlin).

1.1 HISTORIE STATISTIKY

Nejstarší písemné památky statistické povahy pocházejí ze Sumeru (nejstarší stát světa 3000 – 2000 př. n. l., Perský záliv).

Hliněné destičky obsahují záznamy o časových intervalech, počtech osob, počet domácího zvířectva, množství úrody, atd.

1.1 HISTORIE STATISTIKY

Pojem *statistika* pochází z latinského slova *status* (tj. postavení, stav).

Počátky statistických postupů využívány již ve středověku ke zjišťování počtu obyvatelstva, velikosti majetku, území, obchodu, armády, atd.

Statistika jako součást přednášek na středověkých univerzitách =>
(1) UNIVERZITNÍ STATISTIKA.

1.1 HISTORIE STATISTIKY

V 17. století se Angličané John Graunt a William Petty zabývali zkoumání různých *hromadných společenských jevů* za pomocí číselných charakteristik skupin obyvatelstva (např. *počty narozených a zemřelých osob, počtem obyvatel a složením rodin*).

Tyto postupy byly nazvány (2) *POLITICKÁ ARITMETIKA*
(využitelné politicky, používány aritmetické postupy).

1.1 HISTORIE STATISTIKY

17. století:

(3) TEORIE PRAVDĚPODOBNOSTI

- ***Francie*** **(B. Pascal, P. de Fermat, de Moivre, de Laplace, Poisson);**
- ***Holandsko*** **(Ch. Huygens);**
- ***Švýcarsko*** **(J. Bernoulli, Euler);**
- ***Německo*** **(C. F. Gauss)**
- ***Rusko*** **(Čebyšev, Markov, Ljapunov).**

19. století = postupná integrace:

UNIVERZITNÍ STATISTIKA + POLITICKÁ ARITMETIKA + TEORIE PRAVDĚPODOBNOSTI

⇒ MODERNÍ STATISTIKA

Aplikace do praxe, do výzkumu o příčinných vztazích mezi hromadnými jevy

(Belgie, L. A. J. Quételet).

Později pronikání statistiky do přírodních a technických věd

(Anglie, Galton, Pearson a Fisher).

HISTORIE STATISTIKY V ČECHÁCH

Nejstarší dochovaný zápis:

„SOUPIS MAJETKU LITOMĚŘICKÉHO KOSTELA Z ROKU 1058“

(Kapitula sv. Štěpána v Litoměřicích, www.czso.cz).

HISTORIE STATISTIKY V ČECHÁCH

VÝZNAMNÁ DATA

- **13. října 1753** … **patent císařovny Marie Terezie (1717 – 1780) o každoročním sčítání lidu,**
- **6. března 1897** … zřízen **Zemský statistický úřad Království českého,** (první statistický úřad na území dnešní České republiky).
- **1909** … vyšla první **„Statistická příručka království Českého".**

HISTORIE STATISTIKY V ČECHÁCH

VÝZNAMNÁ DATA

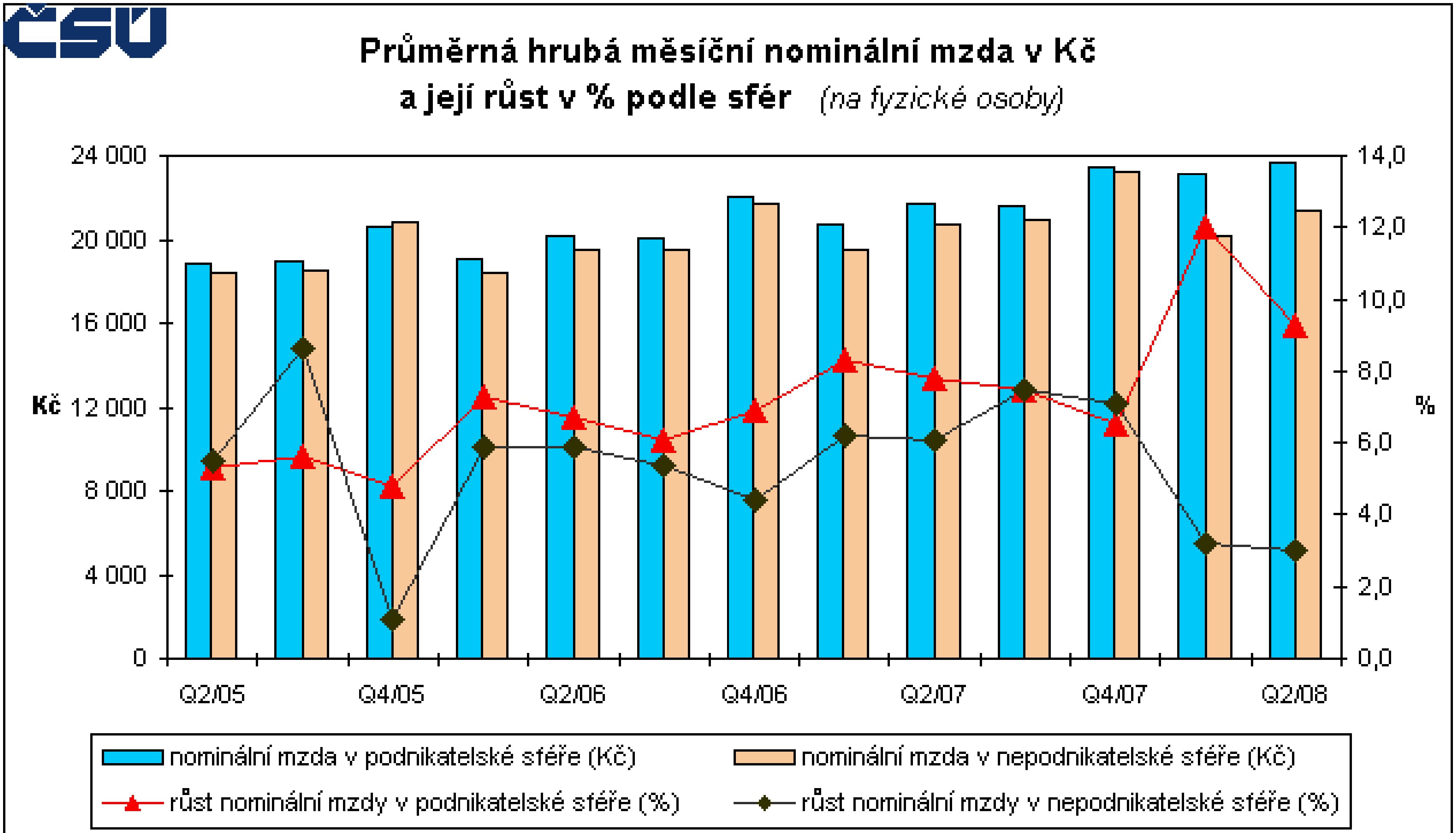
1918 (vznik Československa) => **zákon č. 49 o organizaci statistické služby (1919).**

1919 … založen **STÁTNÍ ÚŘAD STATISTICKÝ** (SÚS) jako orgán pověřený **celostátními statistickými šetřeními** (např. sčítání lidu).

1.1.1993 (vznik ČR) všechny kompetence převzal **ČESKÝ STATISTICKÝ ÚŘAD (ČSÚ).**

Český statistický úřad (ČSÚ)

NEJŽÁDANĚJŠÍ INFORMACE: inflace, makroekonomické údaje, obyvatelstvo, regiony, města, obce, ročenky, sčítání lidu, volební výsledky, základní údaje o ČR.



1.1.2 POJEM A STRUKTURA STATISTIKY

STATISTIKA OBECNĚ

Obor zabývající se

zpracováním, rozborem a zveřejňováním

informací,

které kvantitativně charakterizují

zákonitosti společenského života

(Encyklopedický slovník, 1982).

1.1.2 POJEM A STRUKTURA STATISTIKY

MATEMATICKÁ STATISTIKA

Matematický obor zabývající se

zpracováním dat a rozborem statistických charakteristik

popisovaného statistického souboru

(Encyklopedický slovník, 1982)

Základy statistiky = opravdu jen ZÁKLADY!
(viz příklad)

Např. ***Pravděpodobnost a statistika*** (Friesl, 2004).

Definice náhodného jevu:

Je-li dána množina Ω (všech výsledků náhodného pokusu, tj. pokusu, jehož výsledek není jednoznačně určen podmínkami, za kterých je prováděn), pak náhodným jevem (v Ω) nazýváme každou podmnožinu množiny Ω.

(MATEMATICKÁ) STATISTIKA

1. DESKRIPTIVNÍ

(popisná)

2. ANALYTICKÁ

(inferentní, induktivní, srovnávací)

1. DESKRIPTIVNÍ STATISTIKA

se zabývá zpracováním a popisem dat.

Poskytuje metody umožňující přehledné a názorné zpracování dat, např. v podobě:

- **tabulek,**
- **grafů (znázornění rozložení četností),**
- **výpočtu základních statistických charakteristik** (např. aritmetický průměr nebo korelační koeficient).

2. ANALYTICKÁ (INFERENTNÍ) STATISTIKA

vychází z výsledků deskriptivní statistiky (zpracování dat), **umožňuje nám data analyzovat, tzn. vyhodnotit.**

Tedy např. posoudit, zda

diference mezi středními hodnotami (M) výsledků testu „skok daleký z místa" tréninkových skupin A a B je statisticky (věcně) významná,

což může být vysvětleno vlivem různých tréninkových metod.

SYMBOLICKÉ ZNÁZORNĚNÍ FUNKCE STATISTIKY

STATISTIKA = ZPRACOVÁNÍ + POPIS + ANALÝZA DAT

1.1.3 ZÁKLADNÍ STATISTICKÉ POJMY

STATISTICKÝ SOUBOR
je souhrn (množina) statistických jednotek stejného druhu

Rozlišujeme pojmy základní soubor a výběrový soubor.

Rozsah základního souboru N, výběrového souboru n.

Základní soubor **(populace, N) je soubor všech statistických jednotek, které teoreticky mohou být předmětem sledování.**

Např. 1) všichni studenti oboru TV a sport v ČR, Evropě,

2) všichni členové fotbalové reprezentace v roce 2022,

3) všechny pětileté děti v ČR narozené k 1.1. 2022, …

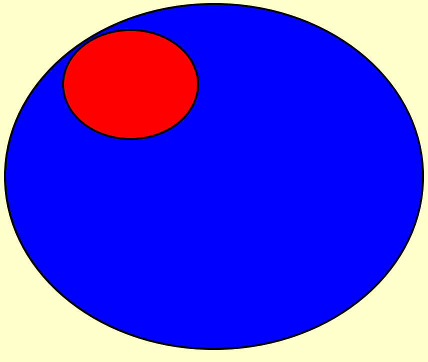
ZÁKLADNÍ SOUBOR (ZS)

(stručné opakování z Metodologie)

ZS má zpravidla značný rozsah, zjištění zkoumaných vlastností všech prvků je buďto nemožné nebo je příliš časově a ekonomicky náročné.

Výzkumné šetření (zjištění) se proto provádí u vybraných jednotek ze základního souboru => výběrový soubor (n).

Výběrový soubor **je náhodnou podmnožinou prvků** ***základního souboru*** **a reprezentuje jej.**



VÝBĚROVÝ SOUBOR získáváme tzv. NÁHODNÝM VÝBĚREM.

- ✓**Každý prvek základního souboru má stejnou možnost být vybrán.**
- ✓ **O vybrání či nevybrání do výběrového souboru rozhoduje tedy pouze náhoda.**

Např. **ZS** (N=10 000) = studenti TV v CZ, **VS** (n=100)

Z výsledků výzkum úrovně znaků výběrového souboru *(náhodně vybraného)* **je možno usuzovat – při splnění určitých podmínek – na vlastnosti *základního souboru.***

METODY NÁHODNÉHO VÝBĚRU PRVKŮ DO VÝBĚROVÉHO SOUBORU (VS)

I. LOSOVÁNÍ

- ✓ **losování statistických jednotek *s jejich vracením do osudí*** (u malých souborů),
- ✓ **losování statistických jednotek *bez vracení do osudí*** (u velkých souborů),
- ✓ **Generátor náhodných čísel**

(software, ukázat na příkladu)

II. TABULKA NÁHODNÝCH ČÍSEL

Např. ze základního souboru N=540 máme vybrat n=12

1. **V tabulce zvolíme libovolné číslo, od něj čteme uvedená čísla s potřebným počtem míst** (např. N=540 => trojmístná čísla).
2. **Do výběru zahrnujeme ty jednotky základního souboru, jejichž přiřazená čísla jsou < 540.**
3. **Čísla vyšší než rozsah základního souboru vynecháme.**
4. **Pokračujeme tak dlouho, než dosáhneme požadovaného rozsahu výběrového souboru.**

N=540

n=12

85306	37114	22718	50584	92291	56575	24075	43889
32066	43098	75738	94910	15403	89151	73322	18370
63314	87302	49472	24885	79506	60638	07132	00908
40287	52435	23926	92544	54099	31497	06853	22864
30925	46148	20138	33874	56715	38424	38273	11361
27146	37012	43361	03173	97911	71313	44256	66609
01674	47274	56350	37512	14883	99673	62298	33948
76730	25043	16686	54737	57431	01786	20803	69465
93941	84434	22384	13240	93617	51549	28532	57150
90475	10341	39703	83224	37858	61657	04184	15597
86115	17196	24569	26820	66299	39960	02489	53079
51156	74037	12501	94162	42006	16135	82797	31296
59886	03051	78702	13402	74318	10870	72107	11550
13960	95736	43637	60399	19080	60261	11207	73065
39954	86726	91039	13884	25376	36880	02564	96978
47906	99501	27753	69946	66875	25601	30038	78786
66444	15979	83469	76952	50065	72802	70630	87336
40177	01081	57788	08612	39886	42234	04905	83274
46747	30655	41878	93610	51745	41771	61398	98154
60888	18689	45966	25837	70906	60733	11765	09293

VÝSLEDEK: 936 (mimo), 175, 154, 928, 532, 571, 509, 047, 510, 341, 397, 038, 322, 437, 858, 616, 570, 418.

Možno vyzkoušet pomocí

Excel – Analýza dat – Generátor náhodných čísel:

Typ rozložení - diskrétní

Počet proměnných – dle počtu číslic

jednociferné n = 1

dvouciferné n = 2

atd.

Pro N=54O se počet proměnných rovná 3.

III. SKUPINOVÝ VÝBĚR

... užívá se, je-li základní soubor velmi početný a je uspořádán do skupin (např. třídy ve škole), z nichž vybíráme skupiny – nutný je dostatečný počet skupin.

IV. STRATIFIKOVANÝ VÝBĚR

... vychází z rozdělení základního souboru na skupiny (straty), z každé z nich se pak dělá náhodný výběr.
Je žádoucí proporcionální zastoupení ve výběru ze straty (neproporcionální ve specifických případech).

Př. 1. výzkumný soubor „vysokoškoláci“ (= studující techniky, univerzity, uměleckých vysokých škol, atd.).

Př. 2. výzkumný soubor „učitelé s praxí do ...“ (1. do 5 let, 2. do 10 let, 3. do 15 let, do 20 let, atd.).

V. ZÁMĚRNÝ VÝBĚR

… nerozhoduje náhoda, výzkumník sám vybírá jedince jež považuje za typické (subjektivní výběr).

Výsledky se týkají jen daného výběru (v závěrech výzkumu nutná formulace:

„na daném vzorku se prokázalo. že…“)!!!

Problém: výběr x dobrovolníci (rozdíly - vyšší výkon, motivace, větší potřeba sociálního uznání, …).

Nelze je použít při standardizaci testů!

Další podrobnosti např.

Chrástka, M. (2007). Metody pedagogického výzkumu.

STATISTICKÉ JEDNOTKY
jsou prvky statistického souboru, které mají alespoň jednu společnou vlastnost (znak)

Statistickými jednotkami mohou být např. osoby (subjekty), **věci** (objekty), **resp. události, jejichž vlastnosti nás zajímají.**

Zjišťujeme-li pouze *jeden statistický znak* (např. tělesnou výšku), **hovoříme o *jednorozměrném statistickém souboru.***

Zjišťujeme-li *dva nebo více znaků,* hovoříme o dvourozměrném (výška a hmotnost), **resp. vícerozměrném statistickém souboru** (3 a více znaků).

STATICKÉ ZNAKY

(stručné opakování z Metodologie)

Vyjádření hodnot statistických znaků (proměnných) je možné slovy nebo čísly.

Klasifikace:

1. **SLOVNÍ PROMĚNNÉ (alfabetické, kategoriální) = KVALITATIVNÍ ZNAKY.**
2. **ČÍSELNÉ PROMĚNNÉ (numerické) = KVANTITATIVNÍ ZNAKY.**

STATICKÝ ZNAK

je společná vlastnost jednotek statistického souboru

Statistické znaky vyjadřují vlastnosti statistických jednotek.

1. KVALITATIVNÍ ZNAKY (kategoriální, slovně)

Např. muž/žena, plavec/neplavec, zdravý/nemocný
barva očí: zelené, modré, hnědé, ...,
herní kategorie: žáci mladší, starší, junioři, ...

☯ ***alternativní*** *(binární, dichotomické):*

nabývá-li znak pouze dvou variant (muž/žena)

☯ ***množné*** *(polytomické):*

nabývá-li znak více než dvou variant (barva očí: zelená, modrá, černá).

2. KVANTITATIVNÍ ZNAKY

☯ ***spojité* neboli *kontinuální***

nabývají libovolných **reálných číselných hodnot:**

např. výsledek v běhu na 100 m (10,7 s),

ve skoku vysokém (220 cm).

Mezi 2 hodnotami může být vždy další hodnota:

(10,7 s; 10,72 s; 10,723 s)

☯ ***nespojité* neboli *diskrétní***

(nabývají pouze **konečný počet číselných hodnot**,

nejčastěji z oboru celých nezáporných čísel.

např. *počet úspěšných hodů na koš, leh-sedy, kliky).*

1.2 TEORIE MĚŘENÍ, MĚŘÍCÍ STUPNICE (ŠKÁLY)

1.2.1 ZÁKLADNÍ POJMY TEORIE MĚŘENÍ

Měření … v průběhu historického vývoje lidské společnosti je běžné jeho každodenní užití (např. hodinky, tachometr automobilu, váha, atd.).

Historické počátky měření … porovnávání objektů s počtem *prstů, délkou palce, délkou chodidla, lokte, paže,* tj. primitivní měřící způsoby.

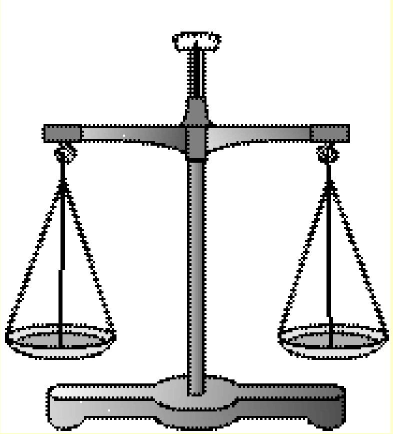
Rozvoj vědy a techniky složitých měřících přístrojích.

1.2 TEORIE MĚŘENÍ, MĚŘÍCÍ STUPNICE (ŠKÁLY)

1.2.1 ZÁKLADNÍ POJMY TEORIE MĚŘENÍ

Problematiku kvantifikace (měření) řeší obor nazývaný TEORIE MĚŘENÍ.

a) Měřitelnost *fyzikálních vlastností* (délka, čas, hmotnost),



b) Měřitelnost *psychických vlastností*
(inteligence, strach, postoje atd.).

REPREZENTAČNÍ TEORIE MĚŘENÍ (Campbell):

… měření jako „přiřazování číslic k reprezentaci vlastností“.

Později doplněna o formulaci *„…za měření lze považovat každé přiřazování číslic k objektům nebo událostem … podle pravidel* *(*Stevens*).*

KLASICKÁ KONCEPCE MĚŘENÍ ROZLIŠUJE

(1) FUNDAMENTÁLNÍ* (ZÁKLADNÍ) *MĚŘENÍ

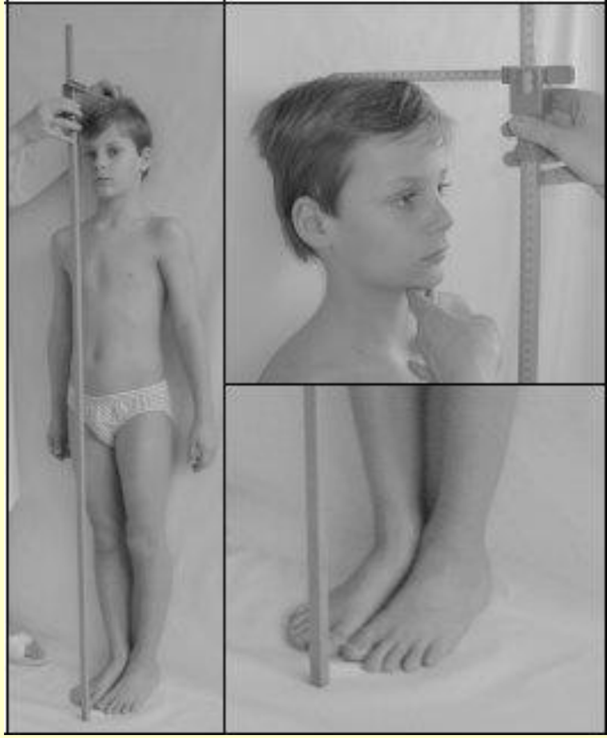
(2) ODVOZENÉ MĚŘENÍ

Další autoři zmiňují

(3) MĚŘENÍ ASOCIATIVNÍ

(Berka, 1977) resp. *asociační* (Blahuš, 1996), označované rovněž jako měření *per fiat, per Definition, by fiat* či *měření na základě konvence.*

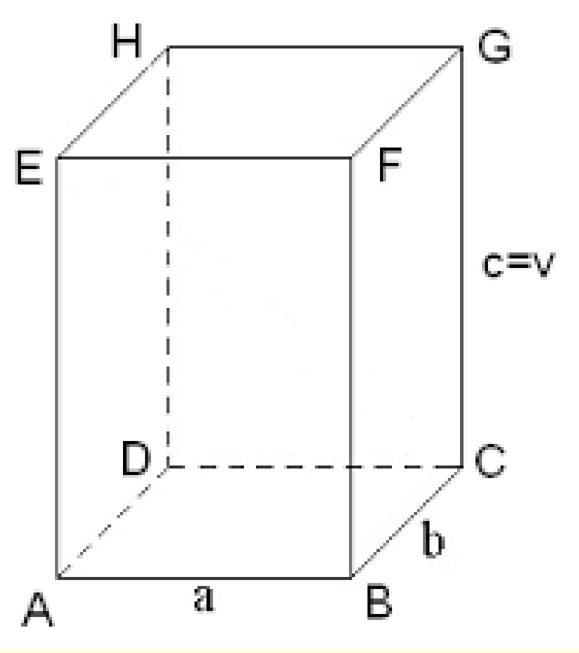
(1) FUNDAMENTÁLNÍ (ZÁKLADNÍ) MĚŘENÍ



***„se vztahuje na bezprostřední měření veličin“* a je to *„každé měření, které nezahrnuje žádná předcházející měření“*.**

Příklad: měření tělesné výšky

(2) ODVOZENÉ MĚŘENÍ

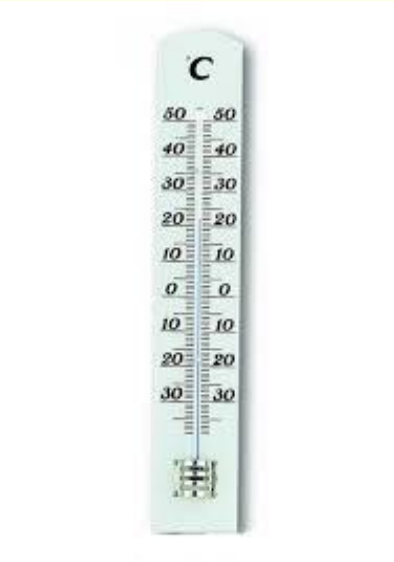


„předpokládá jiná, dříve provedená měření, z nichž je odvozeno na základě vztahů“; a tedy „závisí na předcházejících měřeních“.

Příklad: „měření“ objemu kvádru

(3) ASOCIATIVNÍ MĚŘENÍ (ASOCIAČNÍ)

je takové měření, kdy „je přímo měřená veličina asociována s nepřímo měřitelnou veličinou“.



Příklad 1.

Při měření teploty vycházíme ze závislosti změny objemu kapaliny na teplotě.

Příklad 2.

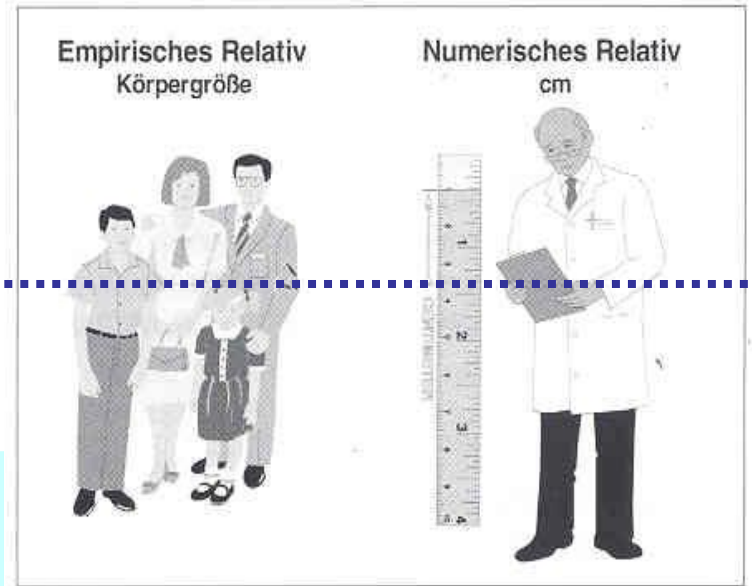
Při testování úrovně vytrvalosti pomocí Cooper testu vycházíme z předpokládané asociace (vztahu) mezi uběhnutou vzdáleností (měřitelná) **a úrovní vytrvalostní schopnosti** (nepřímo měřitelná).

1.2.2 MĚŘÍCÍ STUPNICE (ŠKÁLY)

Empirická proměnná

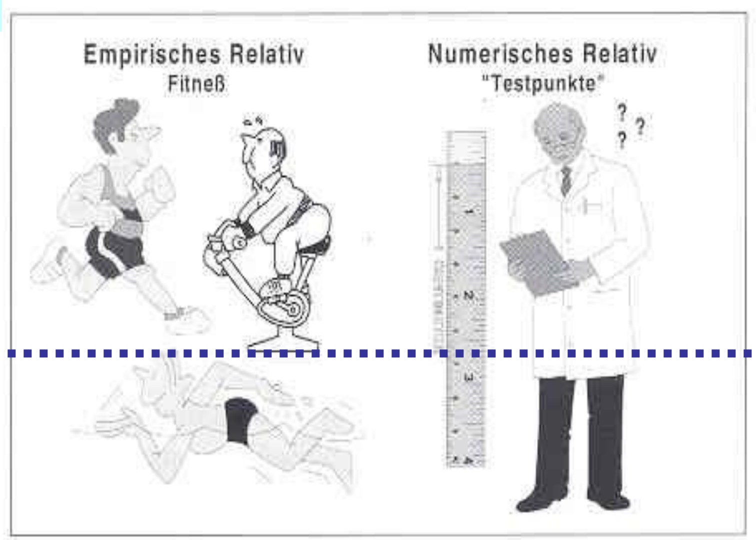
Numerická proměnná

Tělesná výška → **cm**



Zuordnungen von Körpergrößen (ER) zu Zahlenwerten (NR)

Rozdíl ve způsobu měření a přiřazení!



Zuordnung von Fitneßausprägungen (ER) zu Testpunkten (NR)

Empirická proměnná

Numerická proměnná

Kondice → **Testové skore**

TEORII ŠKÁL

(pojem škála, resp. stupnice)

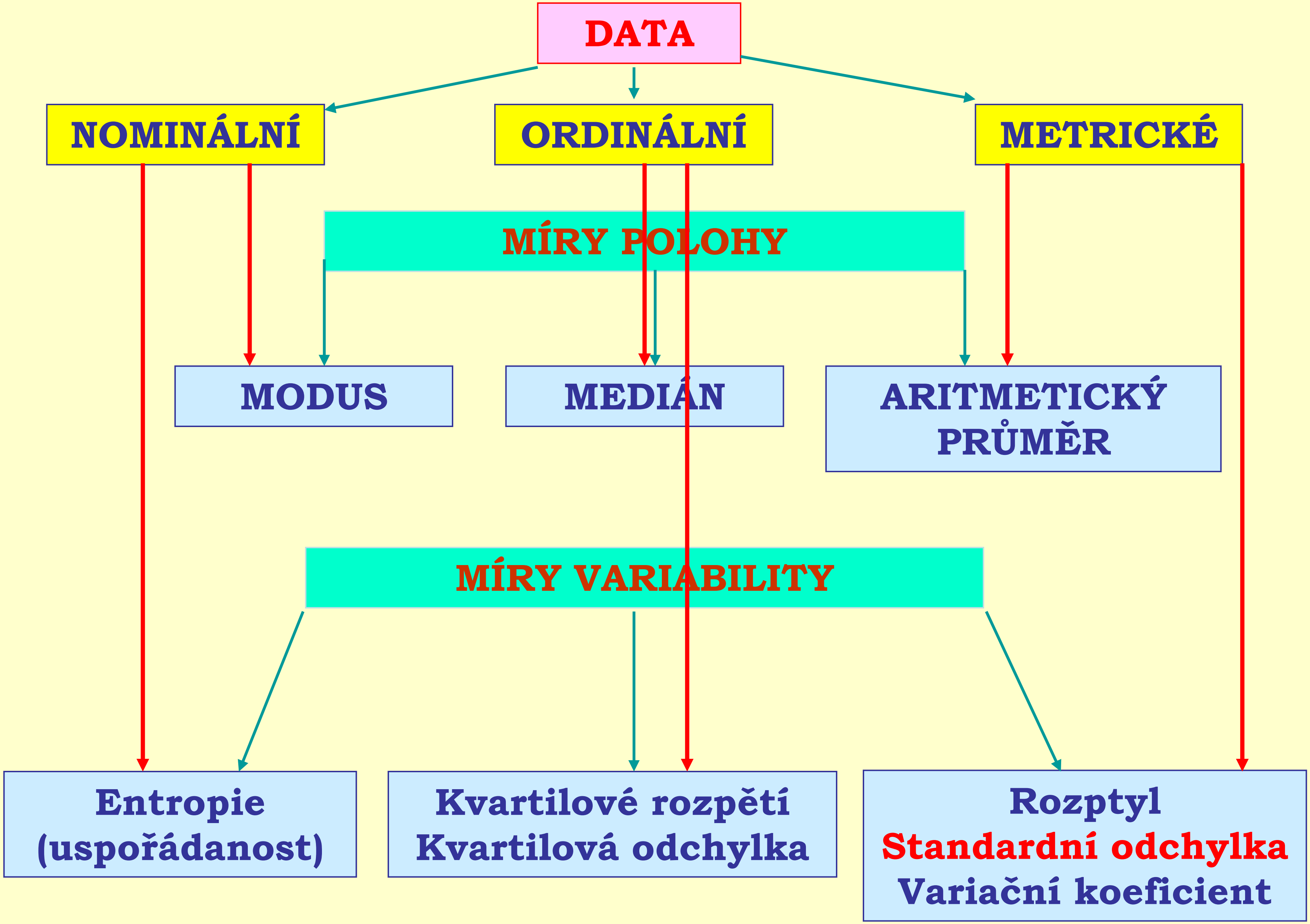
ZÁKLADNÍ DRUHY ŠKÁL (STUPNIC)

1. **NOMINÁLNÍ škála (jmenná, klasifikační)**
2. **ORDINÁLNÍ škála (pořadová)**

NEMETRICKÉ

3. **METRICKÉ škály**
   - **INTERVALOVÁ**
   - **POMĚROVÁ**

METRICKÉ


DATA
NOMINÁLNÍ
ORDINÁLNÍ
METRICKÉ
MÍRY POLOHY
MODUS
MEDIÁN
ARITMETICKÝ PRŮMĚR
MÍRY VARIABILITY
Entropie (uspořádanost)
Kvartilové rozpětí
Kvartilová odchylka
Rozptyl
Standardní odchylka
Variační koeficient

1. NOMINÁLNÍ ŠKÁLA

(jmenná, klasifikační)

… je škála založena na jakémkoliv *přiřazování číslic* ve smyslu pouhého *pojmenování.*

Jde vlastně o *pojmenování* osob či skupin *čísly*, o uspořádání do tříd, které se navzájem *vylučují.*

Např. *pohlaví (M/Ž), kuřák/nekuřák, národnost, čísla hráčů, věkové kategorie (U10–U18)*

1. NOMINÁLNÍ ŠKÁLA

Třídění na znaky:

1. **alternativní** (binární, dichotomické) **= 2 možnosti**

   (plavec/neplavec; kuřák/nekuřák; muž/žena)

2. **množné** (polytomické) **= více než 2 možností**

(oči zelené, modré, hnědé;

věkové kategorie: žáci mladší, starší, junioři)

Základní empirická operace: *„určení rovnosti"*.

Možné relace: =, ≠,

Zpracování znaků: neparametrické statistické metody

2. ORDINÁLNÍ ŠKÁLA (pořadová)

... předpokládá přirozené uspořádání objektů vzhledem k nějaké vlastnosti.

Škála umožňuje *uspořádání* objektů *do pořadí,* je možno určit vztah větší či menší, těžší či lehčí, atd.

Nejsou známy odstupy (intervaly) mezi znaky (čísly) !!!

Např. *školní známky, stupnice tvrdosti, pořadí v cíli.*

Základní empirické operace:

„*určením rovnosti*“ a „*určením vztahu více nebo méně*“.

Relace: =, ≠, >, <,

Zpracování znaků: neparametrické statistické metody.

3. METRICKÉ ŠKÁLY

(INTERVALOVÁ A POMĚROVÁ)

Je zavedena *jednotka měření,* tzn. jsou známy *odstupy (intervaly)* mezi hodnotami (čísly).

Nutný předpoklad: normální rozložení četností!

3. 1 INTERVALOVÁ ŠKÁLA

… vyžaduje stanovení *měrové jednotky a počátku,* jsou přípustné všechny aritmetické operace.

Nula je zvolená!!! => stanovení počátku dohodou.

Např. letopočet (*Diokleciánův, byzantský, křesťanský, čínsky, atd.*),

teplota ve °C (***bod tání ledu = 0°C a bod varu = 100°C při tlaku vzduchu 1013,25 hPa***).

3. METRICKÉ ŠKÁLY
(INTERVALOVÁ A POMĚROVÁ)

3. 2 POMĚROVÁ ŠKÁLA

… z formálního hlediska vlastně intervalová *škála s přirozeným počátkem,* jsou přípustné všechny aritmetické operace.

Nula je absolutní … (nepřítomnost jevu).

Např. čas, věk, výška, hmotnost, teplotní stupnice dle Kelvina (v podstatě všechny fyzikální jednotky).

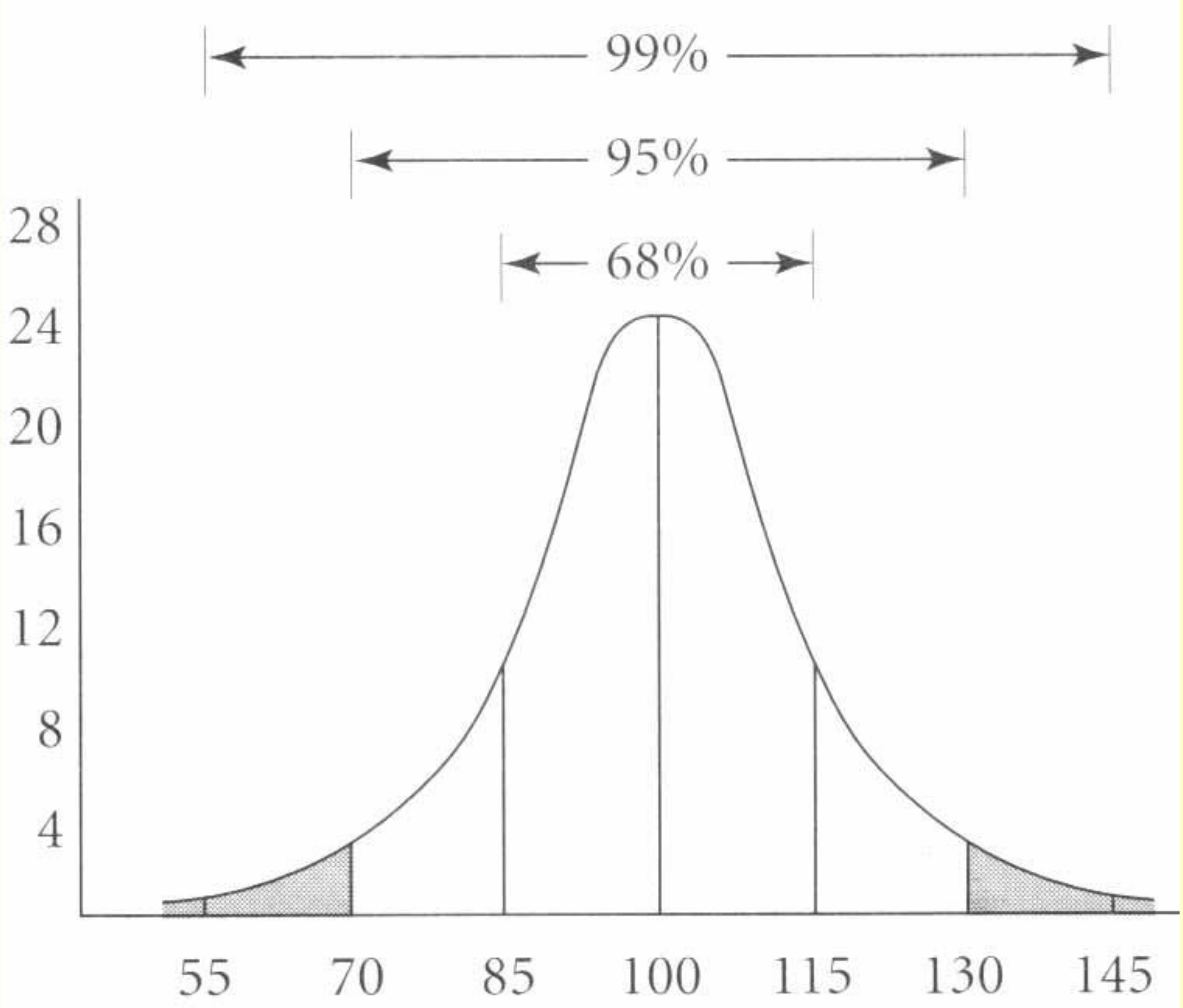
Statistické metody: parametrické i neparametrické.

Přehled typů škál (Bruhn, 1986; Roth, 1995)

TYP ŠKÁLY	NEMETRICKÉ ŠKÁLY		METRICKÉ ŠKÁLY	
	NOMINÁLNÍ	ORDINÁLNÍ	INTERVALOVÁ	POMĚROVÁ
Příklady	Číselné označení barev, psychologického typu, pohlaví, atd.	Školní známky, stupnice tvrdosti, služební pořadí, Richterova stupnice	Teplota ve °C, Fahrenheita, letopočet, inteligenční kvocient	Teplota °Kelvina, věk, váha, výška, velikost úhlu, čas
Operace	$=, \neq$	$=, \neq, >, <$	Navíc: intervaly, nula zvolená	Navíc: nula absolutní
Statistické charakteris.	Modus, absolutní a relativní četnosti	Navíc: medián, kvantily a kvantilové odchylky, procentily	Navíc: arit. Průměr, směrodat.odchylka, šikmost, špičatost	Navíc: koeficient variability, geometr. průměr
Testy Významnosti	χ^2 - test, McNemar test, Cochran test,...	Znaménkový test, Mann-Whitney U-test, Friedmanova pořadová analýza variance, aj.	Parametrické metody: F-test t-test (pro závislé či nezávislé soubory)	Parametrické metody: F-test t-test (pro závislé či nezávislé soubory)
Míry závislosti	Kontingenční a čtyřpolní koeficient	Navíc: pořadová korelace	Navíc: Pearsonova součinová korelace	Navíc: Pearsonova součinová korelace
Statistické metody	Některé neparametrické metody	Všechny neparametrické metody	Všechny neparametrické a parametrické metody	Všechny neparametrické a parametrické metody

Inteligenční kvocient (IQ; Stern, 1912)

je index inteligence, který má normální rozložení s průměrem 100 a standardní odchylkou 15.



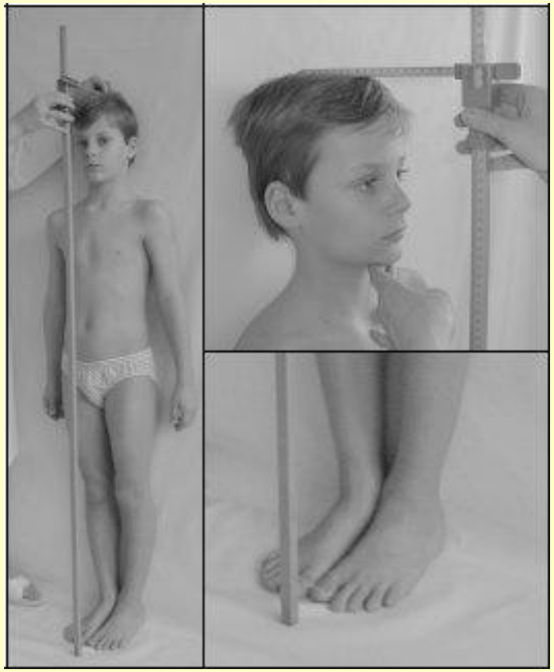
POSTUP PŘI URČENÍ TYPU ŠKÁLY: A. výška (cm)

1. Je známa *jednotka měření?* ANO Škály metrické

2. Je počátek *zvolený* nebo absolutní? absolutní **POMĚROVÁ**

3. Lze *stanovit pořadí?* Nemá smysl zjišťovat

4. Jedná se jen o *pojmenování znaků* čísly? Nemá smysl zjišťovat



Znaky ?

- Kvantitativní
- Výška = spojitý

POSTUP PŘI URČENÍ TYPU ŠKÁLY: B. dějepis (zn)

1. Je známa *jednotka měření?* NE => Nemohou být metrické

2. Je počátek *zvolený* nebo absolutní? Nemá smysl zjišťovat

3. Lze *stanovit pořadí?* ANO => ORDINÁLNÍ

4. Jedná se jen o *pojmenování znaků* čísly? Nemá smysl zjišťovat

Znak?

- Kvantitativní
- Dějepis = spojitý

Pozn. Známky jsou spojitými znaky, i když jsou měřeny pouze na ordinální škále.

Slide 57

Z jakých škál jsou uvedené proměnné?

Studenti sami … potom kontrola!

Klasifikujte znaky obsažené v tabulce – správnou odpověď označte křížkem (X)

Znak	ŠKÁLA				ZNAK	
	Nominální (a)	Ordinální (b)	Intervalová (c)	Poměrová (d)	Spojitý (e)	Diskrétní (f)
1. Pohlaví						
2. Věk						
3. Počet sourozenců						
4. Známka z matematiky						
5. Inteligenční kvocient						
6. Hodnocení v krasobruslení						
7. Výkon ve skoku dalekém						

Pozn. Známky jsou spojitými znaky, i když jsou měřeny pouze na ordinální škále.

Znak	ŠKÁLA				ZNAK	
	Nominální (a)	Ordinální (b)	Intervalová (c)	Poměrová (d)	Spojitý (e)	Diskrétní (f)
1. Pohlaví	■					■
2. Věk				■	■	
3. Počet sourozenců				■		■
4. Známka z matematiky		■			■	
5. Inteligenční kvocient			■		■	
6. Hodnocení v krasobruslení				■	■	
7. Výkon ve skoku dalekém				■	■	

Řešení: 1. a, f; 2. d, e; 3. d, f; 4. b, e; 5. c, e; 6. d, e; 7. d, e.
Pozn. Známky jsou spojitými znaky, i když jsou měřeny pouze na ordinální škále.

ANALÝZA JEDNOROZMĚRNÉHO SOUBORU

METODY DESKRIPTIVNÍ STATISTIKY

1. URČENÍ TYPU ŠKÁLY (nominální, ordinální, metrické)

a) nominální + ordinální ⇒ *neparametrické statistické metody*

b) metrické ⇒ *parametrické statistické metody*

2. ROZLOŽENÍ ČETNOSTÍ ZNAKŮ (NORMÁLNÍ ČI JINÉ)

a) normální ⇒ *parametrické statistické metody*

b) jiné ⇒ *neparametrické statistické metody*

3. VÝPOČET ZÁKLADNÍCH STATISTICKÝCH CHARAKTERISTIK

a) míry centrální tendence (M, Mo, Me)

b) míry variability (s, ...)

c) míry závislosti (r)

2.1 STATISTICKÉ TŘÍDĚNÍ DAT

Výsledkem měření, testování, dotazování jsou *neuspořádaná, neroztříděná a nepřehledná* data (tzv. *hrubé skóre*).

Tabulka 1: Výsledky testování tenistů U12 (n = 10)

Věk	Výška	Váha	BMI	IPR	SH	RB	V	PTC	RRR	RRN	SR
11,0	150,0	36,0	16,0	1,7	24,8	15,0	153,9	43	0,52	0,45	0,69
11,0	155,5	46,0	19,0	2,1	26,7	15,9	159,8	40	0,77	0,49	0,58
11,0	151,0	36,4	16,0	2,7	19,5	15,4	161,2	40	0,63	0,41	0,54
11,0	150,0	39,8	17,7	1,9	23,4	14,4	151,7	40	0,56	0,43	0,59
11,0	144,0	35,0	16,9	2,3	23,7	14,6	153,9	35	0,56	0,54	0,68
11,0	143,0	38,6	18,9	1,4	15,2	14,0	155,3	38	0,69	0,43	0,39
11,0	144,0	41,2	19,9	3,3	20,0	16,2	165,7	17	0,66	0,51	0,49
11,0	153,0	37,0	15,8	2,9	20,0	14,8	158,8	41	0,61	0,46	0,54
11,1	155,0	40,0	16,6	1,4	19,3	15,5	142,4	48	0,47	0,37	0,48
11,1	140,0	32,8	16,7	2,8	20,2	15,0	163,2	37	0,56	0,40	0,62

Chceme-li získat přesnější, smysluplnější, podrobnější informace, je třeba údaje uspořádat: Hovoříme o *statistickém zpracování (třídění) dat.* Nejjednodušším způsobem statistického zpracování dat je tzv. *tabulka rozdělení (rozložení) četností.*

Porodní délka	Pořadí intervalu	Absolutní četnosti	Relativní četnosti	Kumulativní relativní četnosti
42	1	1	0,03	0,03
43	2	0	0	0,03
44	3	1	0,03	0,06
45	4	0	0	0,06
46	5	1	0,03	0,09
47	6	0	0	0,09
48	7	3	0,09	0,18
49	8	4	0,12	0,3
50	9	8	0,25	0,55
51	10	6	0,18	0,73
52	11	4	0,12	0,85
53	12	3	0,09	0,94
54	13	1	0,03	0,97
55	14	0	0	0,97
56	15	1	0,03	1
Celkem		33	1	

2.1.1 JEDNOROZMĚRNÉ ROZDĚLENÍ ČETNOSTÍ

JEDNA VLASTNOST **(např. tělesná výška) statistického souboru je charakterizovaná** ***JEDNÍM STATISTICKÝM ZNAKEM*** ***(170 cm)*** **– jedná se tedy o** ***jednorozměrný statistický soubor.***

Konstrukce tabulky - postup vhodný pro:

(1) nespojité kvantitativní statistické znaky

(např. počet dětí v rodině, úspěšné koše),

(2) spojité statistické znaky s malým počtem výskytu

(např. pro statistické soubory s malým rozsahem).

PŘÍKLAD 1. Při *dvakrát opakovaném testování střelby na koš* byly u deseti osob (n=10) zjištěny výsledky uvedené v tabulce (zaznamenán počet úspěchů z deseti pokusů při 1. resp. 2. testování).

Tabulka (hrubé skóre)

Hráč	A	B	C	D	E	F	G	H	J	K
x_i	7	6	7	8	9	8	8	8	9	10
y_i	4	8	6	8	7	8	7	4	8	10

Pro znaky x_i sestavte (frekvenční) tabulku rozdělení četností.

Posouzení znaků x_i: ***kantitativní, nespojité, poměrová*** **⇒ …**

⇒ … tabulka jednorozměrného rozdělení četností.

Frekvenční tabulka jednorozměrného rozdělení četností.

X_i	Čárkovací metoda	n_i	f_i	Kumulativní četnost N_i	Kumulativní četnost F_i
6	/	1	0.1	1	0.1
7	/ /	2	0.2	3	0.3
8	/ / / /	4	0.4	7	0.7
9	/ /	2	0.2	9	0.9
10	/	1	0.1	10	1.0
Σ		10	1.0	-	-

Vysvětlivky:

n…rozsah souboru x_i…hodnota znaku

n_i…absolutní četnost f_i…relativní četnost   ($f_i = n_i / n$)

Ni … absolutní kumulativní četnost

Fi … relativní kumulativní četnost

Absolutní četnost – vyjadřuje absolutní výskyt jednotlivých znaků, **relativní četnost** – vyjádření v procentech.

Kumulativní relativní četnost – vyjadřuje v % (po vynásobená stem) jaké procento rozsahu souboru má odpovídající variantu a menší hodnotu dané proměnné.

F_i = 0,7 => ***70 % hráčů*** **dosáhlo výsledku** ***8 úspěšných pokusů a méně.***

X_i	Čárkovací metoda	n_i	f_i	Kumulativní četnost	
				N_i	F_i
6	/	1	0.1	1	0.1
7	/ /	2	0.2	3	0.3
8	/ / / /	4	0.4	7	0.7
9	/ /	2	0.2	9	0.9
10	/	1	0.1	10	1.0
Σ		10	1.0	-	-

2. 1. 2 JEDNOROZMĚRNÉ INTERVALOVÉ (SKUPINOVÉ) ROZDĚLENÍ ČETNOSTÍ

Konstrukce tabulky *jednorozměrného intervalového rozdělení četností* je postup vhodný pro:

(1) spojité kvantitativní statistické znaky

(např. výsledky měření běhu na 100 m, tělesné výšky, skoku dalekého),

(2) nespojité statistické znaky s velkým počtem výskytů.

DOPORUČENÁ PRAVIDLA

pro konstrukci tabulky jednorozměrného intervalového rozložení četností

URČENÍ ŠÍŘKY A POČTU INTERVALŮ

Variační rozpětí (R) $R = x_{max} - x_{min}$

Šířka intervalu (h) $h = 0{,}08 \times R$

Počet intervalů (k) $k = \sqrt{n}$

$k \leq 5. \log n$

$k \approx 1 + 3.3 \log n$

(Sturgesovo pravidlo)

Je-li $n < 30$ doporučuje se vytvořit ne více než 6 intervalů.

Je-li $30 < n < 100$ doporučuje se vytvořit 7 až 10 intervalů.

PŘÍKLAD 2. Pro znaky y_i sestavte tabulku

skupinového (intervalového) rozdělení četností.

Hráč	A	B	C	D	E	F	G	H	J	K
x_i	7	6	7	8	9	8	8	8	9	10
y_i	4	8	6	8	7	8	7	4	8	10

Pomocné výpočty pro určení

šířky (h) a počtu intervalů (k)

Variační rozpětí (R) $R = x_{max} - x_{min}$ $R = 10 - 4 = 6$

Šířka intervalu (h) $h = 0{,}08 \times R$ $h = 0.08 \times 6 = 0{,}48 \approx 1$ (pokus)

PŘÍKLAD 2. Pro znaky y_i sestavte tabulku

skupinového (intervalového) rozdělení četností.

Pomocné výpočty pro určení

šířky (h) a počtu intervalů (k)

Počet intervalů (k)

$k = \sqrt{n}$ $k \leq 5.\log n$ $k \approx 1 + 3.3 \log n$

$k = 3.16$ $k \leq 5$ $k \approx 4.3$ $(\log 10 = 1)$

$\Rightarrow$ ***Doporučená šířka intervalu: 1***

$\Rightarrow$ ***Doporučený počet intervalů: 3 až 5***

POZOR !

Intervaly musí být vytvořeny tak, aby

jeden statistický znak nemohl být současně

zařazen do dvou různých intervalů!!!

Intervaly na sebe musejí navazovat!!!

Tabulka skupinového (intervalového) rozdělení četností (znak y_i).

Třída	Interval	Střed	n_i	f_i	N_i	F_i
1	4 – 5	4,5	2	0,2	2	0,2
2	6 – 7	6,5	3	0,3	5	0,5
3	8 – 9	8,5	4	0,4	9	0,9
4	10 –	10,5	1	0,1	10	1,0
Σ	-	-	10	1,0	-	-

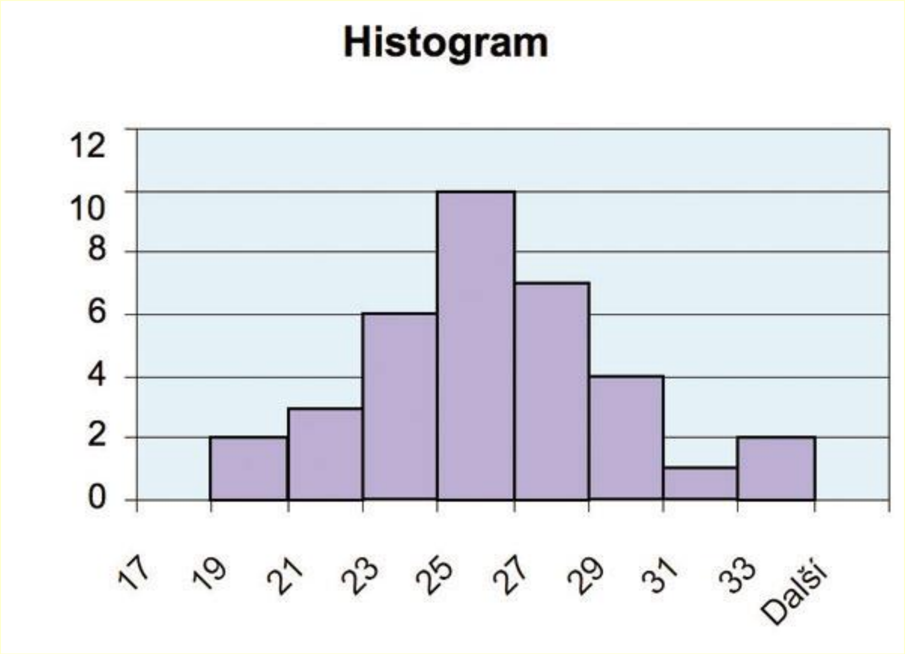
2. 1. 3 GRAFICKÉ ZNÁZORNĚNÍ ROZDĚLENÍ ČETNOSTÍ

Grafické znázornění = **přehlednější a názornější forma znázornění *rozdělení četností.***

1) HISTOGRAM ČETNOSTÍ

(sloupkový diagram, sloupcový graf)

Histogram … **jedna z nejčastěji užívaných forem grafického znázornění *rozdělení četností.***



Histogram je tvořen sloupci

… jejich šířka odpovídá *šířce třídního intervalu,*

… jejich výška odpovídá *absolutní četnosti*

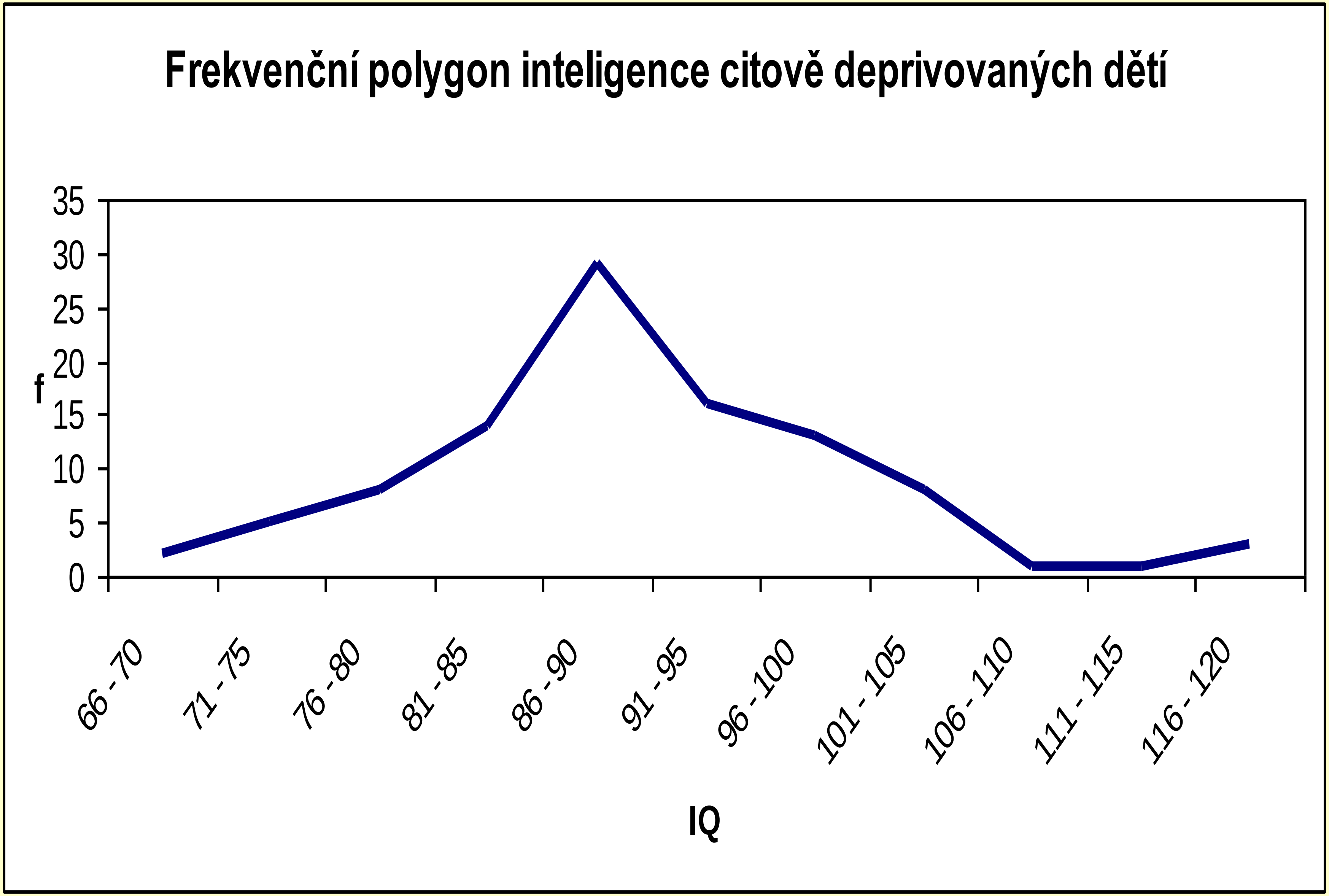
sledovaného statistického znaku.

2) (FREKVENČNÍ) POLYGON

Forma grafického znázornění rozdělení četností, kdy *místo sloupců* použijeme ke znázornění rozdělení četností *lomenou čáru.*

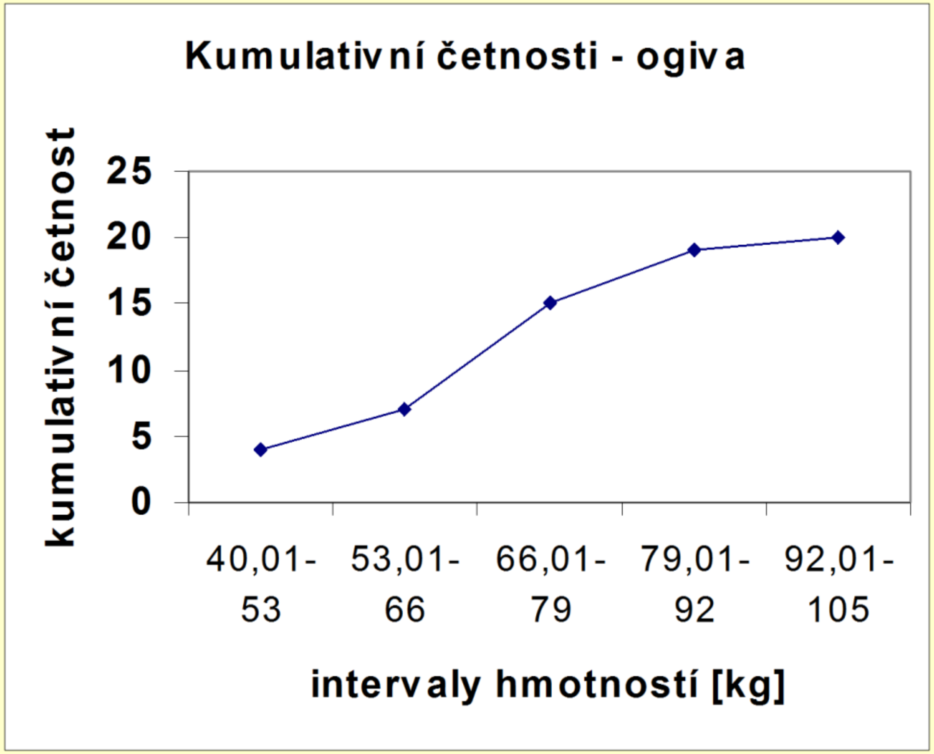
Tato lomená čára je *spojnice bodů* vytvořených v průsečících *středů intervalů* a *příslušných četností.*

2) (FREKVENČNÍ) POLYGON



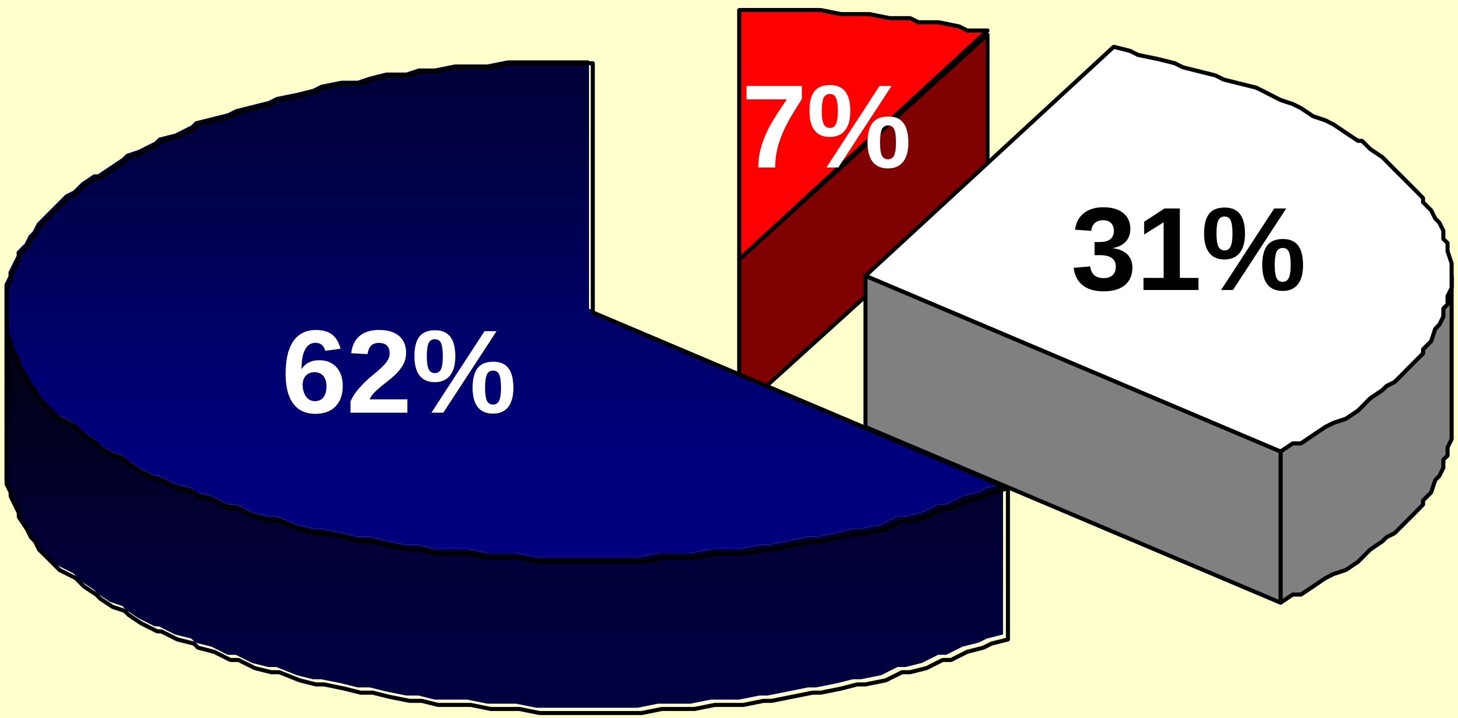
3) (GALTONOVA) OGIVA

Pojem *ogival* je v architektuře používán pro lomený oblouk, ve statistice tento pojem charakterizuje *esovitě lomenou křivku* znázorňující *kumulativní četnosti* *(absolutní nebo relativní)*.

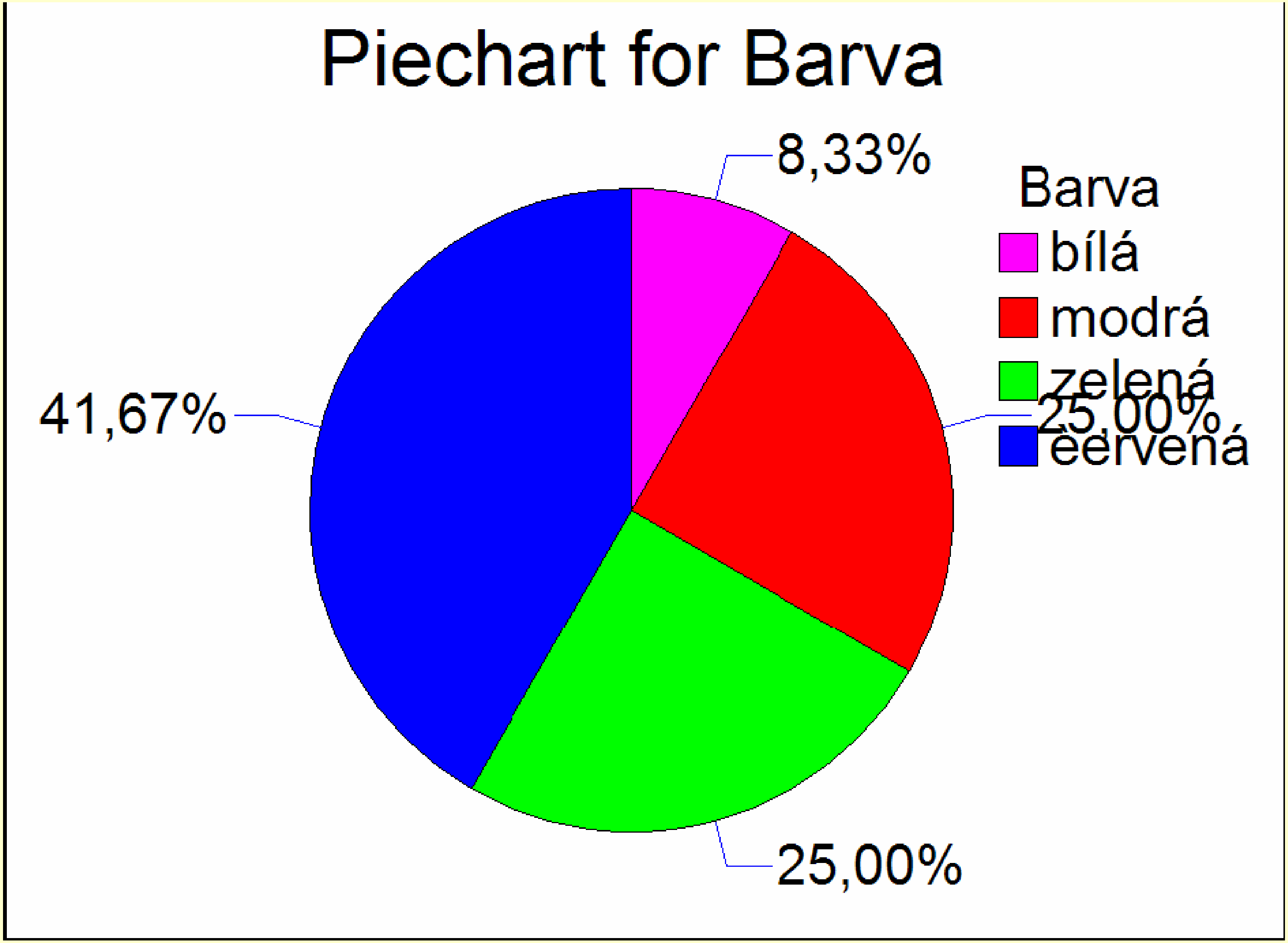


4) VÝSEČOVÝ (SEKTOROVÝ) GRAF

Jedná se o *kruhový graf*, vyjadřující *relativní četnosti* jako charakteristiku struktury daného souboru (nejčastěji v %).



4) VÝSEČOVÝ (SEKTOROVÝ) GRAF

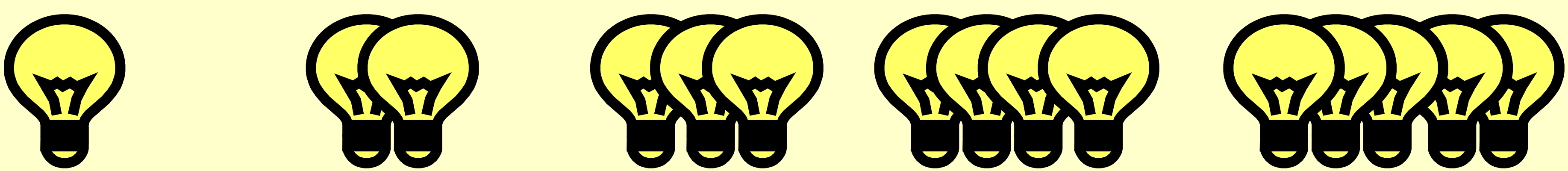


5) PIKTOGRAM

Piktogram = grafický znak znázorňující ***pojem*** nebo ***sdělení*** obrazově (např. dopravní značky), též **piktograf.** Vyjadřuje **absolutní četnosti** bez nároků na přesnost, má spíše informativní charakter a používá obrazových symbolů (např. lokomotiva, váček s penězi, postava vojáka).

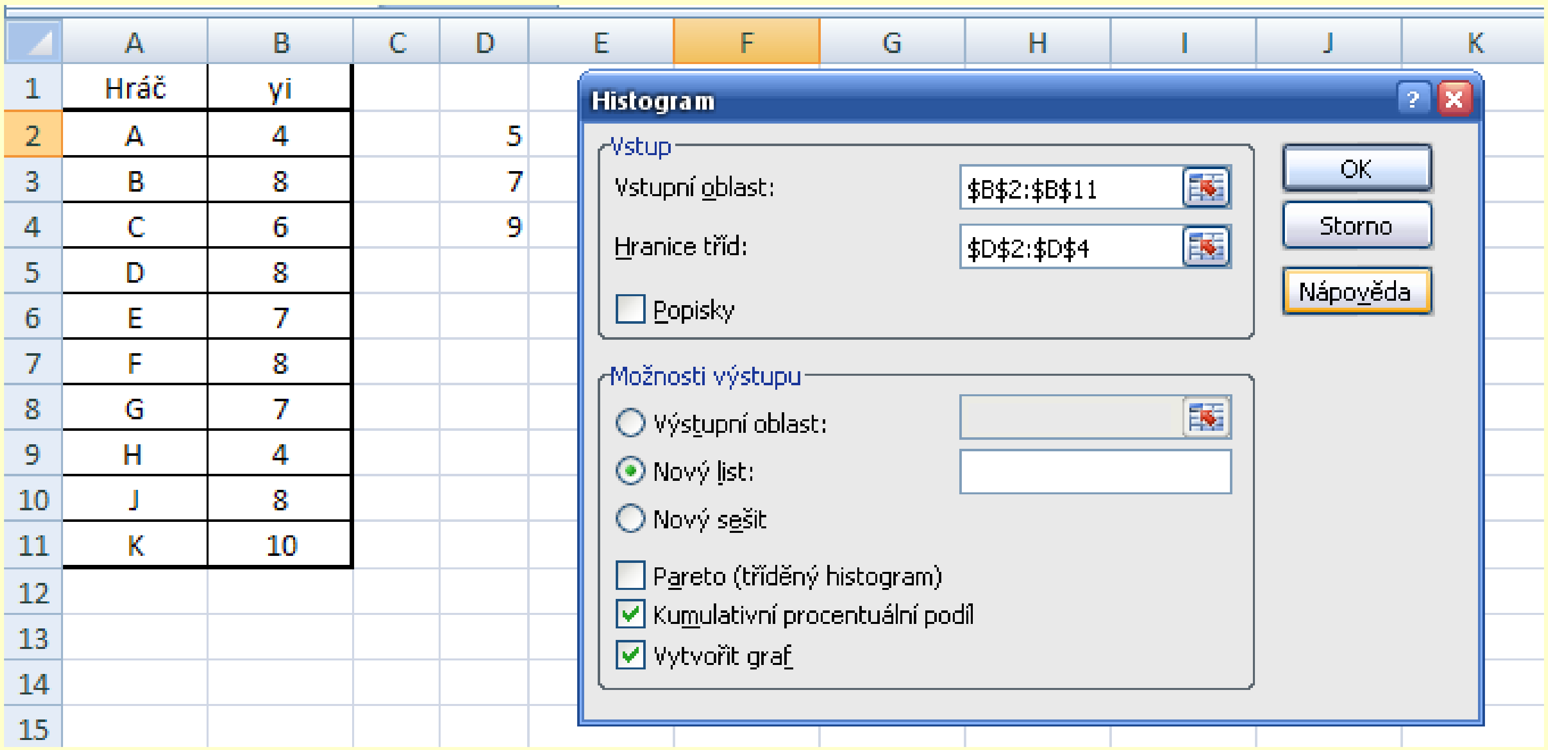
Spotřeba energie v městě X v letech

1960 1970 1980 1990 2000



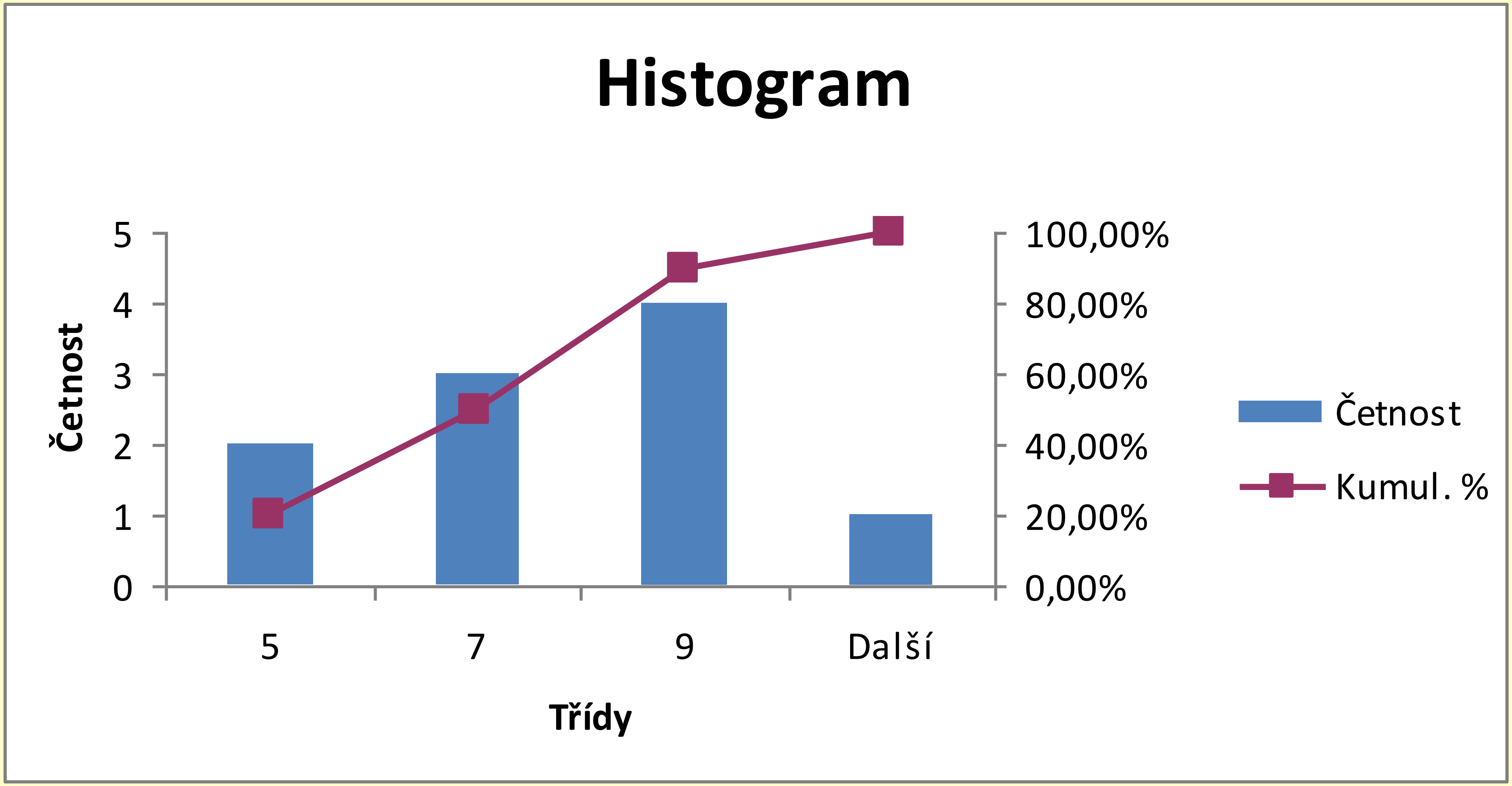
10 MW 22 MW 28 MW 43 MW 52 MW

Pro znaky y sestavte tabulku skupinového (intervalového) rozdělení četností.

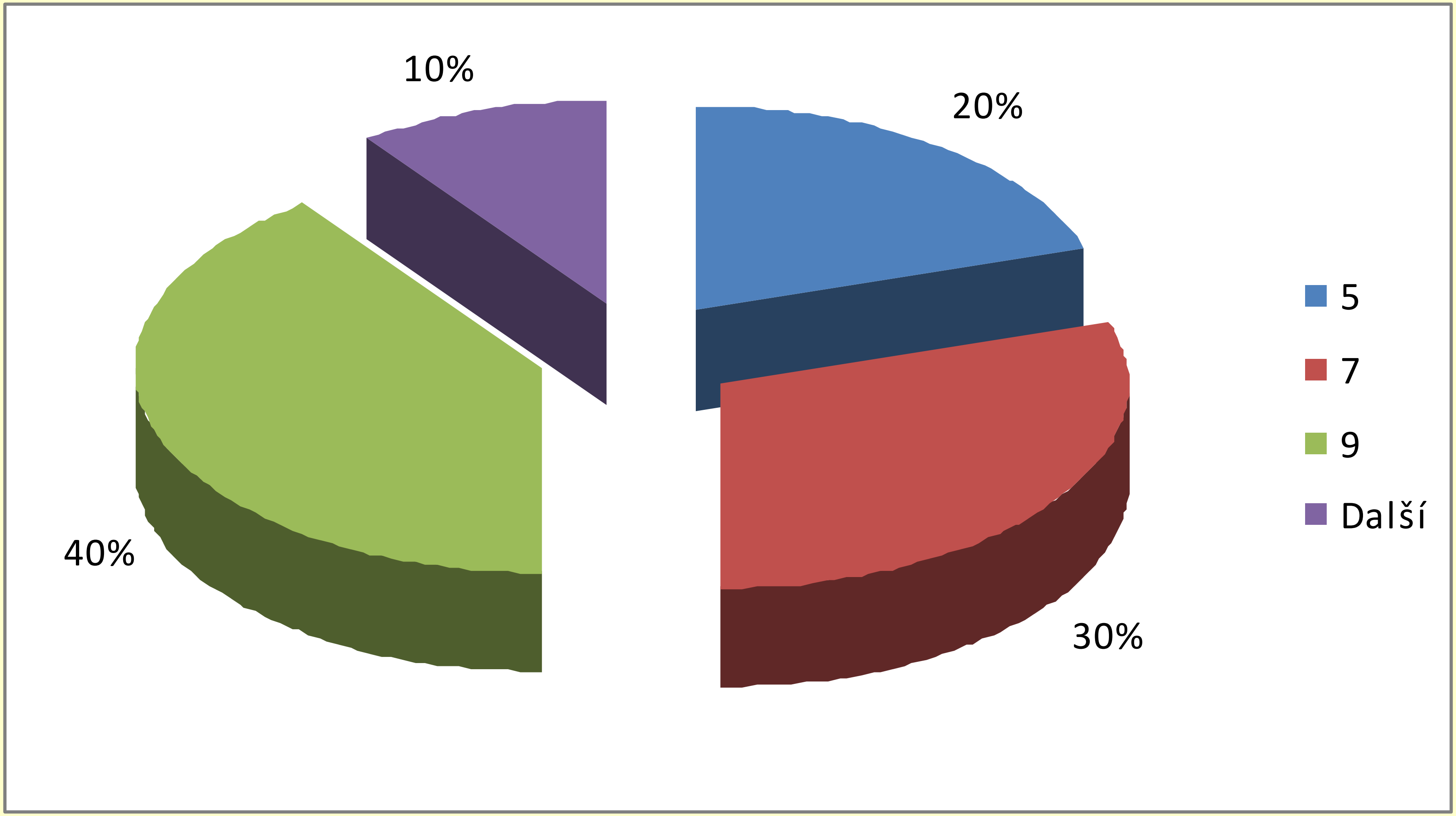


Třídy	*Četnost*	*Kumul. %*
5	2	20,00%
7	3	50,00%
9	4	90,00%
Další	1	100,00%

Histogram četností



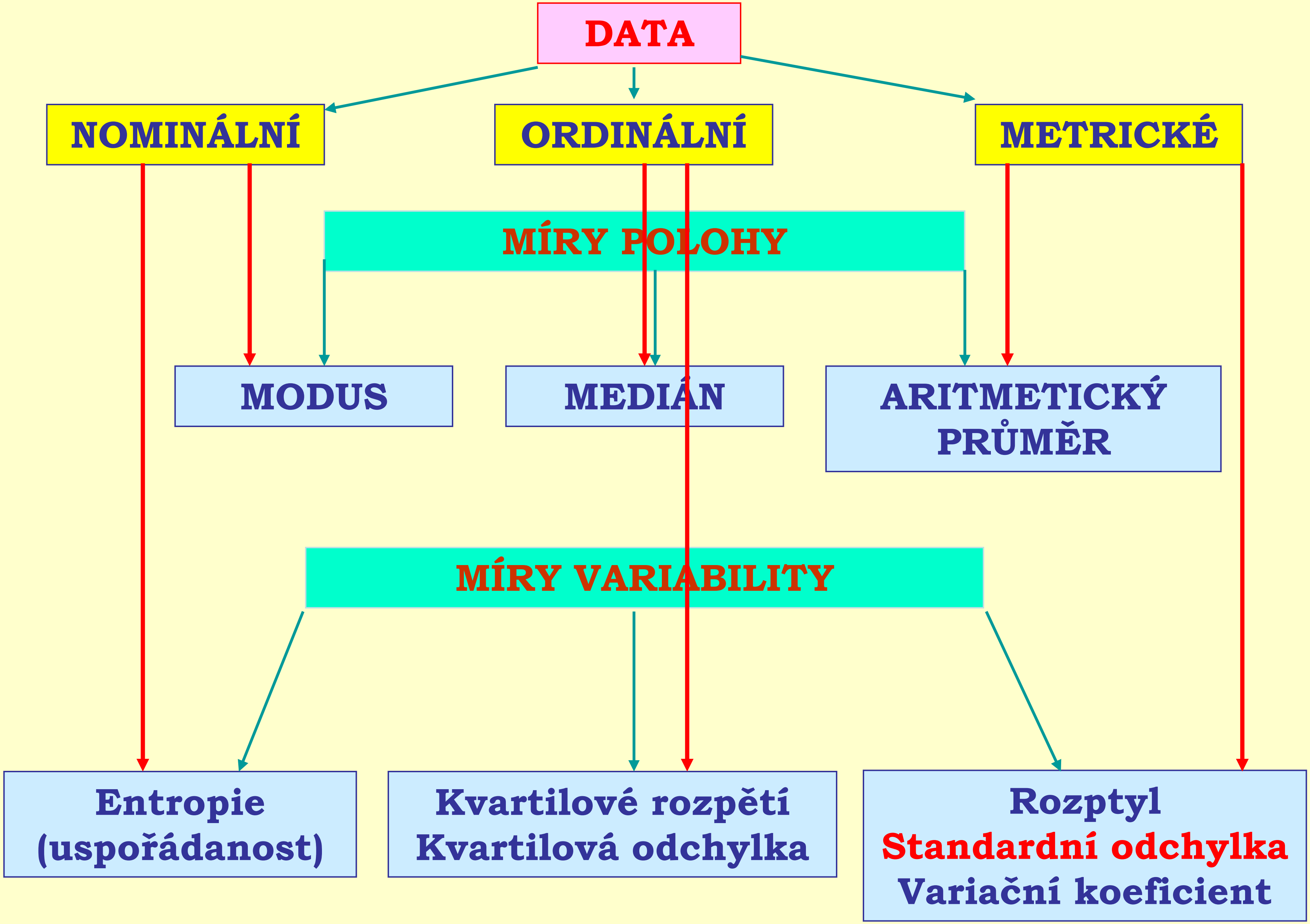
Výsečový (sektorový) graf



ZÁKLADNÍ STATISTICKÉ CHARAKTERISTIKY

PRO DATA ZÍSKANÁ NA ŠKÁLE

NOMINÁLNÍ, ORDINÁLNÍ, METRICKÉ


DATA
NOMINÁLNÍ
ORDINÁLNÍ
METRICKÉ
MÍRY POLOHY
MODUS
MEDIÁN
ARITMETICKÝ PRŮMĚR
MÍRY VARIABILITY
Entropie (uspořádanost)
Kvartilové rozpětí
Kvartilová odchylka
Rozptyl
Standardní odchylka
Variační koeficient

2. 2 MÍRY POLOHY

Míry polohy **(neboli *míry centrální tendence*)**

- **charakterizují *úroveň* statistického souboru z hlediska *jeho střední hodnoty,***

- **zevšeobecňují, zastupují, reprezentují jednotlivé hodnoty sledovaného statistického znaku,**

- **umožňují srovnání polohy dvou či více rozdělení četností, resp., srovnání střední úrovně dvou či více souborů.**

Hod na koš (n=10): 6, 7, 6, 7, 8, 9, 8, 8, 8, 9, 10

NEJČASTĚJI POUŽÍVANÉ MÍRY POLOHY

1. NOMINÁLNÍ STUPNICE (DATA)

MODUS (Mo) označuje *nejčastěji se vyskytující* hodnotu statistického souboru (hodnota s největší četností).

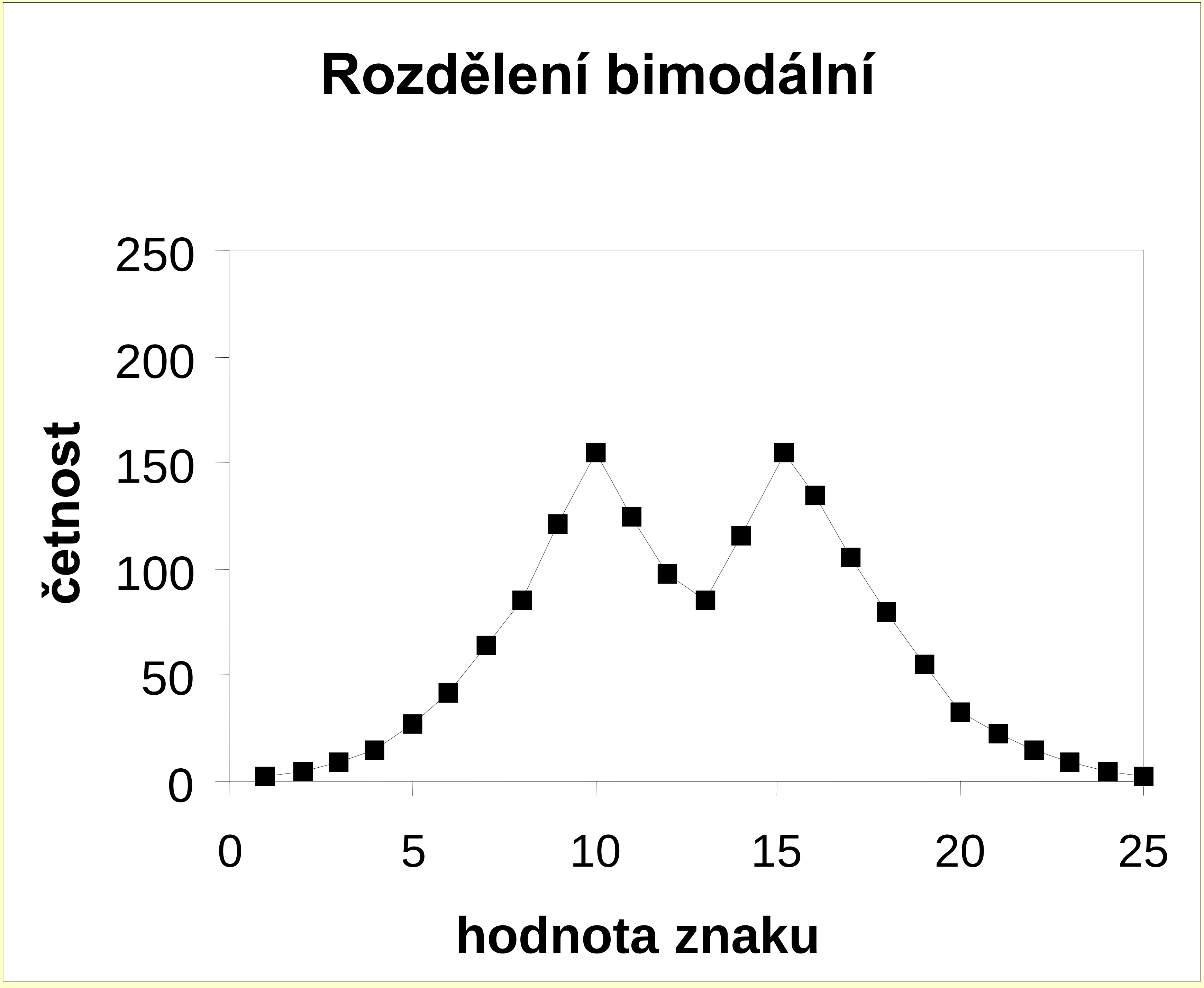
Modus je nejsnáze zjistitelná míra polohy.

Soubor může mít jeden či více modů

(soubor bimodální, resp. trimodální).

Modus je použitelný pro nominální stupnice

(a všechny vyšší).


Rozdělení bimodální
250
200
150
100
50
0
četnost
0
5
10
15
20
25
hodnota znaku

2. ORDINÁLNÍ STUPNICE (DATA)

MEDIÁN (Me) označuje *prostřední člen variační řady* (dělí výsledky seřazené podle velikosti na polovinu).

Medián není citlivý na velikost krajních hodnot.

Medián je použitelný pro ordinální stupnici (a vyšší).

Ukázka výpočtu pro sudý a lichý počet dat:

x_i : **6 7 7 8 8 8 8 9 9 10** (sudý počet, n = 10)

Mo = 8 Me = 8

y_i : **6 7 7 8 8 8 8 9 9 10 10** (lichý počet, n = 11)

3. METRICKÉ STUPNICE (DATA)

ARITMETICKÝ PRŮMĚR $\bar{x}$ (Mean, M) nejpoužívanější míra polohy, použitelný (pouze!) pro metrické škály.

Výpočet:* *součet všech hodnot statistického souboru dělený rozsahem souboru (n).

a) Aritmetický průměr prostý (jednoduchý) $\bar{x}$

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

x - statistický znak n - rozsah souboru
x_i - hodnota statistického znaku

Ø … takto nikdy !

b) Vážený aritmetický průměr

- **Používá se u početnějších souborů, výpočet vychází z rozdělení četností,**
- ***vážený* se nazývá proto, že jednotlivým hodnotám znaku je přisuzována *váha* odpovídající počtu výskytů.**

$$\bar{x} = \frac{x_1 w_1 + x_2 w_2 + \ldots + x_m w_m}{w_1 + w_2 + \ldots + w_m} = \frac{\sum_{i=1}^{m} x_i w_i}{\sum_{i=1}^{m} w_i}$$

w_i … váha (počet výskytů)

n … rozsah souboru (počet hodnot).

$$n = \sum_{i=1}^{m} w_i$$

b) Vážený aritmetický průměr – příklad využití

Přijímací řízení FSpS 2015–2017

Výsledky testu běh na 100m

2015 (n = 350)	M = 13,0
2016 (n = 230)	M = 12,5
2017 (n = 120)	M = 12,0

Jaký je průměrný výkon v běhu na 100 m v letech 2015–2017?

??? 13,0 + 12,5 + 12,0 = 37,5/3 = 12,5 ??? NE!

$$\bar{x} = \frac{x_1 w_1 + x_2 w_2 + \ldots + x_m w_m}{w_1 + w_2 + \ldots + w_m} = \frac{\sum_{i=1}^{m} x_i w_i}{\sum_{i=1}^{m} w_i}$$

$$\bar{x} = \frac{13{,}0 \text{ x } 350 + 12{,}5 \text{ x } 230 + 12{,}0 \text{ x } 120}{350 + 230 + 120} = 12{,}7$$

RUHR-UNIVERSITÄT BOCHUM

ZeDI Zentrum für Diagnostik und Intervention im Sport

Věk (tenis): 13 let (muži, n=104), 14 let (muži, n=98)

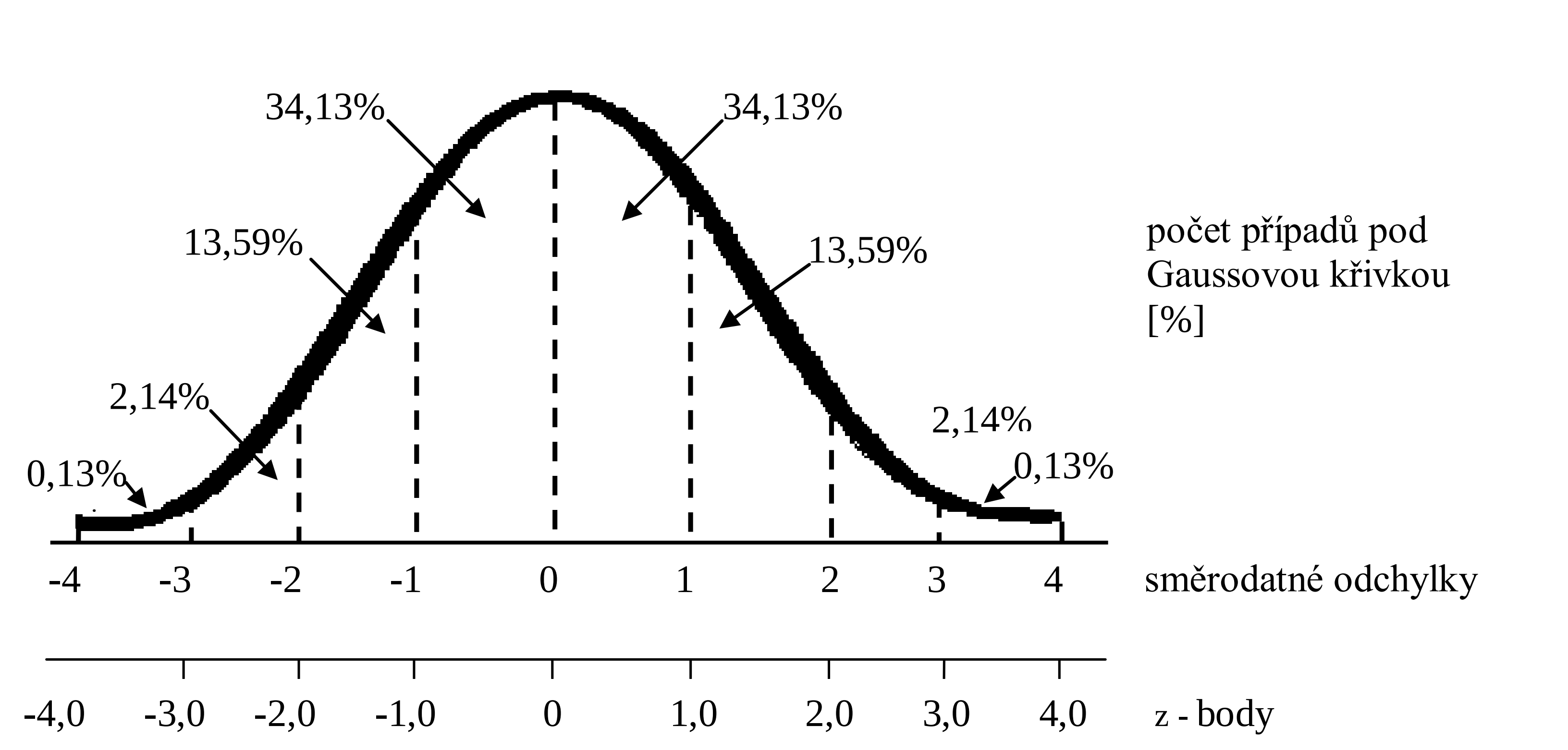
Znak	Výška	Hmotnost	BMI	HGSD	HGSN
M (104)	165,6	51,2	18,5	31,2	26,8
M (98)	173,4	59,6	19,7	37,0	31,8
M	*169,5*	*55,4*	*19,1*	*34,1*	*29,3*
M_V	**169,38**	**?**	**?**	**?**	**?**

Vypočet váženého aritmetického průměru (výška):

M_V = 165,6 x 104 + 173,4 x 98/104 + 98 = 17222,4 + 16993,2/202 = **169,38**

Poznámky k rozložení četností a měr polohy

- **Mají-li data prokazatelně normálním (Gaussovo) rozložení četností znaků jsou vypočítané střední hodnoty** (aritmetický průměr, modus, medián) **přibližně stejně velké.**
- **Čím více se střední hodnoty liší, tím více je rozložení asymetrické** (nejde o normální rozložení četností).



PŘÍKLAD 3

Výpočet: modus, medián, aritmetický průměr.

Hráč	A	B	C	D	E	F	G	H	J	K
x_i	7	6	7	8	9	8	8	8	9	10
y_i	4	8	6	8	7	8	7	4	8	10

Variační řada znaku x_i: 6, 7, 7, 8, 8, 8, 8, 9, 9, 10

Mo = ? **Me = ?** **M = ?** **Vážený AP = ?**

Mo = 8 **Me = 8** **M = 8** **Vážený AP = 8**

SAMI: Variační řada znaku y_i: 4, 8, 6, 8, 7, 8, 7, 4, 8, 10

Mo = ? **Me = ?** **M = ?** **Vážený AP = ?**

Mo = 8 **Me = 7,5** **M = 7** **Vážený AP = 7**

Pomocí Excelu – Statistické funkce

Výpočet: modus, medián, aritmetický průměr.

	A	B	C
1	Hráč	yi	
2	A	4	
3	B	8	
4	C	6	
5	D	8	
6	E	7	
7	F	8	
8	G	7	
9	H	4	
10	J	8	
11	K	10	
12			
13			
14			
15			

MODE

MEDIAN

PRŮMĚR

Argumenty funkce

MODE

Číslo1 B2:B11 = {4|8|6|8|7|8|7|4|8|10}

Číslo2 = matice

= 8

Vrátí hodnotu, která se v matici nebo v oblasti dat vyskytuje nejčastěji.

Číslo1: číslo1;číslo2;... je 1 až 255 čísel nebo názvů, matic či odkazů obsahujících čísla, jejichž modus chcete zjistit.

Výsledek = 8

Nápověda k této funkci

OK Storno

Argumenty funkce

MEDIAN

Číslo1 B2:B11 = {4|8|6|8|7|8|7|4|8|10}

Číslo2 = číslo

= 7,5

Vrátí medián, střední hodnotu množiny zadaných čísel.

Číslo1: číslo1;číslo2;... je 1 až 255 čísel, názvů, matic nebo odkazů ob pro která chcete nalézt medián.

Výsledek = 7,5

Nápověda k této funkci

OK

Argumenty funkce

PRŮMĚR

Číslo1 B2:B11 = {4|8|6|8|7|8|7|4|8|10}

Číslo2 = číslo

= 7

Vrátí průměrnou hodnotu (aritmetický průměr) argumentů. Argumenty mohou být čísla či názvy, matice nebo odkazy, které obsahují čísla.

Číslo1: číslo1;číslo2;... je 1 až 255 číselných argumentů, jejichž průměrnou hodnotu chcete zjistit.

Výsledek = 7

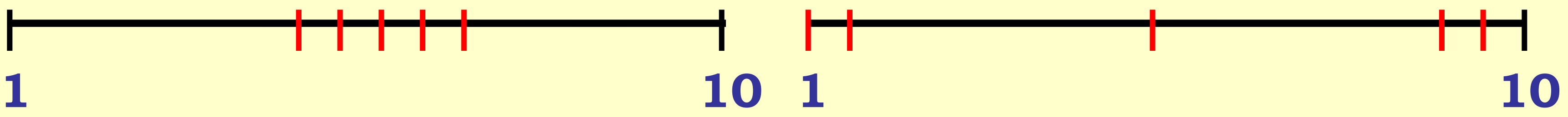
Nápověda k této funkci

OK Storno

2. 3 MÍRY VARIABILITY

Popis statistického souboru pomocí měr polohy (výpočet středních hodnot) **není dostačující - viz příklad!**

Př. 1: 3,4,5,6,7 ⇒ 25/5=5 (M=5) **Př. 2: 1,2,5,8,9 ⇒ 25/5=5 (M=5)**

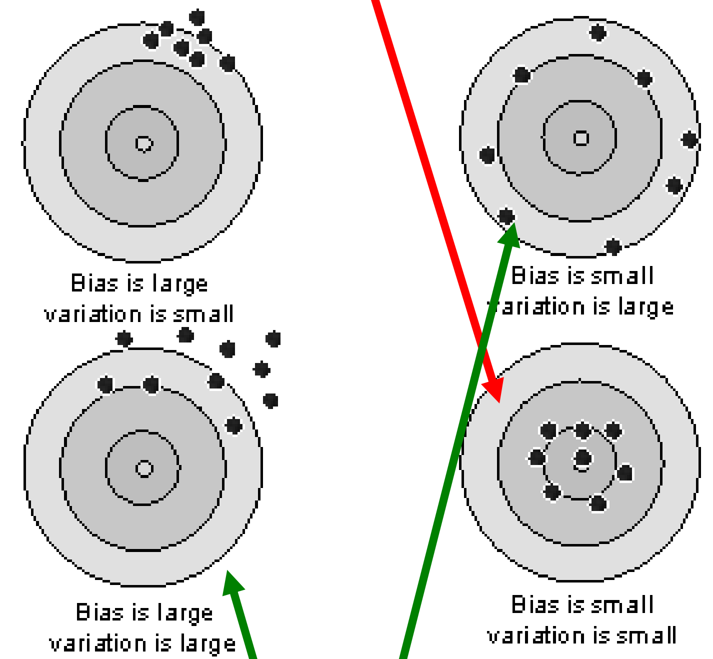


Míry variability *(neboli míry variace, rozptýlení, měnlivosti)* **charakterizují, jak se zkoumané znaky vzájemně odlišují,**

Tedy, je-li zkoumaný soubor

- ✓ **homogenní (stejnorodý),**
- ✓ **heterogenní (nestejnorodý, různorodý).**

Soubor homogenní



Accuracy versus Quality of an Estimator Using Bias and Variation as Measurable Quantities Respectively

Soubor heterogenní

2. 3. 1 KVANTILOVÉ MÍRY VARIABILITY

Využitelné pro stupnici ordinální,
pro stupnice metrické, když nelze prokázat normalitu rozložení četností dat.

VARIAČNÍ ŘADA = znaky statistického souboru seřazené podle velikosti.

VARIAČNÍ ROZPĚTÍ = diference mezi největší a nejmenší hodnotou znaku souboru. $R = x_{max} - x_{min}$

KVANTIL = hodnota kvantitativního statistického znaku, která rozděluje (láme) variační řadu na jisté části (***kvartil, decil, procentil***).

PŘÍKLAD 4 Výpočet: variační řada, variační rozpětí.

Hráč	A	B	C	D	E	F	G	H	J	K
x_i	7	6	7	8	9	8	8	8	9	10
y_i	4	8	6	8	7	8	7	4	8	10

VARIAČNÍ ŘADA znaků $x_i \Rightarrow$ 6,7,7,8,8,8,8,9,9,10

VARIAČNÍ ROZPĚTÍ $R = x_{max} - x_{min} \Rightarrow R = 10 – 6 = 4$

Totéž si sami vypočítat v přednášce pro znaky y_i

VARIAČNÍ ŘADA znaků $y_i \Rightarrow$ 4,4,6,7,7,8,8,8,8,10

VARIAČNÍ ROZPĚTÍ $R = x_{max} - x_{min} \Rightarrow R = 10 – 4 = 6$

DRUHY KVANTILŮ (kvartil, decil, percentil)

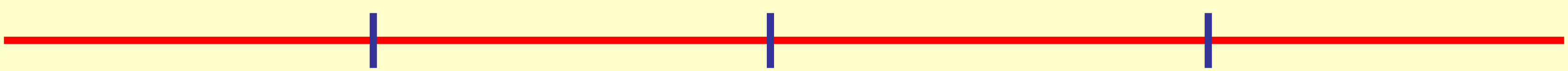
1. KVARTIL (Y) … kvartily rozdělují variační řadu na čtvrtiny, na 4 skupiny.

KOLIK MÁME KVARTILŮ?

Dolní kvartil (Q_1, x_{25})

Horní kvartil (Q_3, x_{75})

(Střední kvartil) = medián



VÝPOČET KVARTILU

$$z_p = \frac{n \cdot p}{100} + 0{,}5$$

z_p - **pořadí kvantilu** x_p

n - rozsah souboru p - kvartil

Příklad : Určete dolní kvartil x_{25}, jestliže rozsah souboru je n = 40

$$z_p = \frac{40 \cdot 25}{100} + 0{,}5 = 10{,}5 \qquad x_{25} = \frac{x_{(10)} + x_{(11)}}{2}$$

Výsledek 10,5 znamená, že dolní kvartil x_{25} je průměrem desáté a jedenácté hodnoty variační řady znaků souboru.

VÝPOČET KVARTILU

$$z_p = \frac{n \cdot p}{100} + 0{,}5$$

z_p - pořadí kvantilu x_p

n - rozsah souboru p - kvartil

Příklad (basketbal) :

Určete dolní (horní) kvartil x_{25} (x_{75}), jestliže rozsah souboru je n = 10 (6,7,7,8,8,8,8,9,9,10)

z_{25} = 10 x 25/100 + 0,5 = 2,5 + 0,5 = 3,0

z_{75} = 10 x 75/100 + 0,5 = 7,5 + 0,5 = 8,0

Výsledek 3,0 znamená, že dolní kvartil x_{25} je <u>třetí</u> (<u>osmá</u>) hodnota variační řady znaků souboru, tedy x_{25} = 7 (x_{75} = 9)

VÝPOČET KVARTILU

$$z_p = \frac{n \cdot p}{100} + 0{,}5$$

z_p - pořadí kvantilu x_p

n - rozsah souboru p - kvartil

Příklad (basketbal):

určete dolní a horní kvartil (x_{25}; x_{75}),

souboru n_i = 4;8;6;8;7;8;7;4;8;5;6;5

dolní kvartil x_{25} = ?

horní kvartil x_{75} = ?

4;4;5;5;6;6;7;7;8;8;8;8

2. DECIL

… decily rozdělují variační řadu na desetiny, tedy na 10 skupin o 10% rozsahu souboru.

Označují se x_{10}, x_{20}, …x_{90}

3. PERCENTIL (PROCENTIL)

… percentily rozdělují variační řadu na setiny, na 100 skupin o 1% rozsahu.

Označují se x_1, x_2, …x_{99}

DALŠÍ KVANTILOVÉ CHARAKTERISTIKY VARIABILITY

KVANTILOVÉ ROZPĚTÍ

- **kvartilové rozpětí** $x_{75} - x_{25}$
- **decilové rozpětí** $x_{90} - x_{10}$
- **percentilové rozpětí** $x_{99} - x_{1}$

KVANTILOVÉ ODCHYLKY

a) kvartilová odchylka

$$Q = \frac{x_{75} - x_{25}}{2}$$

Je polovinou rozpětí krajních hodnot, není ovlivněna jejich extrémy.

b) decilová odchylka

$$D = \frac{x_{90} - x_{10}}{8}$$

Je osminou rozpětí krajních decilů, záleží tedy na rozpětí prostředních 80% prvků souboru.

c) percentilová odchylka

$$C = \frac{x_{99} - x_{1}}{98}$$

Je devadesáti osminou rozpětí krajních percentilů.

2.3.2 MOMENTOVÉ MÍRY VARIABILITY

Předchozí „kvantilové míry variability“ udávají jen rozpětí, v němž se znaky pohybují.

MOMENTOVÉ MÍRY VARIABILITY umožňují výpočet číselných charakteristik,

které umožňují zjistit:

(1) variaci (rozptýlení) ve smyslu vzájemné odlišnosti jednotlivých hodnot znaku mezi sebou,

(2) variaci (rozptýlení) ve smyslu odlišnosti jednotlivých hodnot znaku od průměru.

NEJČASTĚJI POUŽÍVANÉ MOMENTOVÉ MÍRY VARIABILITY

1. ROZPTYL

$$s^2 = \frac{\sum (x_i - M)^2}{n}$$

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

(pro rozsáhlé soubory)

M ... aritmetický průměr

x i ...hodnota znaku

Rozptyl „měří“ variaci ve smyslu odlišnosti jednotlivých hodnot znaku *od průměru.*

Rozptyl (s^2) je aritmetickým průměrem ze čtverců odchylek jednotlivých hodnot znaku od jejich aritmetického průměru *(nepožadováno).*

2. SMĚRODATNÁ (STANDARDNÍ) ODCHYLKA (s)

Symbolický tvar

s = √ s² (var x)

$$s = \sqrt{\frac{\sum (x_i - M)^2}{n}}$$

M … aritmetický průměr

x i …hodnota znaku

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

(pro rozsáhlé soubory)

Směrodatná odchylka (s) je kvadratickým průměrem odchylek jednotlivých hodnot znaku od aritmetického průměru *(nepožadováno).*

3. VARIAČNÍ KOEFICIENT (Coefficient of variation, CV)

$$\mathbf{VK} = \frac{\mathbf{s}}{\mathbf{M}} \qquad \mathbf{VK\ (\%)} = \mathbf{100\ x} \frac{\mathbf{s}}{\mathbf{M}}$$

(s = směrodatná odchylka; M = aritmetický průměr)

- **umožňuje provést *srovnání variability dvou či více souborů*, jejichž znaky jsou měřeny v různých jednotkách (cm, kg, sek; *vysvětlení slide 112*),**

- ***udává poměr směrodatné odchylky k aritmetickému průměru, tedy, kolik % aritmetického průměru tvoří směrodatná odchylka.***

PŘÍKLAD 5

Výpočet: rozptyl, směrodatná odchylka, variační koeficient

Hráč	A	B	C	D	E	F	G	H	J	K
x_i	7	6	7	8	9	8	8	8	9	10
y_i	4	8	6	8	7	8	7	4	8	10

(1) Rozptyl

M = 8

$$s^2 = \frac{(7-8)^2+(6-8)^2+(7-8)^2\ldots\ldots(8-8)^2+(9-8)^2+(10-8)^2}{10}$$

$$= \frac{1+4+1+0+1+0+0+0+1+4}{10} = \frac{12}{10} = \mathbf{1{,}20}$$

(2) Směrodatná odchylka

$s = \sqrt{s^2} = 1{,}09 = 1{,}1$

Sami doma – směrodatná odchylka znaků y_i ...

(3) Variační koeficient VK_1

$$VK_1 = \frac{s}{M} = \frac{1{,}09}{8} = 0{,}14 \quad \text{resp.} \quad VK_1 = \frac{s}{M} \times 100 = 14\ \%$$

Sami doma - variační koeficient VK_2 tj. znaků y_i ...

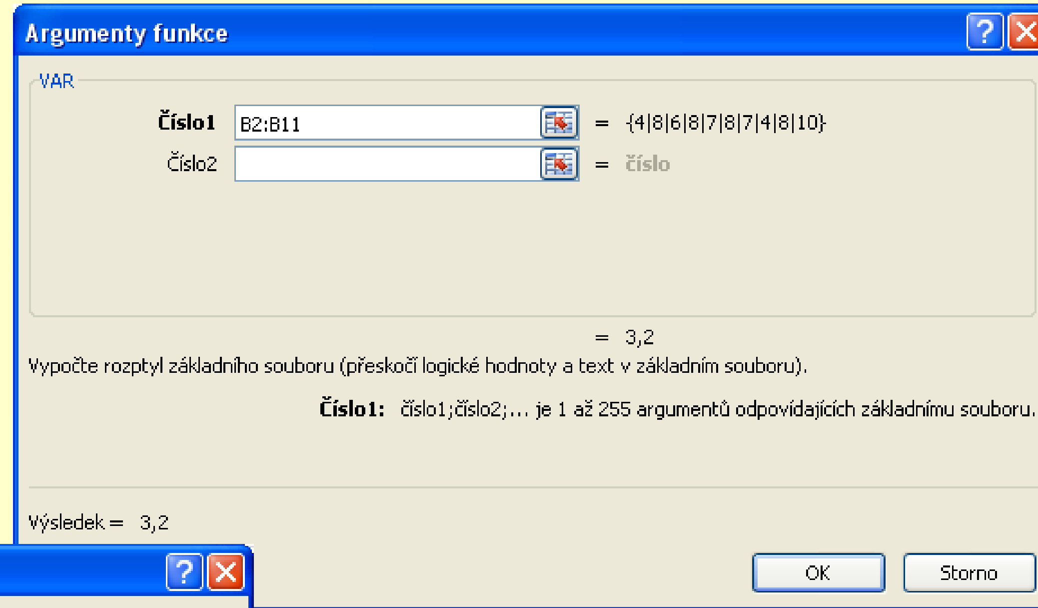
$VK_2 = 0{,}26$ resp. $VK_2 = 26\ \%$ $\Rightarrow$ $VK_1 < VK_2$ **Interpretace ...**

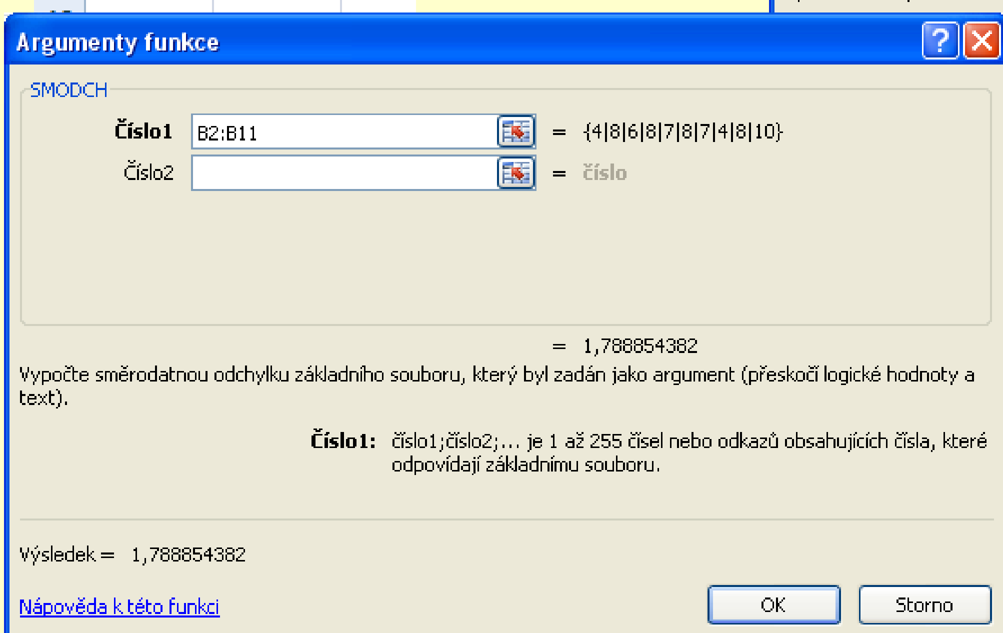
Pomocí Excelu – Statistické funkce
Výpočet: rozptyl, směrodatná odchylka, variační koeficient

	A	B	C
1	Hráč	yi	
2	A	4	
3	B	8	
4	C	6	
5	D	8	
6	E	7	
7	F	8	
8	G	7	
9	H	4	
10	J	8	
11	K	10	

VAR

SMODCH



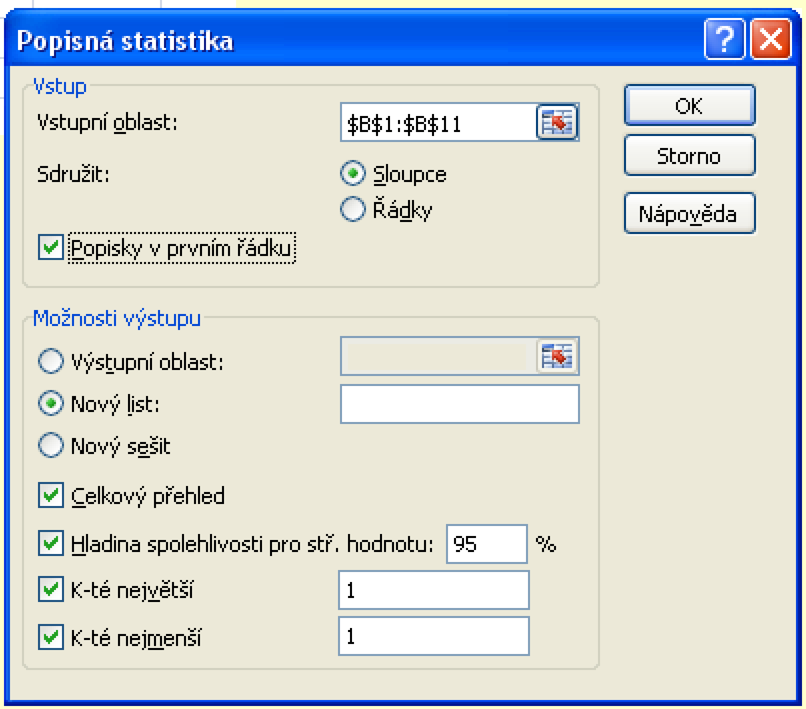


VAR.VÝBĚR
vypočte rozptyl výběru

SMODCH.VÝBĚR
vypočte směrodatnou odchylku výběru

Pomocí Excelu – Analýza dat – Popisná statistika

	A	B
1	Hráč	yi
2	A	4
3	B	8
4	C	6
5	D	8
6	E	7
7	F	8
8	G	7
9	H	4
10	J	8
11	K	10



	yi
Stř. hodnota	7
Chyba stř. hodnoty	0,596285
Medián	7,5
Modus	8
Směr. odchylka	1,885618
Rozptyl výběru	3,555556
Špičatost	-0,05776
Šikmost	-0,49718
Rozdíl max-min	6
Minimum	4
Maximum	10
Součet	70
Počet	10
Největší (1)	10
Nejmenší (1)	4
Hladina spolehlivosti (95,0%)	1,34889

STATISTICKÁ ANALÝZA DAT

Základní statistické charakteristiky

Která proměnná vykazuje největší variabilitu dat?

Tab. 1: Základní statistické charakteristiky souboru tenistek U10 (n = 65)

Proměnné	M	SD	Min	Max	VK (%)
Věk	10,20	0,60	9,0	10,9	5,88
Výška (cm)	145,30	**7,50**	130,0	165,0	5,16
Hmotnost (kg)	36,76	6,10	25,8	53,0	16,59
Síla stisku (P)	18,90	4,82	11,0	36,6	25,50
Síla stisku (L)	16,70	5,03	9,1	39,2	**30,12**

Vysvětlivky: n = počet prvků souboru; M = aritmetický průměr; SD = směrodatná odchylka; Min = minimální hodnota; Max = maximální hodnota; VK = variační koeficient (%); P/L = pravá/levá ruka

METODY DESKRIPTIVNÍ STATISTIKY

1. URČENÍ TYPU ŠKÁLY (nominální, ordinální, metrické)

a) nominální + ordinální ⇒ *neparametrické stat. metody*

b) metrické ⇒ *parametrické statistické metody*

2. ROZLOŽENÍ ČETNOSTÍ ZNAKŮ (NORMÁLNÍ ČI JINÉ)

a) normální ⇒ *parametrické statistické metody*

b) jiné ⇒ *neparametrické statistické metody*

3. ZÁKLADNÍ STATISTICKÉ CHARAKTERISTIKY

a) míry centrální tendence

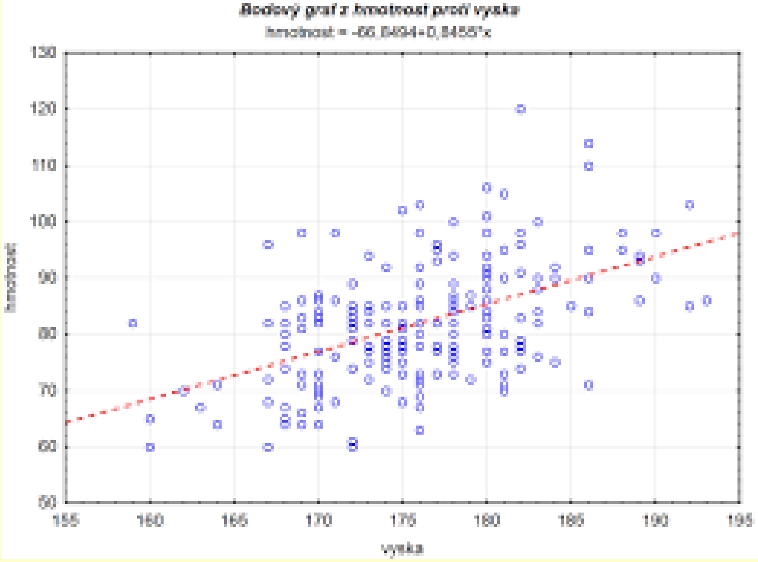
b) míry variability

c) míry závislosti

2.4 MÍRY ZÁVISLOSTI

2.4.1 ZÁVISLOST PEVNÁ & VOLNÁ (STATISTICKÁ)

Statistické soubory **jsou charakterizovány pomocí znaků**



Souvislost mezi znaky**:**

- **rozběhová rychlost x délka skoku,**
- **tělesná hmotnost x délka vrhu koulí,**
- **tělesná výška x hmotnost atd.**

2.5.1 ZÁVISLOST PEVNÁ & VOLNÁ

Míry závislosti se zabývají zkoumáním vztahů mezi dvěma (či více) statistickými znak.

Závislosti můžeme rozdělit na pevnou a volnou:

1) FUNKČNÍ (PEVNÁ) ZÁVISLOST (typická pro vztahy mezi proměnnými v oblasti exaktních věd). **Je výrazem pevného příčinného vztahu. Co to znamená?**

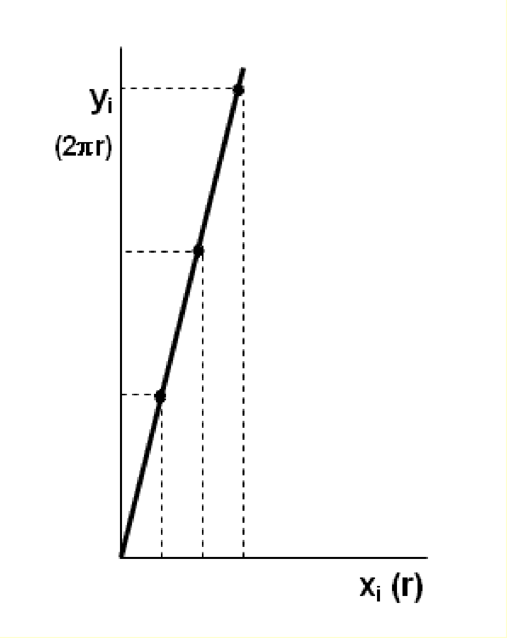
Každé číselné hodnotě jedné proměnné (x_i) odpovídá přesně jedna hodnota druhé proměnné (y_i).

*Jejich **vztah** lze přesně popsat určitou **rovnicí**, např. lineární, kvadratická, hyperbolická, logaritmická, atd.*

PEVNÁ (FUNKČNÍ) ZÁVISLOST

Funkční (pevná) závislost

- **může být tedy charakterizována jediným pozorováním** (větší počet slouží k ověření výsledků a vyloučení chyb),
- setkáváme se s ní při formulování zákonitostí vztahů mezi proměnnými, např. funkční závislosti mezi poloměrem kruhu a jeho obvodem ($y = 2 * \pi * r$), resp. fyzikální zákony (Archimedův).



VOLNÁ (STATISTICKÁ, KORELAČNÍ) ZÁVISLOST

2) VOLNÁ (statistická, korelační) závislost

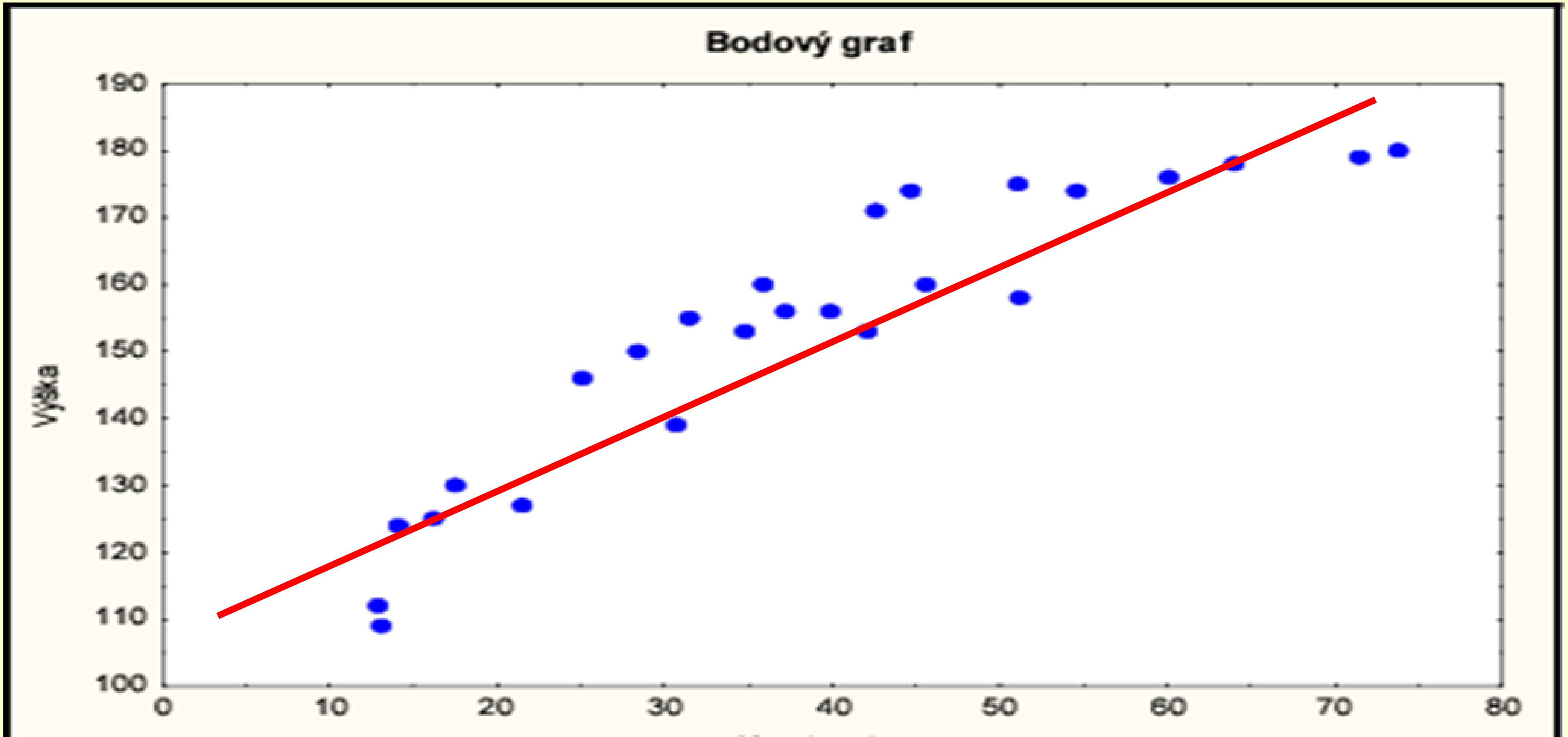
je typická pro vztahy mezi proměnnými (statistickými znaky) sledovanými v biologii, lékařství a dalších neexaktních vědách (např. sportovní vědy). Určitá příčina vede k různým účinkům.

Volnou závislost lze tedy zkoumat pouze na základě **mnoha pozorování, malý počet** pozorování může přinést **nahodilý výsledek** ovlivněný náhodnými faktory.

Příklad: rychlost rozběhu x délka skoku

VOLNÁ (statistická, korelační) závislost.

Výskyt **jednoho jevu** pouze **OVLIVŇUJE** výskyt **druhého jevu.** Každé hodnotě **jedné proměnné** (výška) odpovídají **různé hodnoty jiné proměnné** (hmotnost).



2.5.2 KORELAČNÍ POČET

(regresní a korelační analýza)

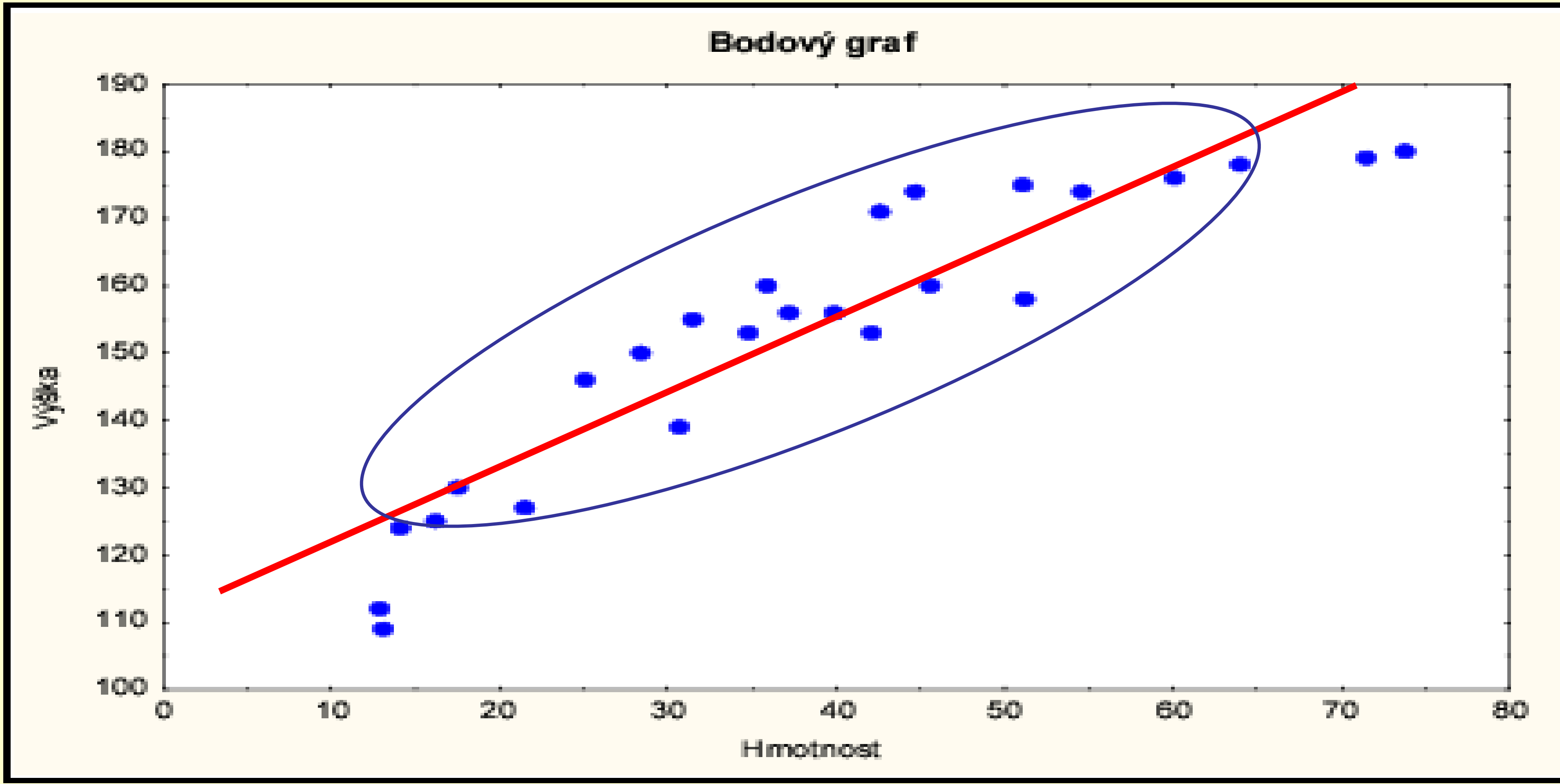
Metody regresní a korelační analýzy **slouží k poznání a matematickému popisu statistických závislostí; jsou souhrnně označované jako** ***korelační počet.***

Hlavní úkoly korelačního počtu:

1. ***postižení povahy korelační závislosti (regresní analýza),***
2. ***měření těsnosti korelační závislosti (korelační analýza).***

1. ***postižení povahy*** ***(regresní analýza),***

2. ***měření těsnosti*** ***(korelační analýza).***



HLAVNÍ ÚKOLY KORELAČNÍHO POČTU

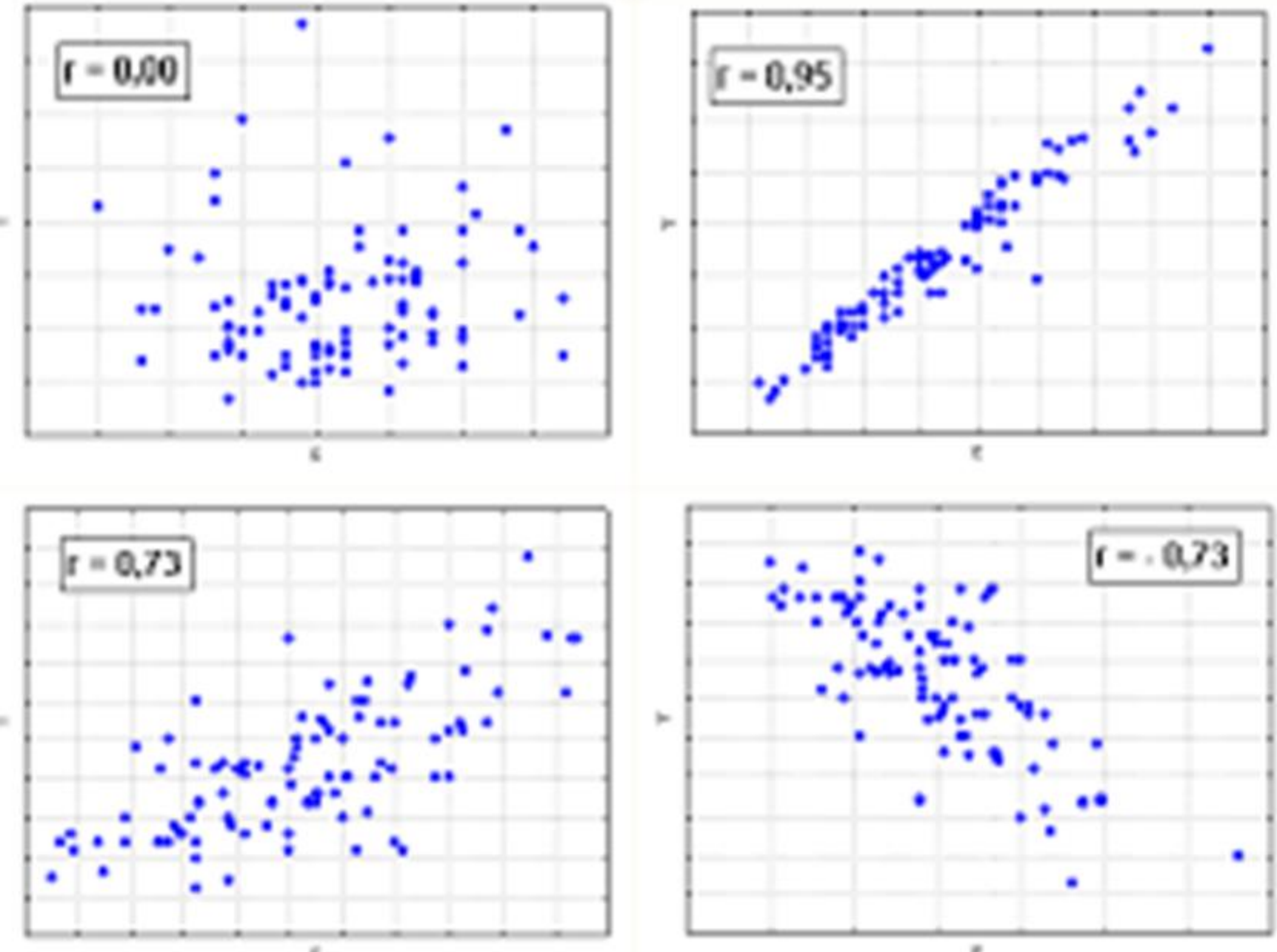
1. postižení povahy **korelační závislosti umožňuje odhady neznámých hodnot** ***závisle proměnné y*** **při známých hodnotách** ***nezávisle proměnné x.***

Povaha korelační závislosti je vyjadřována matematickou funkcí - hovoříme o ***regresní funkci (regresní analýza).***

2. měření těsnosti **korelační závislosti umožňuje posuzovat** ***míru korelační závislosti.***

Korelace je vyjadřována korelačním koeficientem r

- hovoříme o ***vlastní korelaci (korelační analýza).***


r = 0,00
r = 0,95
r = 0,73
r = - 0,73

1. REGRESNÍ ANALÝZA (LINEÁRNÍ)

Jen stručný úvod do problematiky

Regresní analýza **umožňuje** ***postihnout povahu*** **závislosti pomocí** ***regresní funkce*** **nejlépe vyjadřující zkoumané závislosti (je vyjádřena** ***regresní rovnicí*****).**

Regresní funkce **může nabývat mnoha typů:**

- ***přímková (lineární)***
- ***křivková (nelineární),*** např. hyperbolická, logaritmická, parabolická, exponenciální, a další …

LINEÁRNÍ REGRESNÍ FUNKCE je vyjádřena regresní rovnicí $\boxed{y = a + b \cdot x}$

y = závisle proměnná x = nezávisle proměnná

a, b = regresní koeficienty

Pro konstrukci regresní funkce pro konkrétní závislost (např. tělesná výšky a hmotnost) je třeba určit **regresní koeficienty a, b.**

Vycházíme z empirických (měřených) znaků TV a TH).

Vzorce pro výpočet regresních koeficientů a, b:

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \qquad a = \frac{\sum y_i - b \sum x_i}{n}$$

2. KORELAČNÍ ANALÝZA (LINEÁRNÍ)

Úkol: MĚŘENÍ TĚSNOSTI KORELAČNÍ ZÁVISLOSTI

Pojem ***korelace*** pochází z latiny (*co – relation =* souvztažnost), obvykle ji označujeme symbolem „ r “.

Korelace je definována jako volná kvantitativní závislost dvou či více jevů.

Korelace *vyjadřuje míru (stupeň) závislosti a je charakterizována* ***korelačním koeficientem r,*** *který* ***„měří“ těsnost závislosti*** *popsané regresní funkcí.*

VZORCE PRO VÝPOČET KORELAČNÍHO KOEFICIENTU

Symbolická podoba vzorce korelačního koeficientu

kovariance

$$r = \frac{s_{x,y}}{s_x \cdot s_y} = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}\, X \cdot \mathrm{var}\, Y}}$$

součin obou směrodatných odchylek

Korelace je matematicky

podíl kovariance a součinu obou směrodatných odchylek.

Pro **metrická data** (normalita)

PEARSONŮV KOEFICIENT SOUČINOVÉ KORELACE

(vzorec)

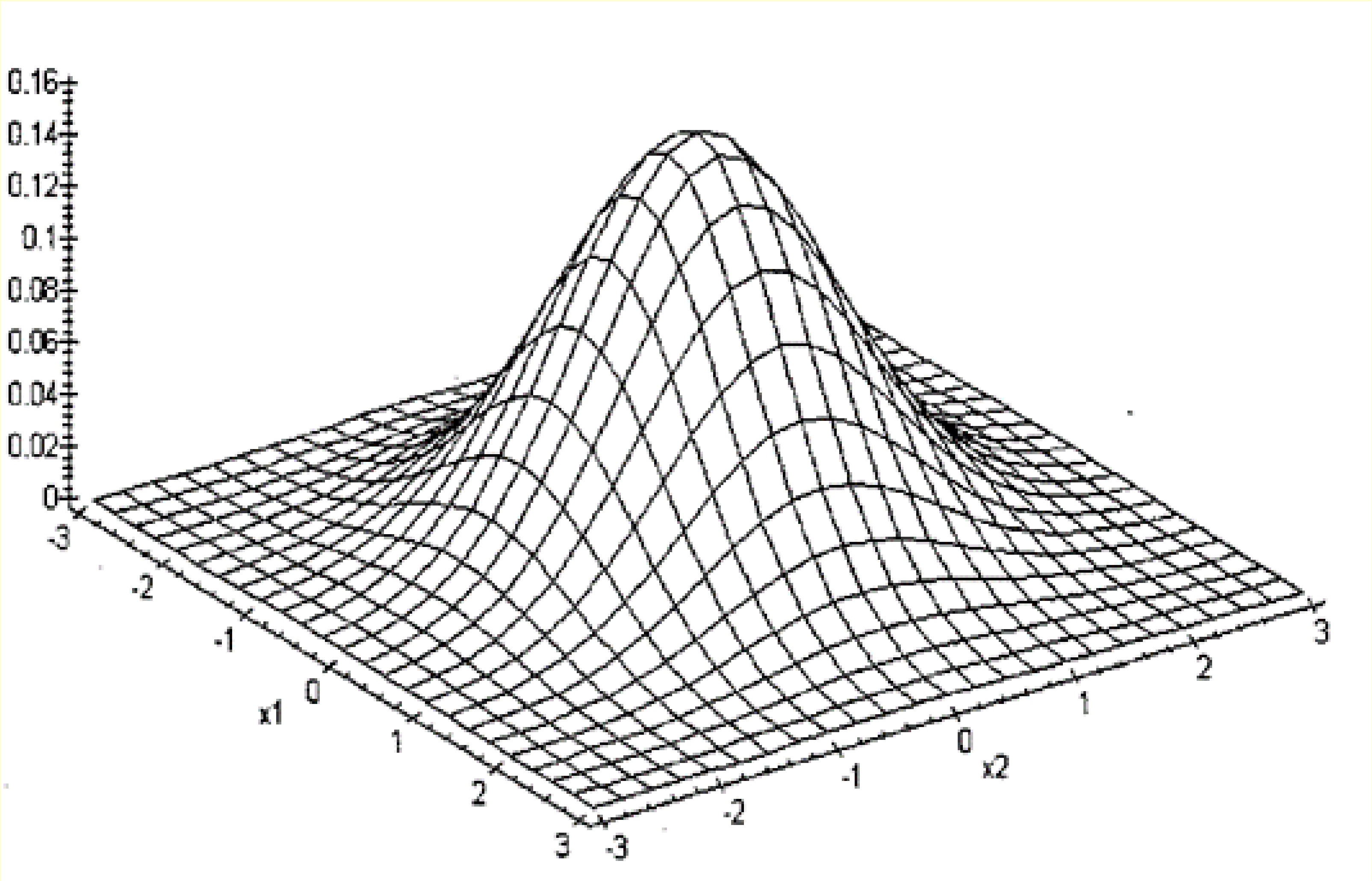
$$r = \frac{\sum (x_i - \bar{x}).(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Pearsonův koeficient - výpočtový tvar

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

PEARSONŮV KORELAČNÍ KOEFICIENT (r)

Podmínkou výpočtu je ověření *dvourozměného normálního rozdělení*



Pro ordinální data

SPEARMANŮV KOEFICIENT POŘADOVÉ KORELACE

(není požadováno normální rozložení četností)

$$r_{xy} = 1 - \frac{6 \cdot \sum (i_x - i_y)^2}{n\,(n^2 - 1)} = 1 - \frac{6}{n\,(n^2 - 1)} \sum d_i^{\,2}$$

VLASTNOSTI KORELACE

1. VELIKOST KORELACE

26.11.24

Korelační koeficient r nabývá hodnot z intervalu <-1 ; 1>

Význam hodnot -1, 0, 1

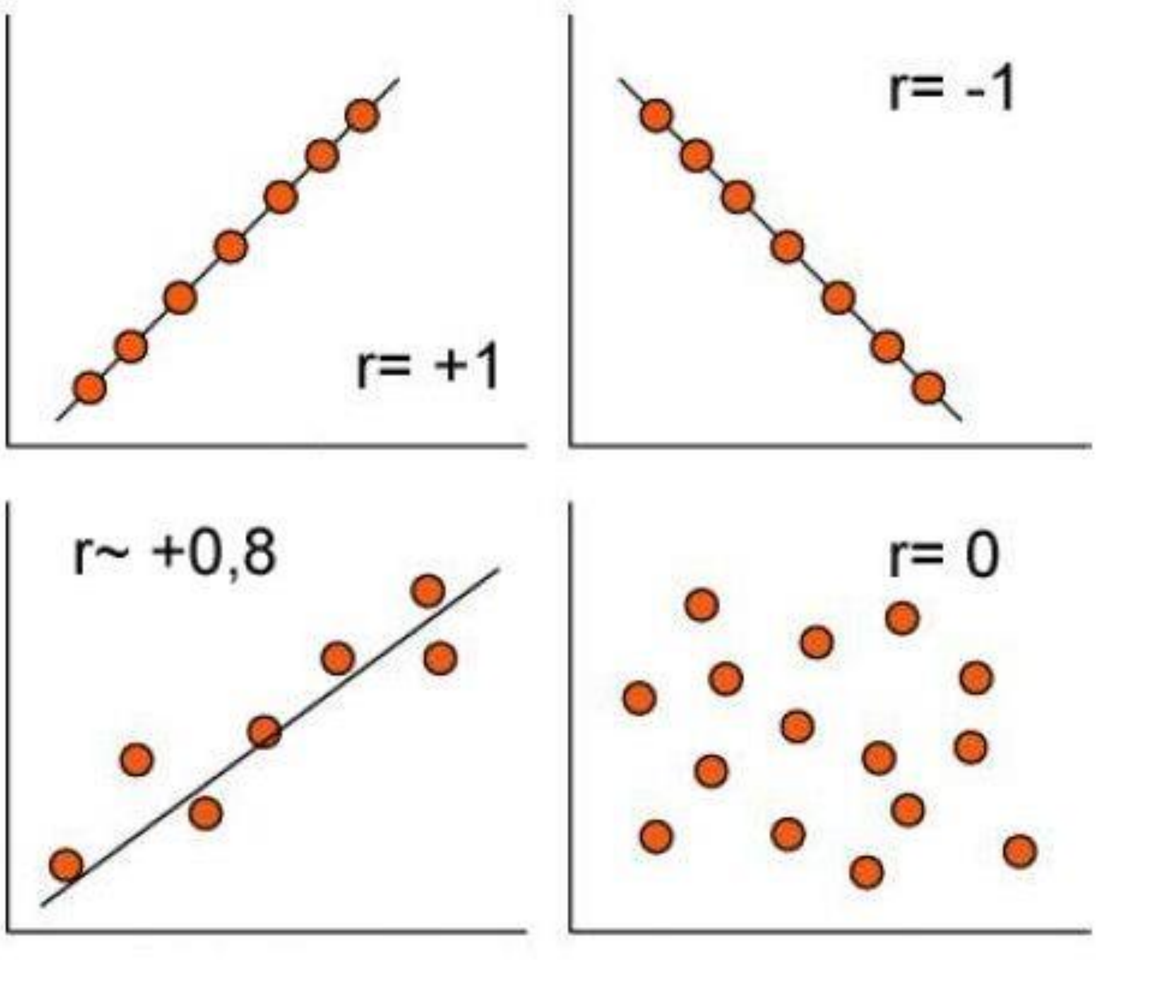
r = 0 ⇐ lineární nezávislost proměnných

r = 1 ⇐ úplná (*funkční)* pozitivní lineární závislost

r = -1 ⇐ úplná (*funkční)* negativní lineární závislost

Čím více se r blíží *hodnotě 1*, tím je *závislost silnější*

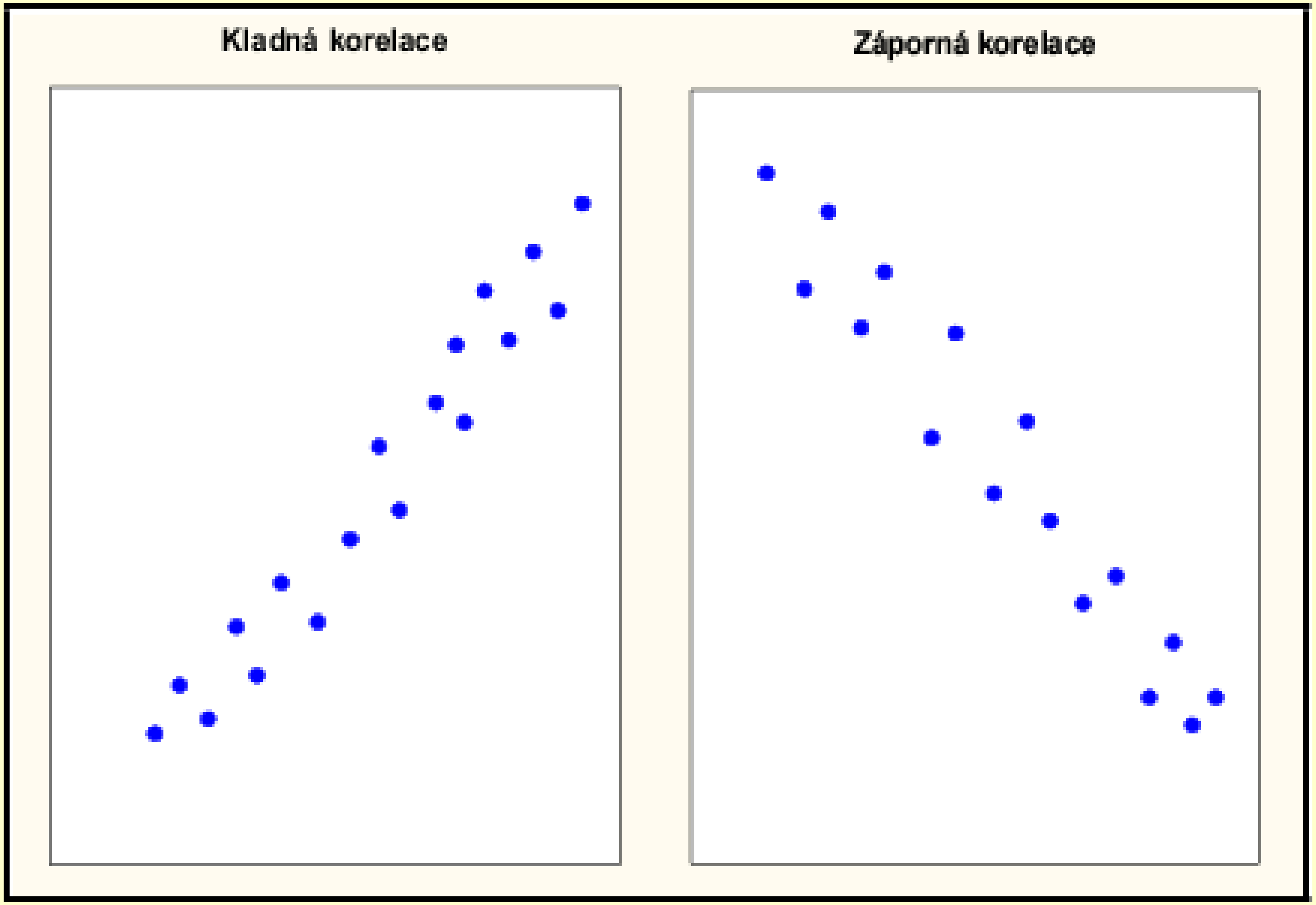
Čím více se r blíží *hodnotě 0*, tím je *závislost slabší*


r= +1
r= -1
r~ +0,8
r= 0

2. SMĚR KORELACE

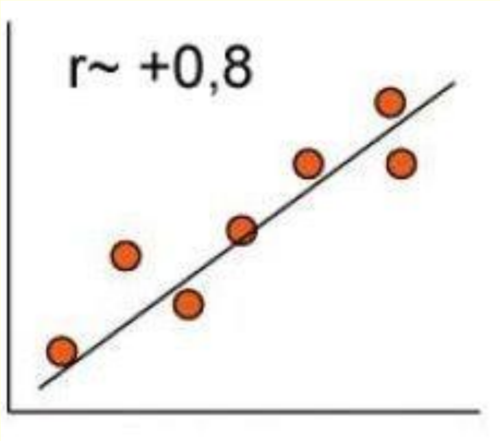
a) kladná (pozitivní) <0;1>

b) záporná (negativní) <-1;0>

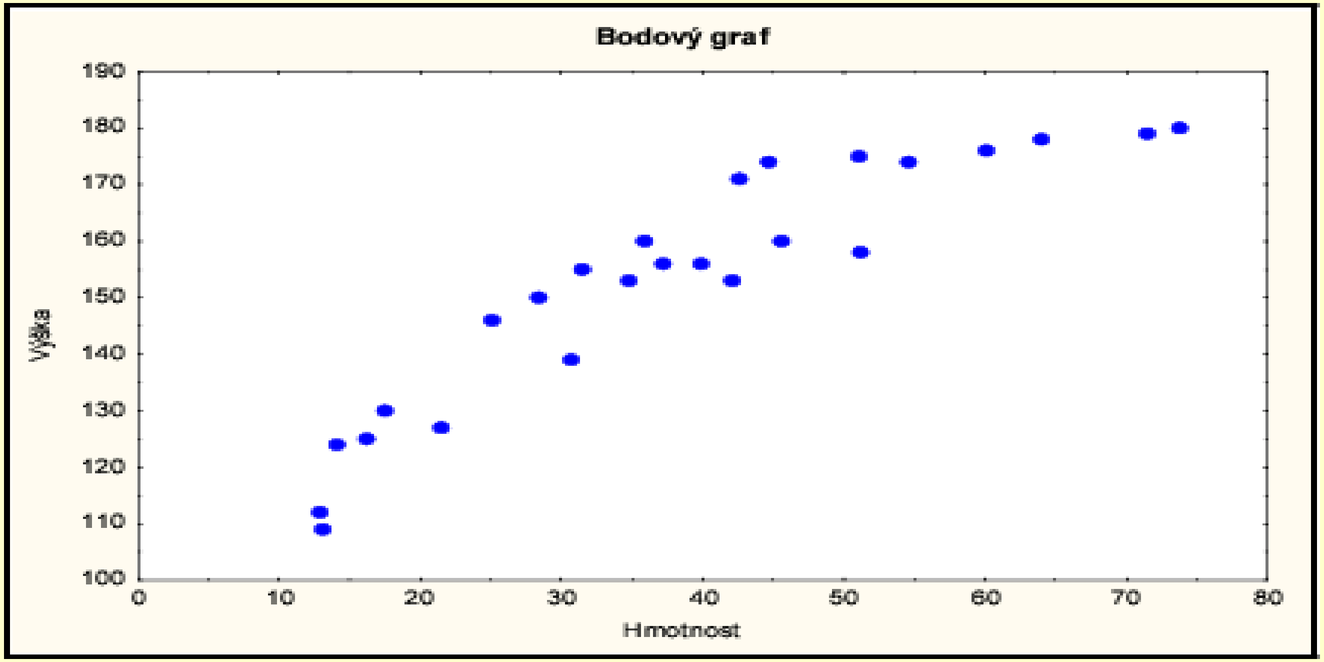


3. TVAR KORELACE

a) lineární (lze dosti dobře proložit přímku)



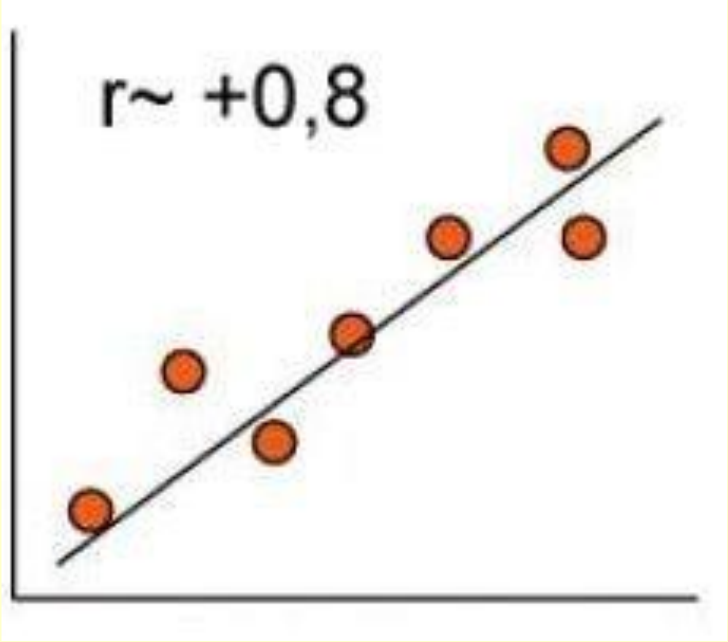
b) nelineární (nelze proložit přímku)



POZNÁMKY KE KORELACÍM

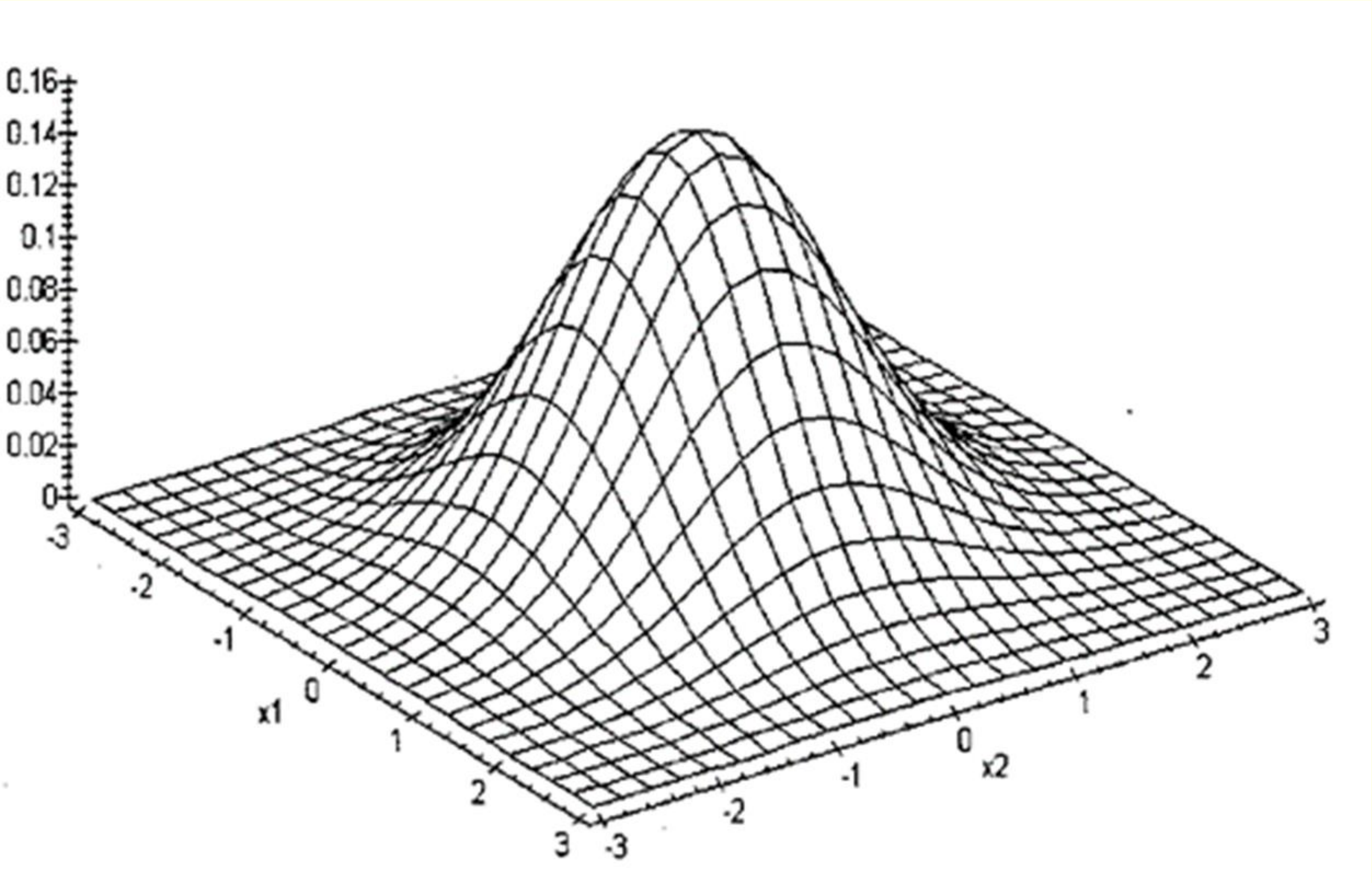
1. Matematicko-statistické předpoklady výpočtu korelačního koeficientu:

a) linearita (korelačním polem lze dosti dobře proložit přímku),



1. Matematicko-statistické předpoklady výpočtu korelačního koeficientu:

b) normalita *(dvojrozměrné normální rozložení četností)*



1. Matematicko-statistické předpoklady výpočtu korelačního koeficientu:

c) dostatečný rozsah souboru (n=200, n=100, n=30)

2. Věcný a formální smysl znaménka korelačního koeficientu

Např. vypočítaná korelační závislost výsledků studentů FSpS (n=185)

v běhu na 100m …

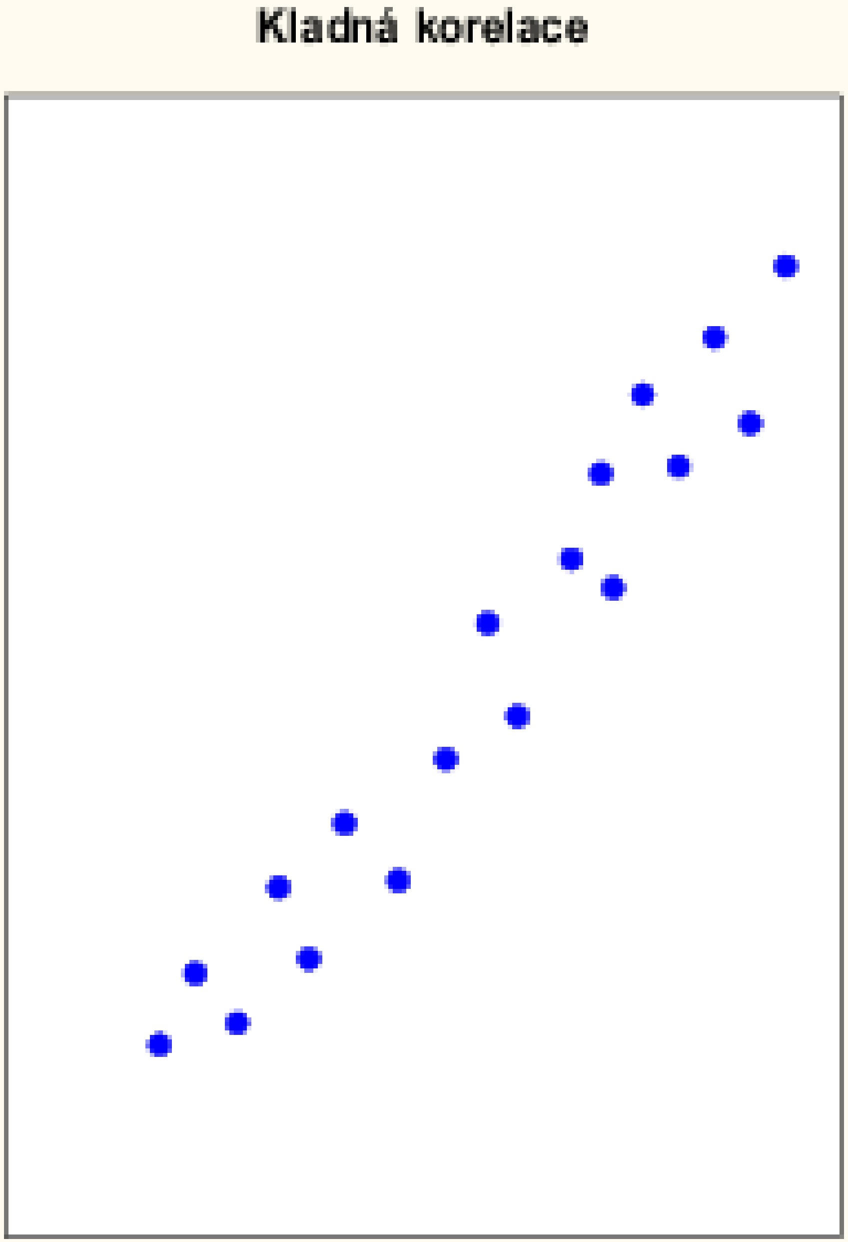
… a ve skoku dalekém je

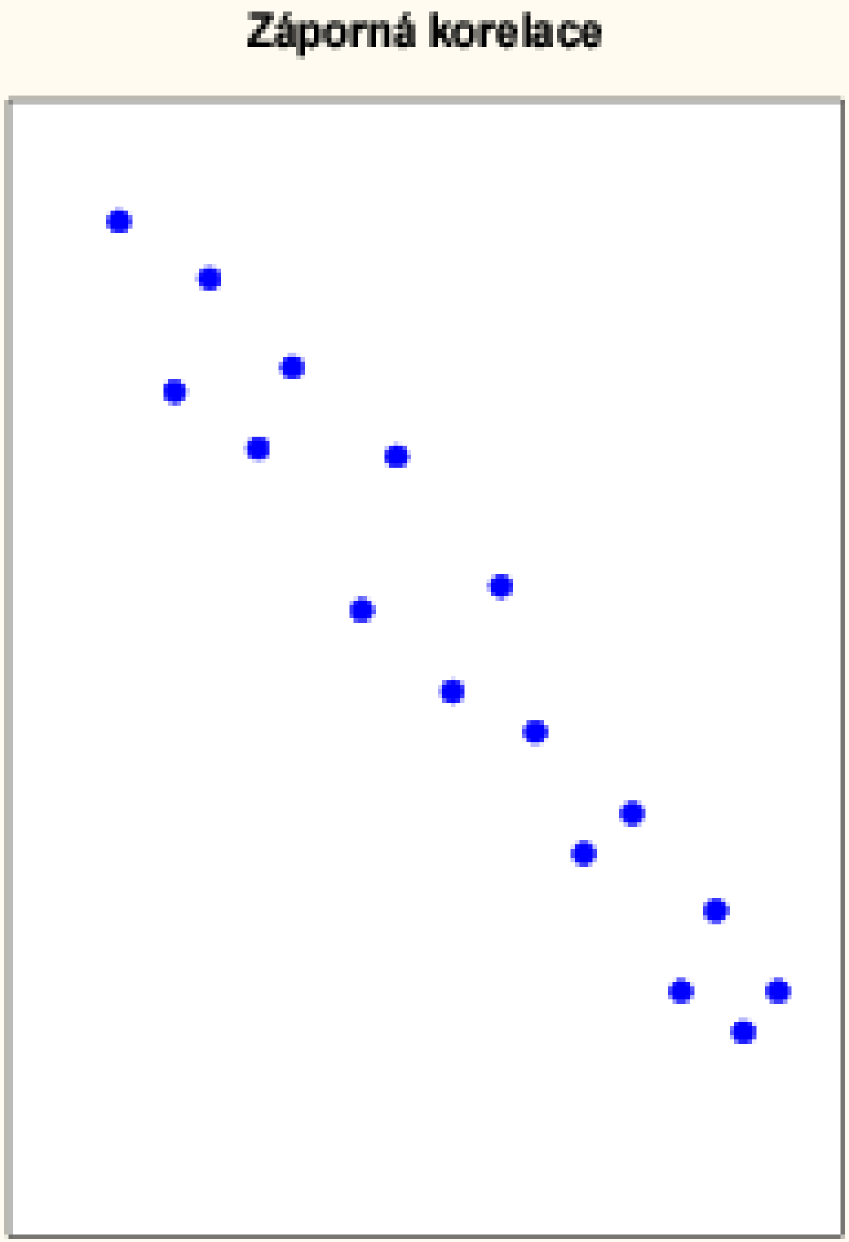
$$r = -0{,}80$$

<-1 ; 1>

Co to znamená z hlediska interpretace?

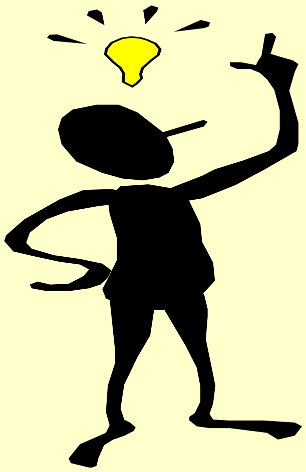
a) kladná (pozitivní) <0;1> **b) záporná (negativní) <-1;0>**





To by ovšem znamenalo, že kdo je <u>lepší</u> v běhu na 100 m, ten je <u>horších</u> výsledků ve skoku dalekém.

To je ovšem……odborně i věcně NESMYSL!



PROČ ???

PROTOŽE …

…jakou „hodnotu“ má výsledek v běhu na 100 m

10,7 s versus 12,3 s?

…jakou „hodnotu“ má výsledek ve skoku dalekém

570 cm versus 430 cm?

3. Koeficient determinace r^2

… určuje *jaká část rozptylu výkonu* *v jednom testu* *je dána proměnlivostí (variabilitou) výkonů* *v druhém testu.*

Např. výše uvedená korelační závislost výsledků studentů FTK (n=185) v běhu na 100m a ve skoku dalekém r = 0,8 znamená, že…

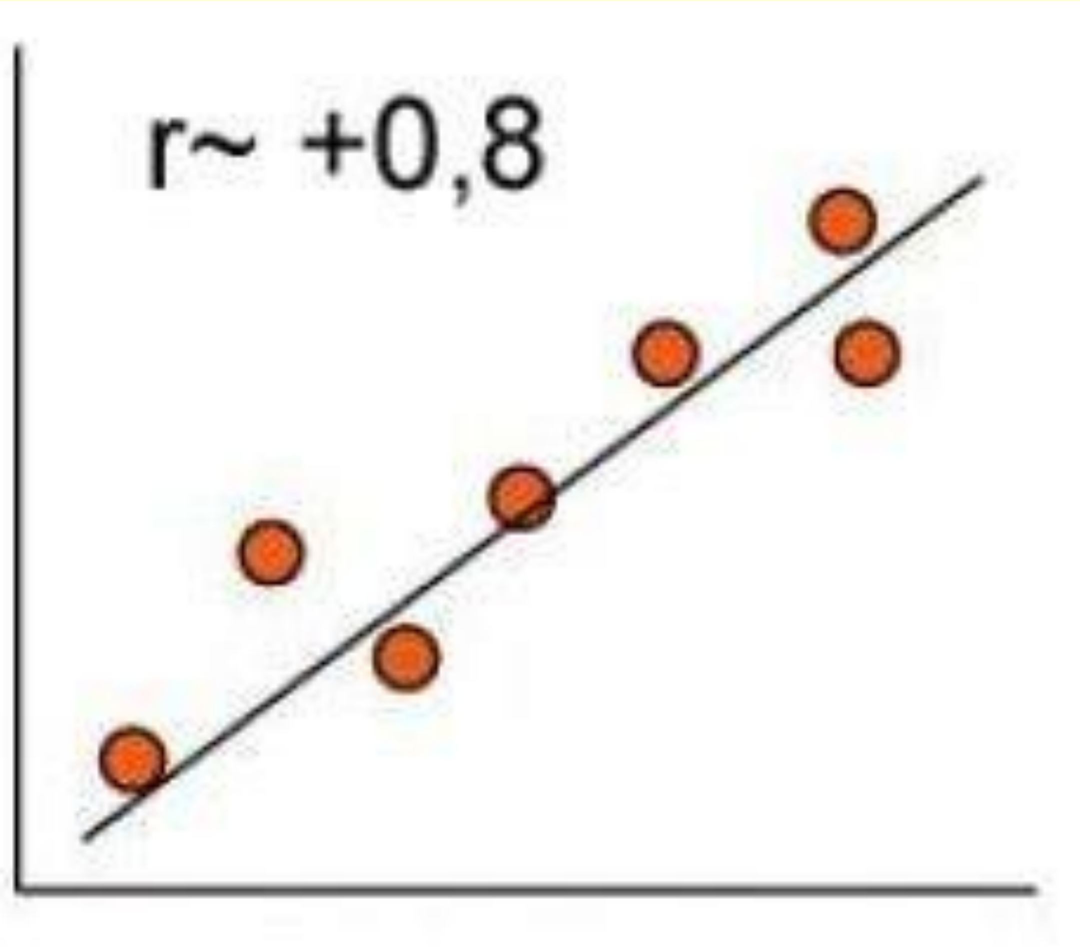
Koeficient determinace $r^2 = 0,64$ *(64 %).*

Tedy 64 % rozptylu výkonu ve skoku dalekém je ovlivněno (determinováno) proměnlivostí (variabilitou) výkonů v běhu na 100m.

REGRESNÍ ANALÝZA (1. úkol korelačního počtu)

PŘÍKLAD 7. Výpočet - koeficientů regresní přímky

Regresní přímka Y = a + b . X



REGRESNÍ ANALÝZA (1. úkol korelačního počtu)

PŘÍKLAD 7. Výpočet - koeficientů regresní přímky

Hráč	A	B	C	D	E	F	G	H	J	K
x_i	7	6	7	8	9	8	8	8	9	10
y_i	4	8	6	8	7	8	7	4	8	10

Regresní přímka $Y = a + b \cdot x$

POMOCNÁ TABULKA

Hráč	X_i	Y_i	X_i^2	Y_i^2	$X_i \cdot Y_i$

REGRESNÍ ANALÝZA (1. úkol korelačního počtu)

PŘÍKLAD 7. Výpočet - koeficientů regresní přímky

POMOCNÁ TABULKA

Hráč	X_i	Y_i	X_i^2	Y_i^2	$X_i . Y_i$
A	7	4	49	16	28
B	6	8	36	64	48
C	7	6	49	36	42
D	8	8	64	64	64
E	9	7	81	49	63
F	8	8	64	64	64
G	8	7	64	49	56
H	8	4	64	16	32
J	9	8	81	64	72
K	10	10	100	100	100
Σ	80	70	652	522	569

Pořadí osob	X_i	Y_i	X_i^2	Y_i^2	$X_i . Y_i$
Σ	80	70	652	522	569

Statistické charakteristiky: $AP_x = 8$ $AP_y = 7$ $s_x = 1,1$ $s_y = 1,8$

$$b = \frac{10 . 569 - 80 . 70}{10 . 652 - (80)^2} = \frac{90}{6520 - 6400} = \frac{90}{120} = 0,75$$

$$a = \frac{70 - 0,75 . 80}{10} = \frac{10}{10} = 1$$

$$Y = a + b . x = 1 + 0,75 . x$$

Konstrukce regresní přímky za pomocí regresní rovnice

$$Y = a + b \cdot x = 1 + 0{,}75 \cdot x$$

Pozn. x…nezávisle proměnná   y…závisle proměnná

Volba x $\Rightarrow$

X	8	10

$Y_1 = 1 + 0{,}75 \cdot 8 = 7$ $Y_2 = 1 + 0{,}75 \cdot 10 = 8{,}5$

Výpočet y $\Rightarrow$

Y	7	8,5

Z_1 (8; 7) Z_2 (10; 8,5)

2) $X = a + b \cdot Y$ **SAMI !!!**

Pozn. y…nezávisle proměnná   x…závisle proměnná

Pořadí osob	X_i	Y_i	X_i^2	Y_i^2	$X_i . Y_i$
$\sum$	80	70	652	522	569

Statistické charakteristiky: $AP_x = 8$ $AP_y = 7$ $s_x = 1,1$ $s_y = 1,8$

Regresní přímka X = a + b . y

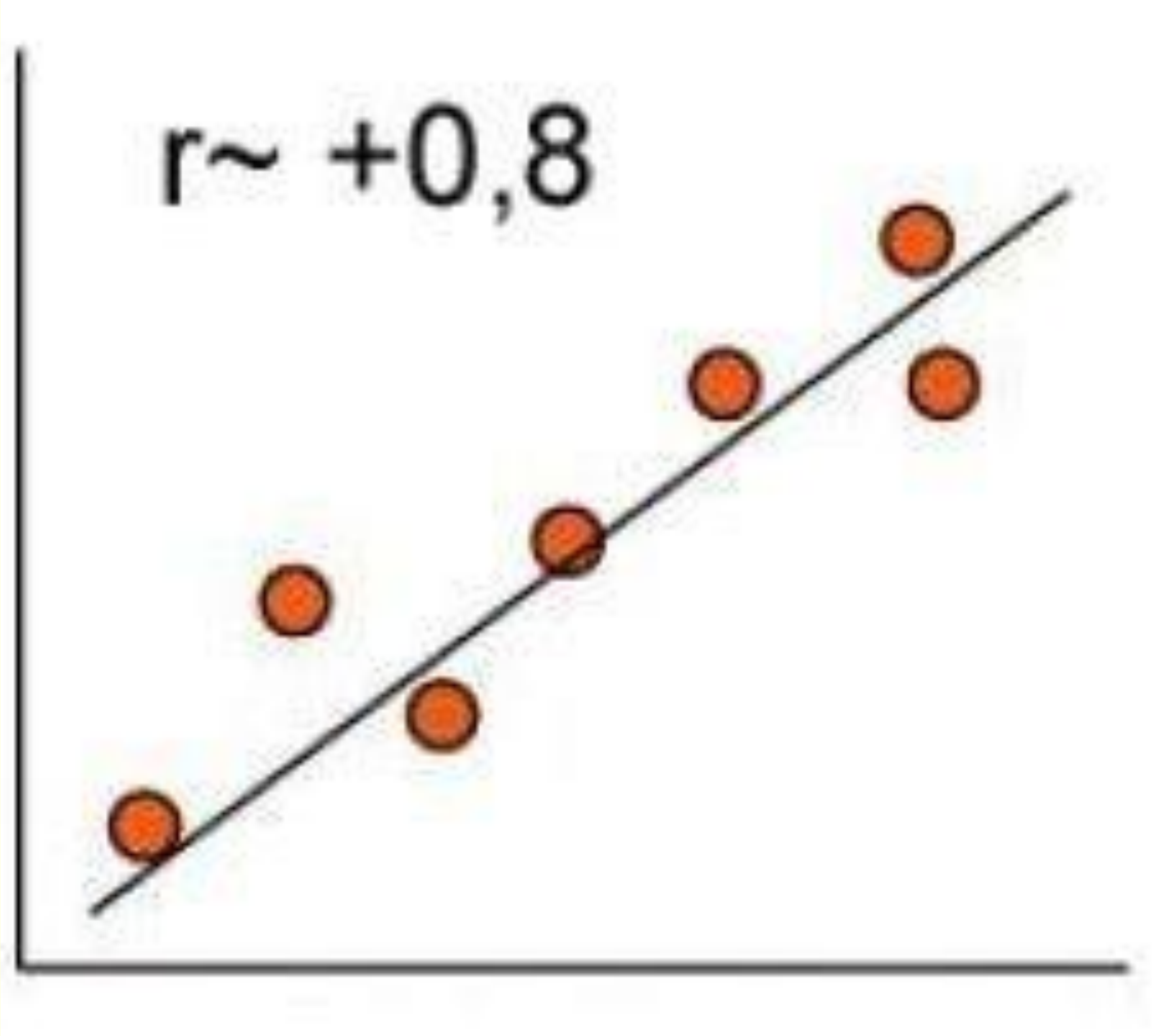
$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum y_i^2 - (\sum y_i)^2} \qquad a = \frac{\sum x_i - b . \sum y_i}{n}$$

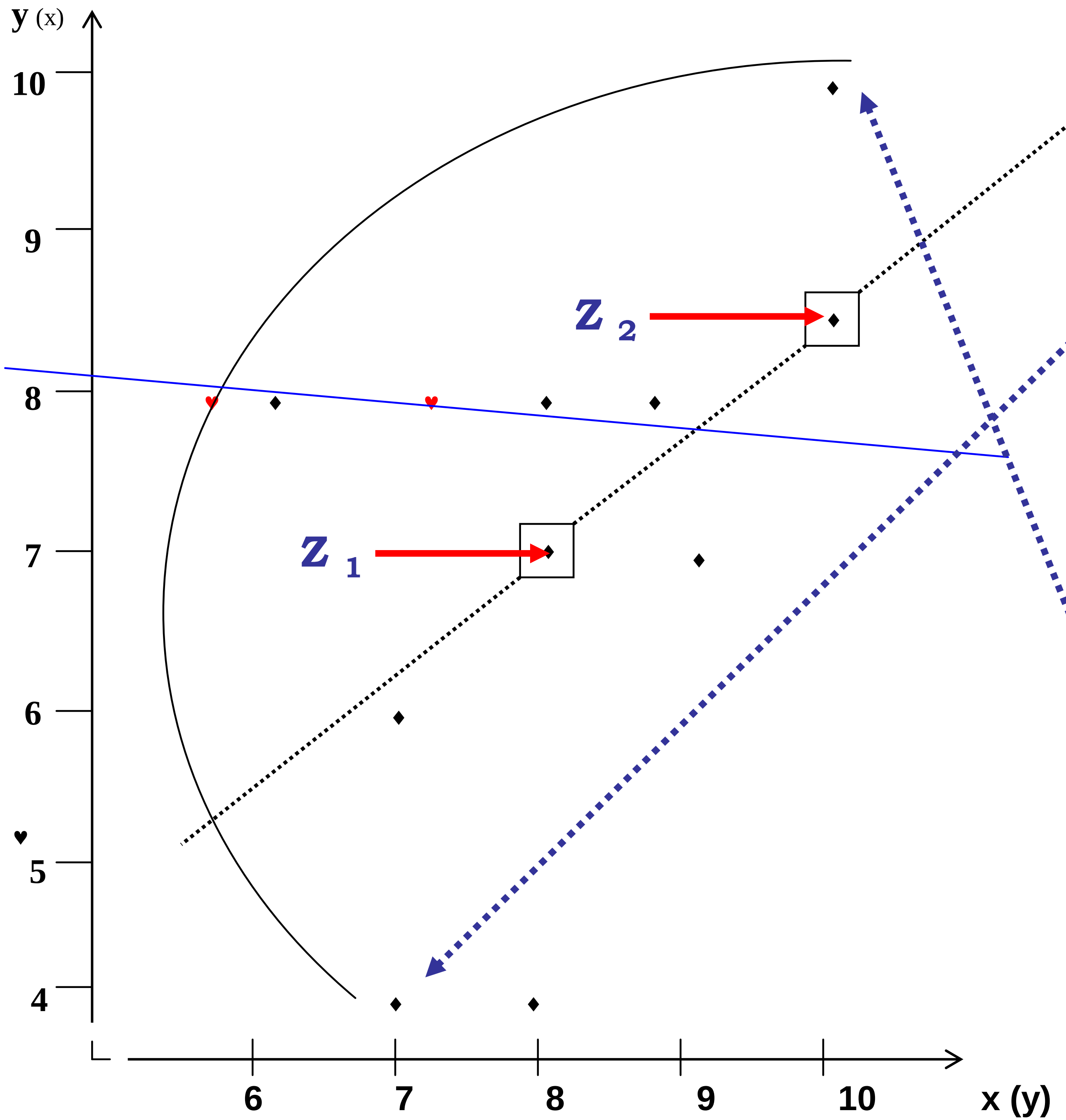
$$b = \frac{10.569 - 80.70}{10.522 - (70)^2} = \frac{5690 - 5600}{5220 - 4900} = \frac{90}{320} = 0,28$$

$$a = \frac{80 - 0,28.70}{10} = \frac{60,4}{10} = 6$$

$$X = a + b . y = 6 + 0,28 . y$$

Graf korelační závislosti (= korelogram) - konstrukce



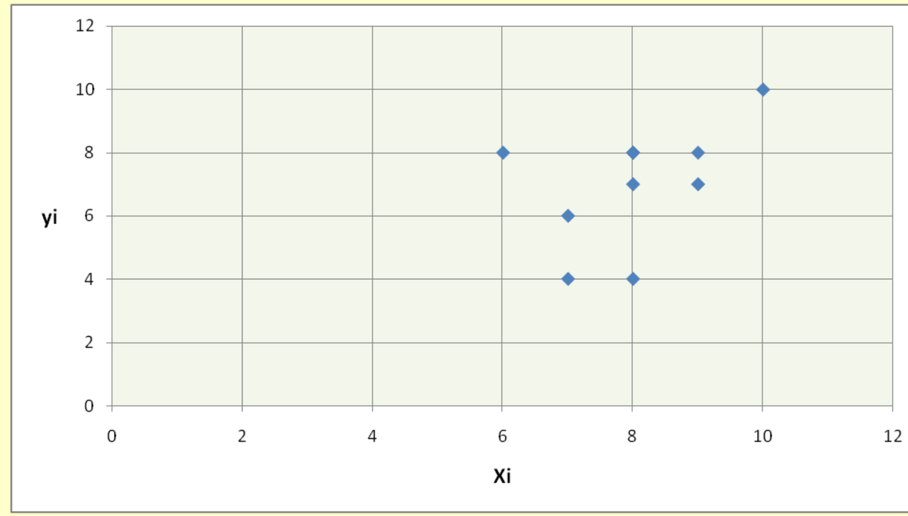


	x_i	y_i
A	7	4
B	6	8
C	7	6
D	8	8
E	9	7
F	8	8
G	8	7
H	8	4
J	9	8
K	10	10

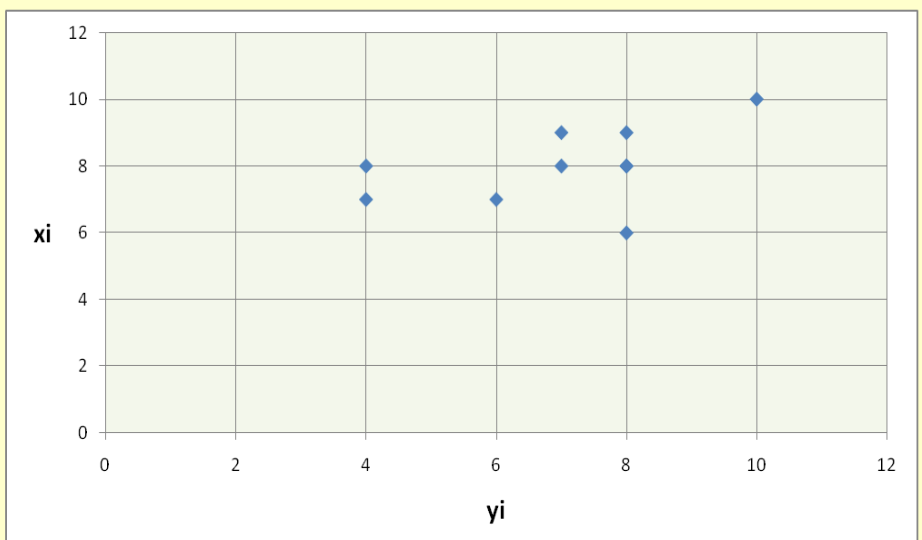
Pomocí Excelu – Statistické funkce
Výpočet koeficientů regresní přímky

	A	B	C	D
1	Hráč	xi	yi	
2	A	7	4	
3	B	6	8	
4	C	7	6	
5	D	8	8	
6	E	9	7	
7	F	8	8	
8	G	8	7	
9	H	8	4	
10	J	9	8	
11	K	10	10	
12				

1) Y = a + b . X

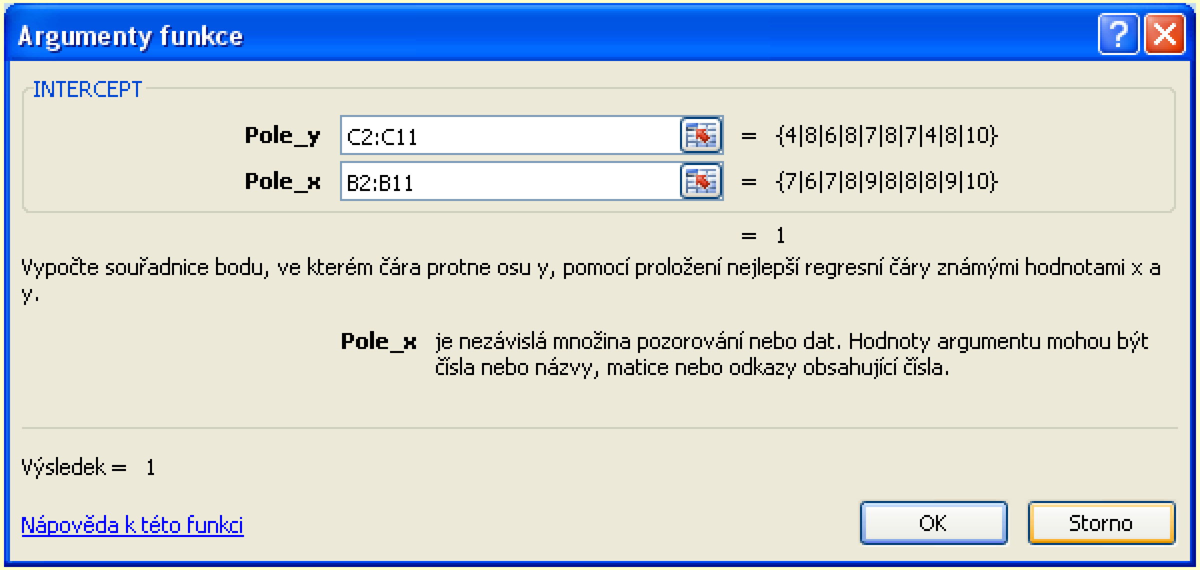


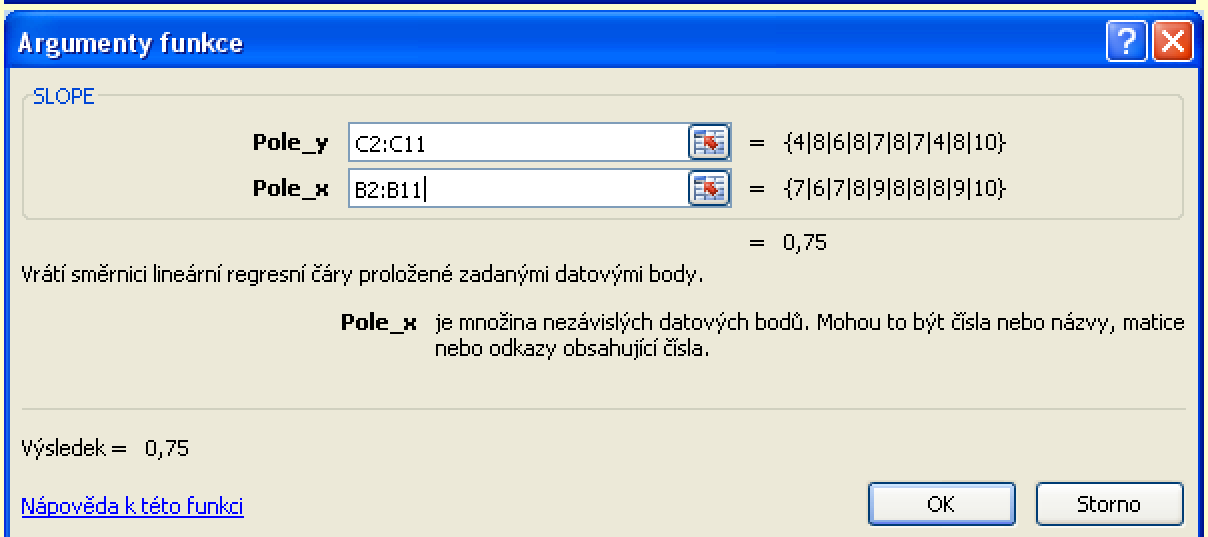
2) X = a + b . Y



Pomocí Excelu – Statistické funkce

Výpočet koeficientů regresní přímky





INTERCEPT
odhad parametru a

SLOPE
odhad parametru b

1) $Y = a + b \cdot X$

$Y = 1 + 0{,}75 \cdot X$

2) $X = a + b \cdot Y$

$X = 6 + 0{,}28 \cdot Y$

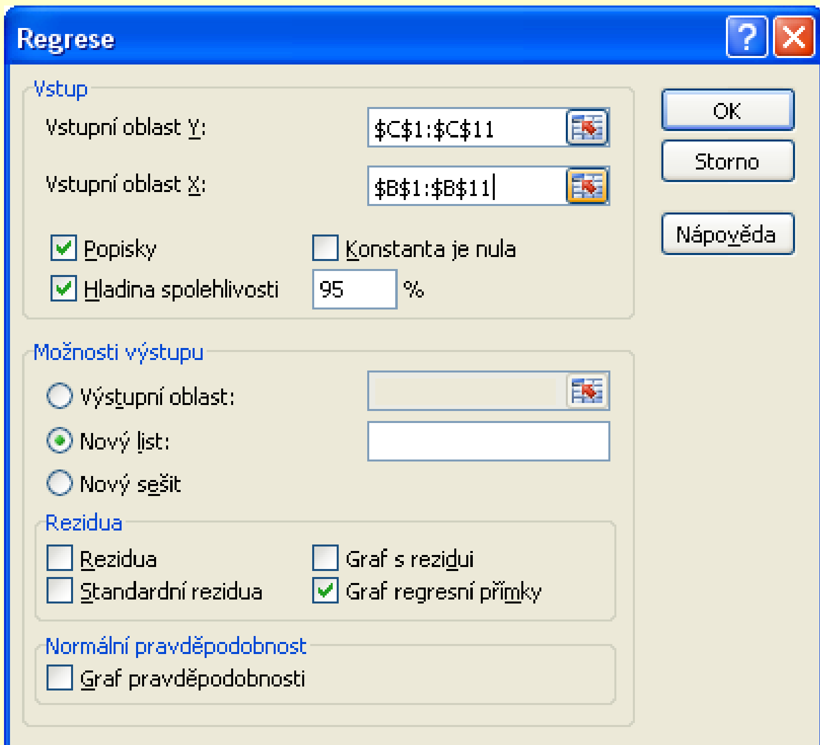
Pomocí Excelu – Analýza dat – Regrese

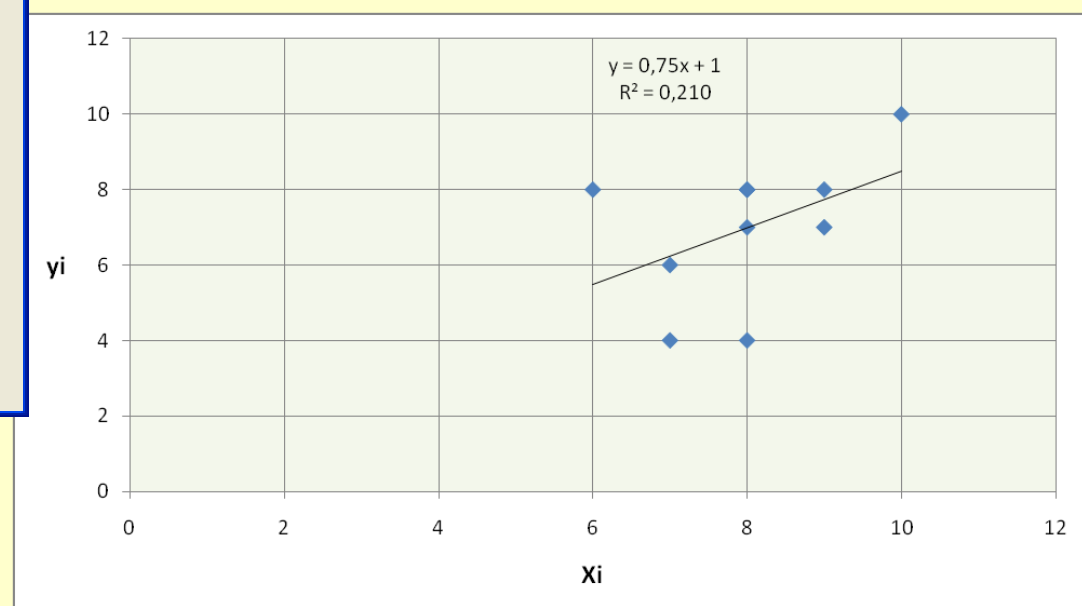
Výpočet koeficientů regresní přímky

1) Y = a + b . X

Y = 1 + 0,75 . X

	A	B	C	D
1	Hráč	xi	yi	
2	A	7	4	
3	B	6	8	
4	C	7	6	
5	D	8	8	
6	E	9	7	
7	F	8	8	
8	G	8	7	
9	H	8	4	
10	J	9	8	
11	K	10	10	
12				





Pomocí Excelu – Analýza dat – Regrese
Výpočet koeficientů regresní přímky

VÝSLEDEK

Regresní statistika	
Násobné R	0,459279327
Hodnota spolehlivosti R	0,2109375
Nastavená hodnota spolehlivosti R	0,112304688
Chyba stř. hodnoty	1,7765838
Pozorování	10

korelační koeficient
koeficient determinace

Významnost F < α = 0,05 →

model je statisticky vhodný

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	6,75	6,75	2,138613861	0,181775314
Rezidua	8	25,25	3,15625		
Celkem	9	32			

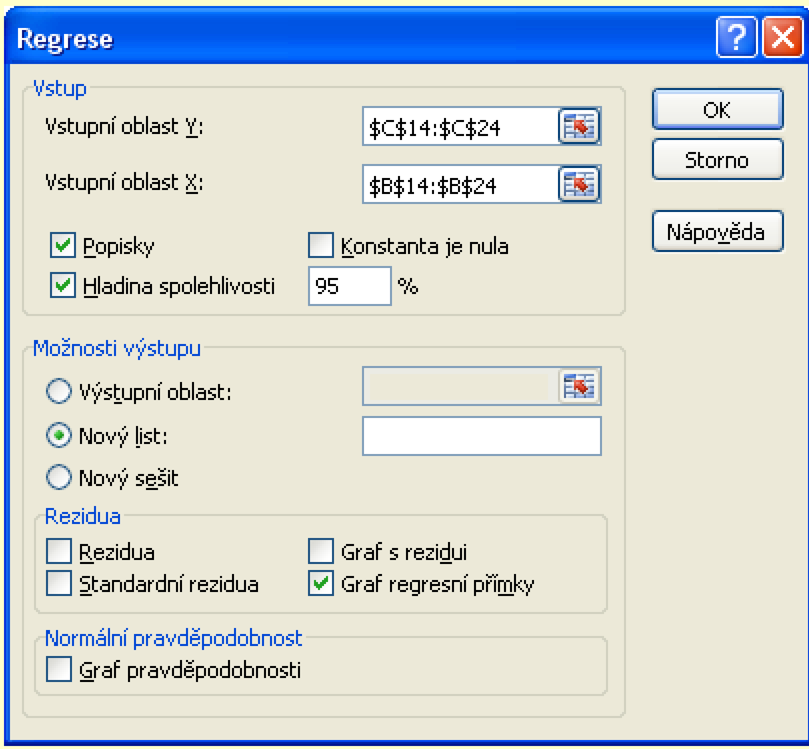
		Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*
Hranice	**a**	1	4,141130079	0,241479978	0,815257536	-8,549463079	10,54946308
xi	**b**	0,75	0,512855568	1,462400035	0,181775314	-0,432647059	1,932647059

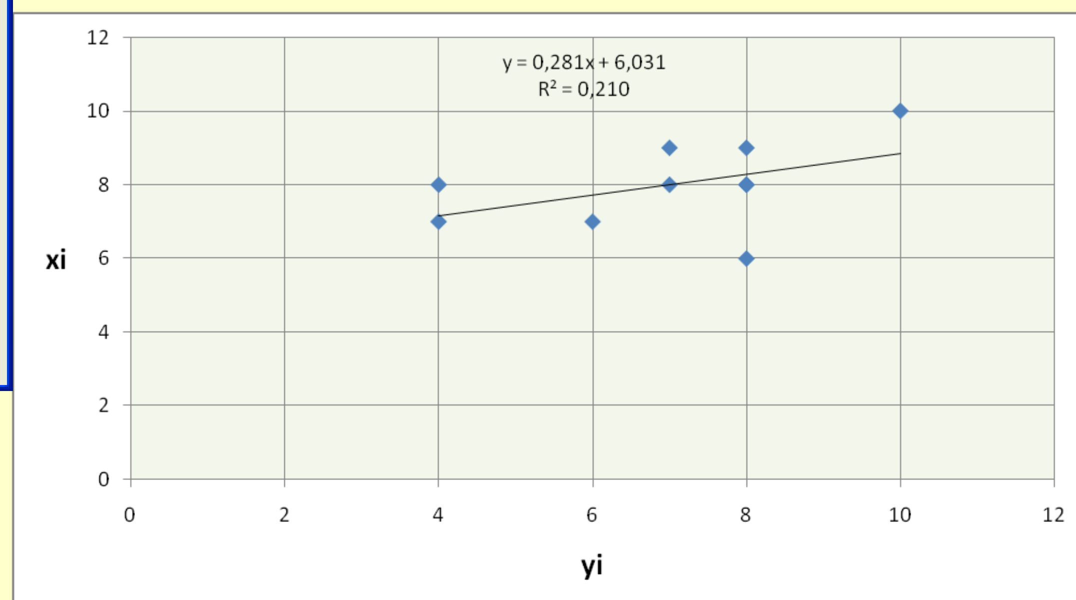
Pomocí Excelu – Analýza dat – Regrese
Výpočet koeficientů regresní přímky

2) X = a + b . Y

X = 6 + 0,28 . Y

	Hráč	yi	xi
14	Hráč	yi	xi
15	A	4	7
16	B	8	6
17	C	6	7
18	D	8	8
19	E	7	9
20	F	8	8
21	G	7	8
22	H	4	8
23	J	8	9
24	K	10	10
25			





Pomocí Excelu – Analýza dat – Regrese
Výpočet koeficientů regresní přímky

VÝSLEDEK

Regresní statistika	
Násobné R	0,459279327
Hodnota spolehlivosti R	0,2109375
Nastavená hodnota spolehlivosti R	0,112304688
Chyba stř. hodnoty	1,087930949
Pozorování	10

korelační koeficient
koeficient determinace

Významnost F < α = 0,05 →

model je statisticky vhodný

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	2,53125	2,53125	2,138613861	0,181775314
Rezidua	8	9,46875	1,18359375		
Celkem	9	12			

		Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*
Hranice	**a**	6,03125	1,389509735	4,340559729	0,002476521	2,827034807	9,235465193
yi	**b**	0,28125	0,192320838	1,462400035	0,181775314	-0,162242647	0,724742647

KORELAČNÍ ANALÝZA (2. úkol korelačního počtu)

PŘÍKLAD 7. Výpočet (Pearsonova) korelačního koeficientu

Pořadí osob	X_i	Y_i	X_i^2	Y_i^2	$X_i . Y_i$
$\sum$	80	70	652	522	569

Výpočtový tvar

$$r_{x,y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]}}$$

$$r = \frac{10.\ 569 - 80.70}{\sqrt{(10.\ 652 - 6400)\ .\ (10.\ 522 - 4900)}} = 0{,}46$$

Pomocí Excelu – Statistické funkce
Výpočet (Pearsonova) korelačního koeficientu

	A	B	C	D
1	Hráč	xi	yi	
2	A	7	4	
3	B	6	8	
4	C	7	6	
5	D	8	8	
6	E	9	7	
7	F	8	8	
8	G	8	7	
9	H	8	4	
10	J	9	8	
11	K	10	10	
12				

CORREL
výpočet korelačního koeficientu

Argumenty funkce

CORREL

Pole1 B2:B11 = {7|6|7|8|9|8|8|8|9|10}

Pole2 C2:C11 = {4|8|6|8|7|8|7|4|8|10}

= 0,459279327

Vrátí korelační koeficient mezi dvěma množinami dat.

Pole2 je druhá oblast buněk s hodnotami. Hodnoty mohou být čísla, názvy, matice nebo odkazy obsahující čísla.

Výsledek = 0,459279327

Nápověda k této funkci

OK Storno

$r = 0{,}46$

POSOUZENÍ A INTERPRETACE KORELAČNÍ ZÁVISLOSTI

Jak „významná“ je korelační závislost r = 0,46?

Vzhledem k intervalu r <0;1> resp. <-1;0> lze hovořit o ***střední míře*** *(velikosti)* ***závislosti*** *(asociace, Hendl, 2012).*

1. **Korelační závislost (r = 0,46) platí pouze pro konkrétní soubor (výběr) s konkrétními osobami, nelze tedy považovat tento vztah za obecně platný!**

2. **Chceme-li zobecnit platnost vypočítané korelace (r) na základní soubor (populaci), musíme ověřit (testovat) hypotézu o statistické významnosti korelačního koeficientu.**

POSOUZENÍ A INTERPRETACE KORELAČNÍ ZÁVISLOSTI

3. Při testování hypotézy o statistické významnosti „r“ (tedy významnosti jeho odlišnost od nuly), zjišťujeme, zda je tento výběrový korelační koeficient statisticky významný (s ohledem na rozsah souboru).

4. Zamítnutí (či nezamítnutí) nulové hypotézy provádíme s určitou pravděpodobností na tzv. hladině významnosti (*p, resp. a).*

Obvykle volíme *p = 0,05, resp. p = 0,01)*

PŘÍKLAD: r = 0,46; n = 10 (basketbal, výsledky 1. a 2. pokusů)

Tab. Kritické hodnoty pro Pearsonův korelační koeficient

Počet dvojic	Kritické hodnoty (na α=0,05, α=0,01)	
n	α=0,05	α=0,01
9	0,666	0,798
10	0,632	0,765
11	0,602	0,735
30	0,361	0,463
50	0,279	0,361
80	0,220	0,286
100	0,197	0,257

Korelační koeficient r = 0,46 je menší, než kritická hodnota (α) je tedy STATISTICKY NEVÝZNAMNÝ.

Závěr: mezi výsledky 1. a 2. pokusů nebyla zjištěna závislost, nelze tvrdit, že... CO? Interpretace!

Testujeme nulovou hypotézu H_0 (r = 0), zda korelační koeficient r je statisticky významný (tedy se významně liší od nuly na hladině významnosti 0,05, resp. 0,01).

Tabulka VIII – Kritické hodnoty pro Pearsonův korelační koeficient (oboustranný test)

	α			α			α	
n	0,05	0,01	n	0,05	0,01	n	0,05	0,01
3	0,9969	0,9999	14	0,5324	0,6614	25	0,3961	0,5052
4	0,9500	0,9900	15	0,5140	0,6411	30	0,3610	0,4629
5	0,8783	0,9587	16	0,4973	0,6226	35	0,3338	0,4296
6	0,8114	0,9172	17	0,4822	0,6055	40	0,3120	0,4026
7	0,7545	0,8745	18	0,4683	0,5897	45	0,2940	0,3801
8	0,7067	0,8343	19	0,4555	0,5751	50	0,2787	0,3610
9	0,6664	0,7977	20	0,4438	0,5614	60	0,2542	0,3301
10	0,6319	0,7646	21	0,4329	0,5487	70	0,2352	0,3060
11	0,6021	0,7348	22	0,4227	0,5368	80	0,2199	0,2864
12	0,5760	0,7079	23	0,4732	0,5256	90	0,2072	0,2702
13	0,5529	0,6835	24	0,4044	0,5151	100	0,1966	0,2565

Zdroj: Anděl, Jiří. *Statistické metody*. 2. vyd. Praha: MATFYZPRESS, 2003

SPEARMANŮV KOEFICIENT POŘADOVÉ KORELACE

Kdy je vhodný pro výpočet korelační závislosti?

1. u znaků získaných na ordinální stupnici (ordinálních znaků)

2. u souborů o nevelkém rozsahu (cca $n < 20$)

3. jestliže znaky nemají (či nelze prokázat) normální rozložení četností

Vzorec pro výpočet Spearmanova koeficientu pořadové korelace:

$$\boxed{r_{xy} = 1 - \frac{6 \cdot \sum (i_x - i_y)^2}{n(n^2 - 1)}}$$

kde i_x resp. i_y je *index pořadí znaků* x resp. y

Příklad. **Výpočet Spearmanova koeficientu pořadové korelace**

Pořadí	x_i	y_i	i_x	i_y	$(i_x - i_y)^2$
1	7 2,5.	4 1.5	*2,5*	*1,5*	1
2	6 1.	8 7,5.	*1*	*7,5*	42,25
3	7 2,5.	6 3	*2,5*	*3*	0,25
4	8	8 7,5.	*5,5*	*7,5*	4
5	9	7 4,5.	*8,5*	*4,5*	16
6	8	8 7,5.	*5,5*	*7,5*	4
7	8	7 4,5.	*5,5*	*4,5*	1
8	8	4 1.5	*5,5*	*1,5*	16
9	9	8 7,5.	*8,5*	*7,5*	1
10	10 10.	10 10.	*10*	*10*	0
$\sum$	-	-	-	-	85,5

$$r = 1 - \frac{6 \cdot \sum (i_x - i_y)^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 85{,}5}{10(100-1)} = 1 - \frac{513}{990} = 0{,}48$$

Spearmanův koeficient pořadové korelace

r = 0,48

Pearsonův koeficient součinové korelace

r = 0,46

POSOUZENÍ A INTERPRETACE ZÁVISLOSTI

...viz Pearsonův koeficient součinové korelace

Tab. XVIII.6. Kritické hodnoty $r_S(\alpha)$
Spearmanova korelačního koeficientu r_S;
$P\{|r_S| > r_S(\alpha)\} \leq \alpha$

n	α	
	0,05	0,01
5	0,900 0	—
6	0,828 6	0,942 9
7	0,745 0	0,892 9
8	0,690 5	0,857 1
9	0,683 3	0,816 7
10	0,636 4	0,781 8
11	0,609 1	0,754 5
12	0,580 4	0,727 3
13	0,554 9	0,697 8
14	0,534 1	0,674 7
15	0,517 9	0,653 6
16	0,500 0	0,632 4
17	0,485 3	0,615 2
18	0,471 6	0,597 5
19	0,457 9	0,582 5
20	0,445 1	0,568 4
21	0,435 1	0,554 5
22	0,424 1	0,542 6
23	0,415 0	0,530 6
24	0,406 1	0,520 0
25	0,397 7	0,510 0
26	0,389 4	0,500 2
27	0,382 2	0,491 5
28	0,374 9	0,482 8
29	0,368 5	0,474 4
30	0,362 0	0,466 5

Příklad. Výpočet Spearmanova koeficientu pořadové korelace

	A	B	C
1	VÝROBEK	ODBORNÍCI	LAICI
2	1	7	8
3	2	9	9
4	3	8	7
5	4	10	10
6	5	6	6
7	6	5	4
8	7	3	5
9	8	4	3
10	9	2	2
11	10	1	1
12			

Tab. XVIII.6. Kritické hodnoty $r_S(\alpha)$ Spearmanova korelačního koeficientu r_S; $P\{|r_S| > r_S(\alpha)\} \leq \alpha$

n	α	
	0,05	0,01
5	0,900 0	—
6	0,828 6	0,942 9
7	0,745 0	0,892 9
8	0,690 5	0,857 1
9	0,683 3	0,816 7
10	0,636 4	0,781 8

F	G	H	I
ix	iy	ix - iy	(ix - iy)^2
7	8	-1	1
9	9	0	0
8	7	1	1
10	10	0	0
6	6	0	0
5	4	1	1
3	5	-2	4
4	3	1	1
2	2	0	0
1	1	0	0

$$r = 1 - \frac{6\,.\sum(i_x - i_y)^2}{n\,(n^2 - 1)} = 1 - \frac{6\,.\,8}{10\,(100 - 1)} = 1 - \frac{48}{990} = 0{,}95$$

Kritické hodnoty z tabulek α = 0,05 0,6364

α = 0,01................ 0,7818

Hypotézu H_0 : $\rho = 0$ o nezávislosti zamítáme

METODY DESKRIPTIVNÍ STATISTIKY

Vše probráno = STATISTIKA 1

1. URČENÍ TYPU ŠKÁLY (nominální, ordinální, metrické)

a) nominální + ordinální ⇒ *neparametrické stat. metody*

b) metrické ⇒ *parametrické statistické metody*

2. ROZLOŽENÍ ČETNOSTÍ ZNAKŮ (NORMÁLNÍ ČI JINÉ)

a) normální ⇒ *parametrické statistické metody*

b) jiné ⇒ *neparametrické statistické metody*

3. VÝPOČET ZÁKLADNÍCH STATISTICKÝCH CHARAKTERISTIK

a) míry centrální tendence

b) míry variability

c) míry závislosti

ZÁKLADY STATISTIKY

Přednášku prezentovat ze samostatného souboru Statistika 2

3. ANALYTICKÁ STATISTIKA

3.1 Základní [illegible]r, výběrový sou[illegible]áhodný výběr, závisl[illegible]y

3.2 Hypotézy [illegible]

3.3 Věcná a [illegible]stická významnost

3.4 Testování statistických hypotéz

ZÁKLADY STATISTIKY – KONTROLNÍ TÉMATA

Charakterizujte stupnice nominální, ordinální a metrické.

Uveďte co je korelace a jakých hodnot může nabývat.

Charakterizujte míry polohy (centrální tendence).

Popište rozdíl mezi měřením základním, odvozeným a asociativním.

Charakterizujte míry variability (rozptýlení).

Uveďte příklady nulové a alternativní hypotézy.

Charakterizujte základní typy hypotéz.

Vysvětlete problematiku použití statistické a věcné významnosti.

**Děkuji
za pozornost**