ZÁKLADY STATISTIKY

Doc. RNDr. Jiří Zháněl, Dr.

aneb

PRŮVODCE

STATISTICKÝM ZPRACOVÁNÍM

KVANTITATIVNÍCH DAT

Přednášky

np4001+nk4001

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

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

2.1.1 Jednorozměrné rozdělení četností

2.1.2 Jednorozměrné intervalové rozdělení četností

2.1.3 Grafické znázornění rozdělení četností

2.2 Míry polohy

2.3 Míry variability

2.3.1 Kvantilové míry variability

2.3.2 Momentové míry variability

PROGRAM PŘEDNÁŠEK

2.4 Standardní skóre

2.5 Míry závislosti

2.5.1 Závislost pevná, volná, statistická a korelační

2.5.2 Lineární korelace a lineární regrese

2.5.3 Součinová a pořadová korelace

3. ANALYTICKÁ STATISTIKA

3.1 Věcná a statistická významnost

3.2 Testování statistických hypotéz

1.1 HISTORIE STATISTIKY

„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“

(součást zakládací listiny kapituly sv. Štěpána v Litoměřicích, Český statistický úřad, 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

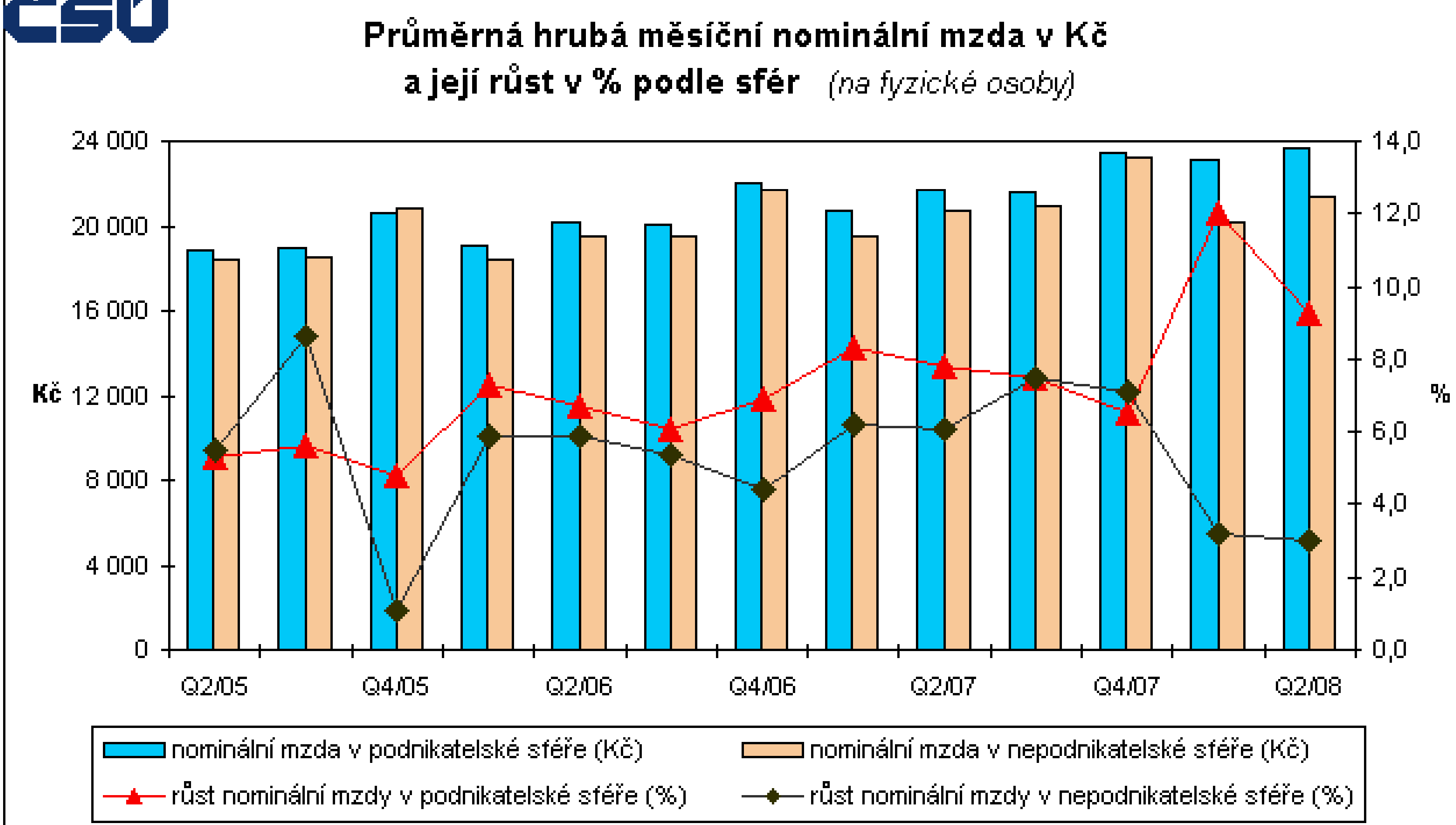
1918 (vznik Československa) => zákon č. 49 o organizaci statistické služby (1919).

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.

Jev Ω nazýváme jistý, jev Ø nemožný. Jednotlivé výsledky ω patří do Ω se nazývají elementární jevy.

(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. pomocí statistických metod posoudit, zda střední hodnoty (M) výsledků testu „skok daleký z místa“ tréninkových skupin A a B vykazují významný rozdíl, který může být vysvětlován použitou tréninkovou metodou.

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

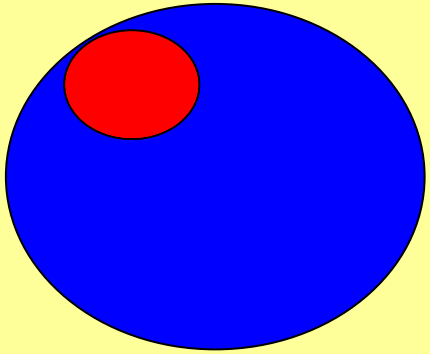
... např. všichni studenti oboru TV a sport = v ČR, Evropě, na světě, ...

ZÁKLADNÍ SOUBOR (ZS)

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 (VS) získáváme tzv. NÁHODNÝM VÝBĚREM.

Každý prvek základního souboru má stejnou možnost být vybrán.

O vybrání či nevybrání do výběrového souboru rozhoduje tedy pouze náhoda.

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

Výzkum u prvků výběrového souboru (TV, H, síla):

na základě poznání vlastností *výběrového souboru* se usuzuje – 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 https://supermartas.cz/aplikace/online/nahodna-cisla/)

II. Tabulka náhodných čísel

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.

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

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 Excelu – 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, věci, události, jejichž vlastnosti nás zajímají (student, klub).

Zjišťujeme-li u každé statistické jednotky 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. o vícerozměrném statistickém souboru (3 a více znaků).

STATICKÉ ZNAKY

Vyjádření hodnot statistických znaků (proměnných) slovy nebo čísly.

Klasifikace:

1. **Slovní proměnné = alfabetické (kategoriální) se nazývají kvalitativní znaky.**
2. **Číselné proměnné = numerické se nazývají kvantitativními znaky.**

STATICKÝ ZNAK
je společná vlastnost jednotek statistického souboru

Statistické znaky tedy **vyjadřují** ***vlastnosti statistických jednotek.***

1. **KVALITATIVNÍ**

   (kategoriální, vyjádřeny slovně)

Např. **muž/žena, plavec/neplavec, zdravý/nemocný**

barva očí: zelené, modré, hnědé, …,

herní kategorie: žáci mladší, starší, junioři, …

1. KVALITATIVNÍ ZNAKY

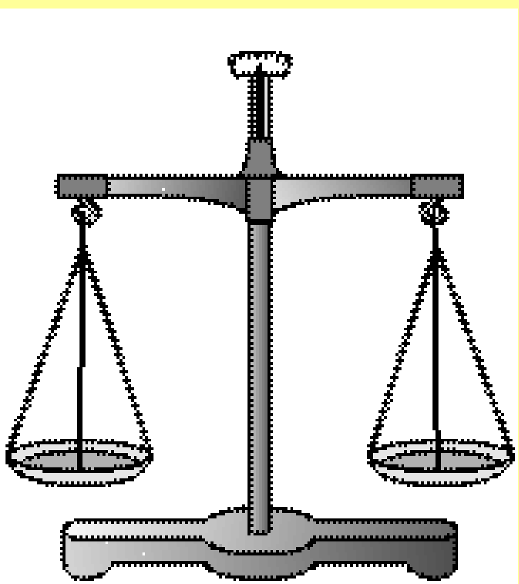
☯ ***alternativní (binární, dichotomické):***

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

☯ ***množné (polytomické):***

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

2. KVANTITATIVNÍ (vyjádřeny číselně, např. věk 37 let, tělesná výška 183 cm, hmotnost, čas, ...)



2. KVANTITATIVNÍ ZNAKY

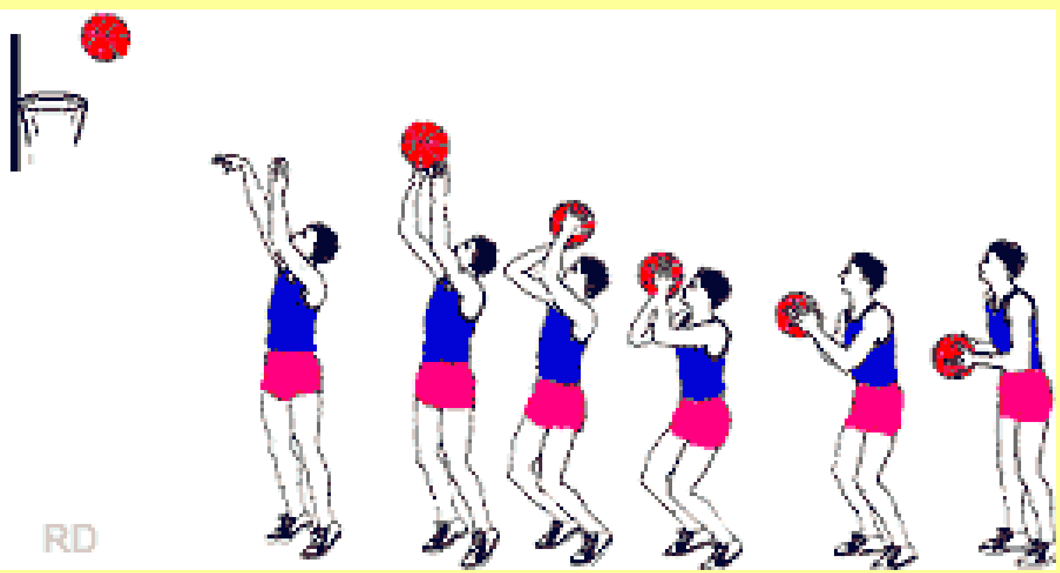
- ***spojité*** **neboli *kontinuální***

(nabývají libovolných reálných číselných hodnot, např. výsledek v běhu na 100 m, ve skoku vysokém, mezi 2 hodnotami vždy může být další),

2. KVANTITATIVNÍ ZNAKY

☯ ***nespojité* neboli *diskrétní***

(nabývají pouze konečný počet číselných hodnot, nejčastěji z oboru celých nezáporných čísel. např. počet úspěšných hodů na koš, leh-sedy, atd.).



1.2 TEORIE MĚŘENÍ, MĚŘÍCÍ STUPNICE (ŠKÁLY)

1.2.1 ZÁKLADNÍ POJMY TEORIE MĚŘENÍ

Měření … v průběhu historického vývoje lidské společnosti běžná každodenní procedura (užití hodinek, tachometru automobilu, atd.).

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

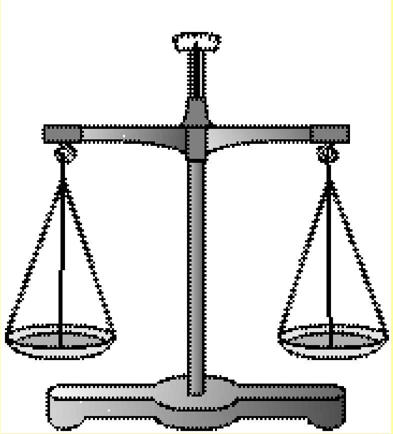
Rozvoj vědy a techniky složitých měřících přístrojích.

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),

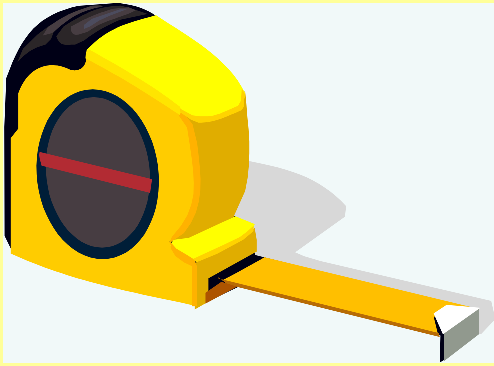


b) Měřitelnost *psychických vlastností* (inteligence, strach, postoje).

***REPREZENTAČNÍ TEORIE MĚŘENÍ* (Campbell):**

… měření jako „přiřazování číslic k reprezentaci vlastností“.

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



Klasická koncepce měření rozlišuje

(1) fundamentální měření

(2) odvozené měření

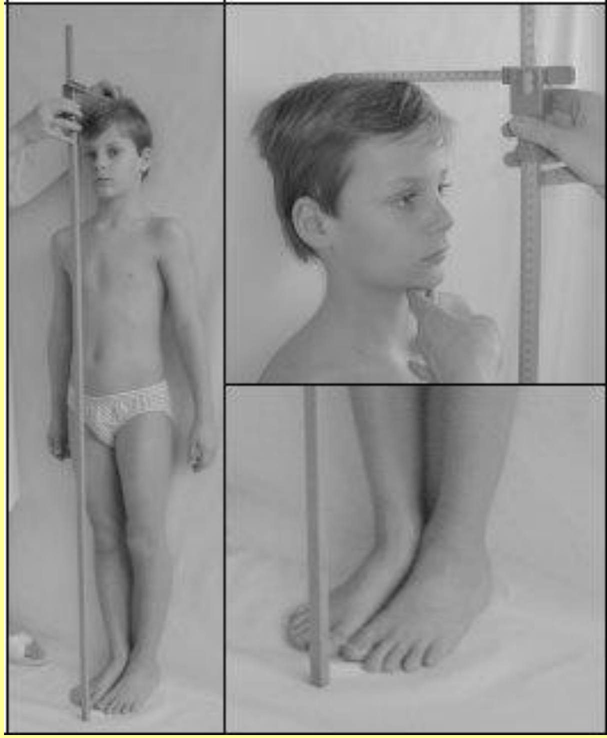
Někteří autoři zmiňují ještě

(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 MĚŘENÍ

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

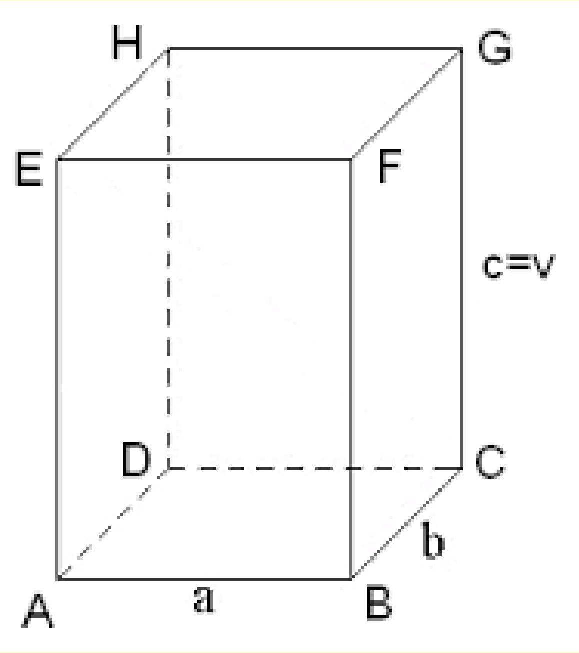
Příklad: měření tělesné výšky



(2) ODVOZENÉ MĚŘENÍ

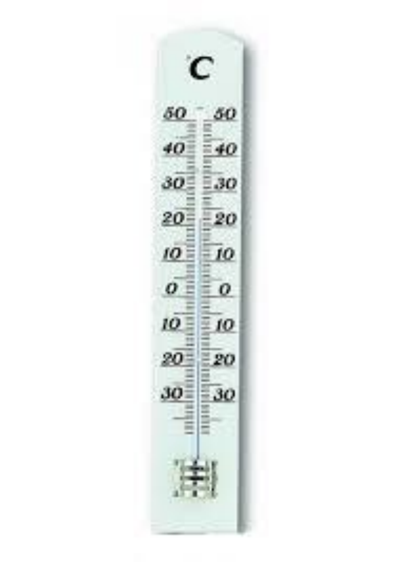
„předpokládá jiná, dříve provedená měření, z nichž je odvozeno na základě vztahů“; a tedy „závisí na předcházejících měřeních“.

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.

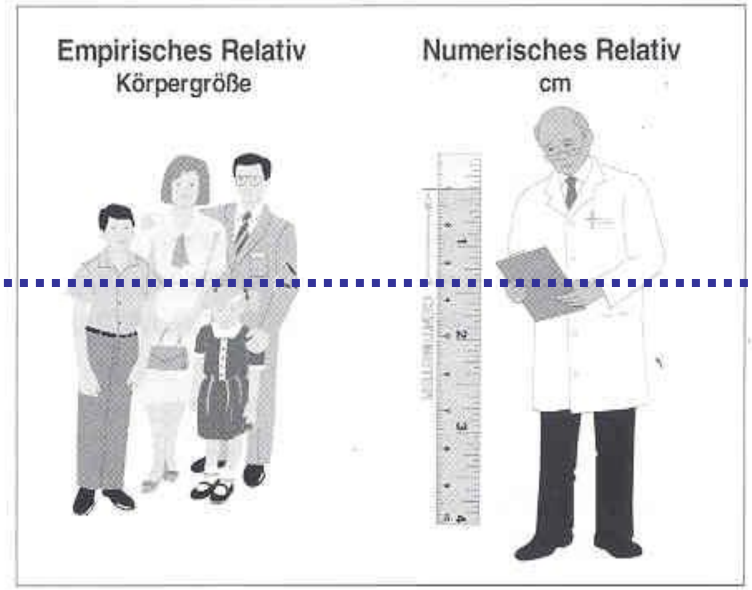
Příklad 2.

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)

Rozdíl v měření

Empirická proměnná



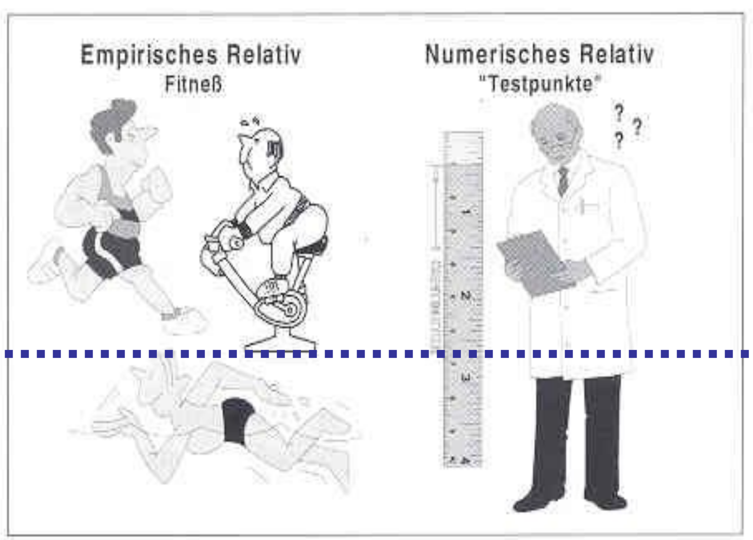
Zuordnungen von Körpergrößen (ER) zu Zahlenwerten (NR)

Numerická proměnná

Tělesná výška → **cm**

Rozdíl ve způsobu měření a přiřazení!

Empirická proměnná



Zuordnung von Fitneßausprägungen (ER) zu Testpunkten (NR)

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

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, atd.

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**

Základní empirickou operací je *„určení rovnosti“*.

Možné relace =, $\neq$,

Zpracování znaků = neparametrické statistické metody

2. ORDINÁLNÍ ŠKÁLA (pořadová)

… předpokládá přirozené uspořádání objektů vzhledem k nějaké vlastnosti.

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

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

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

Základní empirické operace (2):

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

Relace =, ≠, >, <,

Zpracování znaků=neparametrické statistické metody.

3. METRICKÉ ŠKÁLY (INTERVALOVÁ A POMĚROVÁ)

NOVÁ VLASTNOST: je zavedena *jednotka měření,* tzn. jsou známy *odstupy (intervaly)* mezi hodnotami (čísly).

3. 1 INTERVALOVÁ ŠKÁLA

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

Nula je <u>zvolená!!!</u> => stanovení počátku dohodou.

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

teplota °C (bod tání ledu 0 °C a bod varu 100 °C při 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).

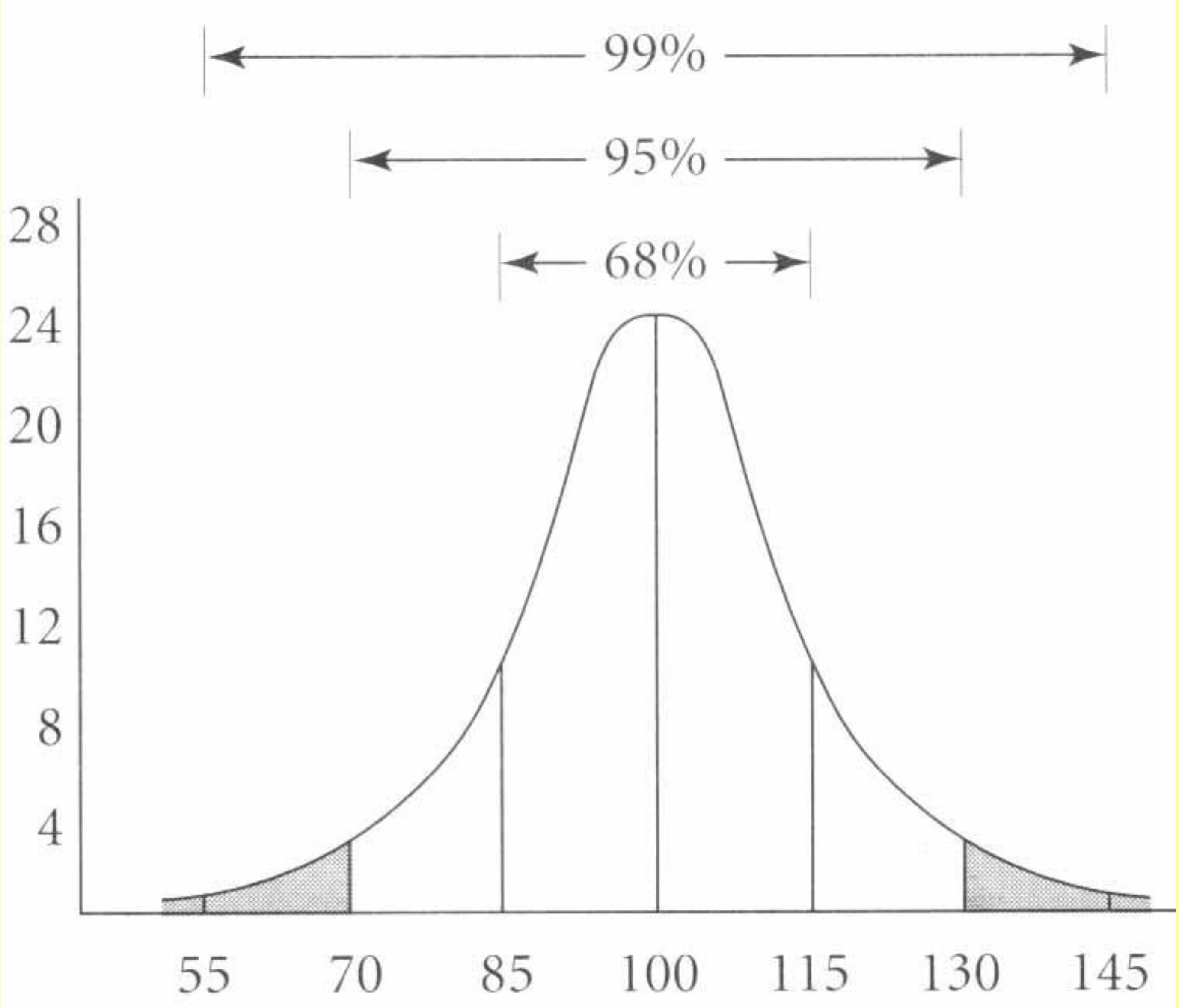
Např. čas, věk, výška, hmotnost, teplotní stupnice dle Kelvina (v podstatě všechny fyzikální jednotky).

Statistické metody: parametrické i neparametrické.

Přehled typů škál (upraveno, 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) je index inteligence, který má normální rozložení s průměrem 100 a standardní odchylkou 15.



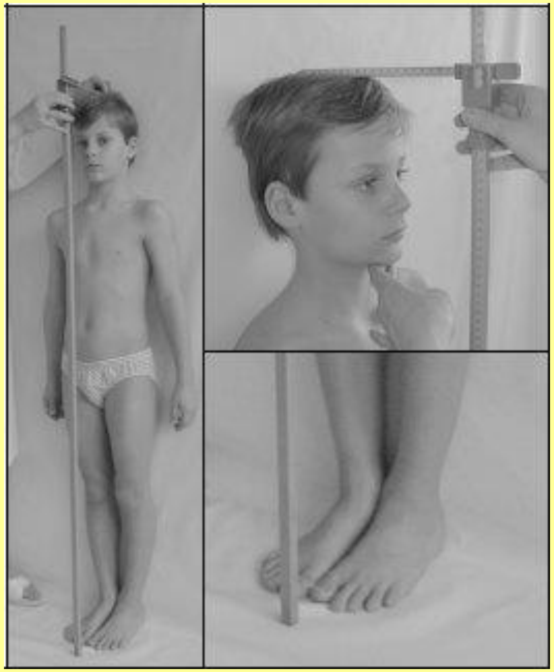
POSTUP PŘI URČENÍ TYPU ŠKÁLY: A. výška (cm)

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 (známka)

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.

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 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 (M, Mo, Me)

b) míry variability (s, ...)

c) míry závislosti (r)

2. DESKRIPTIVNÍ STATISTIKA

2.1 STATISTICKÉ TŘÍDĚNÍ DAT

Výsledkem měření, testování (statistického šetření) jsou *neuspořádané, neroztříděné a nepřehledné* statistické údaje

Hody na koš (n=10): 7; 6; 7; 8; 9; 8; 8; 8; 9; 10

Chceme-li získat podrobnější informace, je třeba údaje uspořádat - tato činnost se nazývá *statistické zpracování (třídění) dat.* Nejjednodušším způsobem statistického zpracování dat je tzv. *tabulka rozdělení (rozložení) četností.*

2.1.1 JEDNOROZMĚRNÉ ROZDĚLENÍ ČETNOSTÍ

Je-li *JEDNA VLASTNOST* statistického souboru charakterizovaná *JEDNÍM STATISTICKÝM ZNAKEM*, hovoříme o *jednorozměrném statistickém souboru (o jednorozměrném rozdělení resp. rozložení četností)*.

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: ***kvantitativní, 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.

POZOR !

Intervaly musí být vytvořeny tak,

aby jeden statistický znak

nemohl být současně zařazen

do dvou různých intervalů!!!

Intervaly na sebe musejí navazovat!!!

POZOR !

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)

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

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

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

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

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	-	-

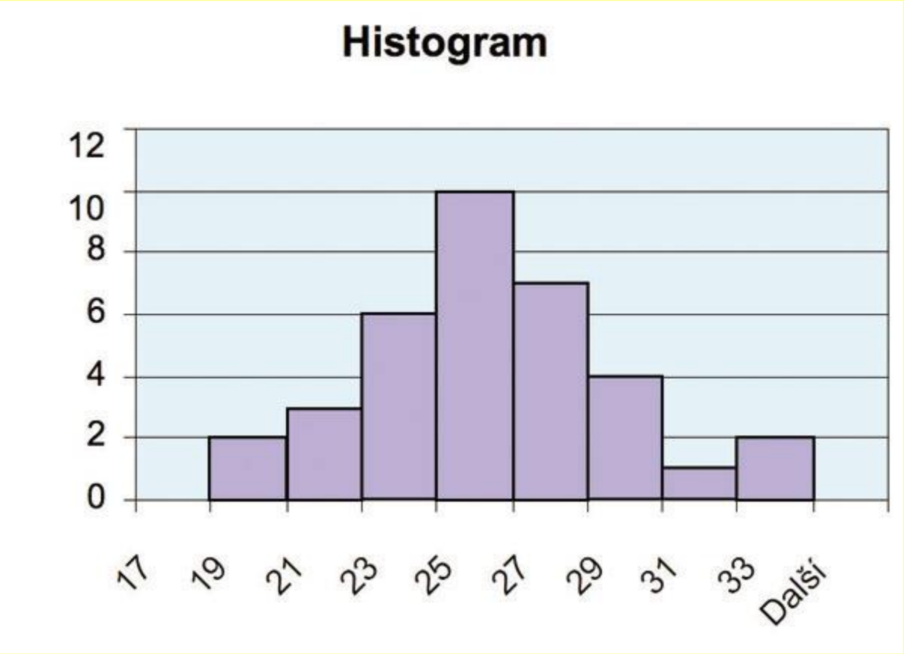
2. 1. 3 GRAFICKÉ ZNÁZORNĚNÍ ROZDĚLENÍ ČETNOSTÍ

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*,

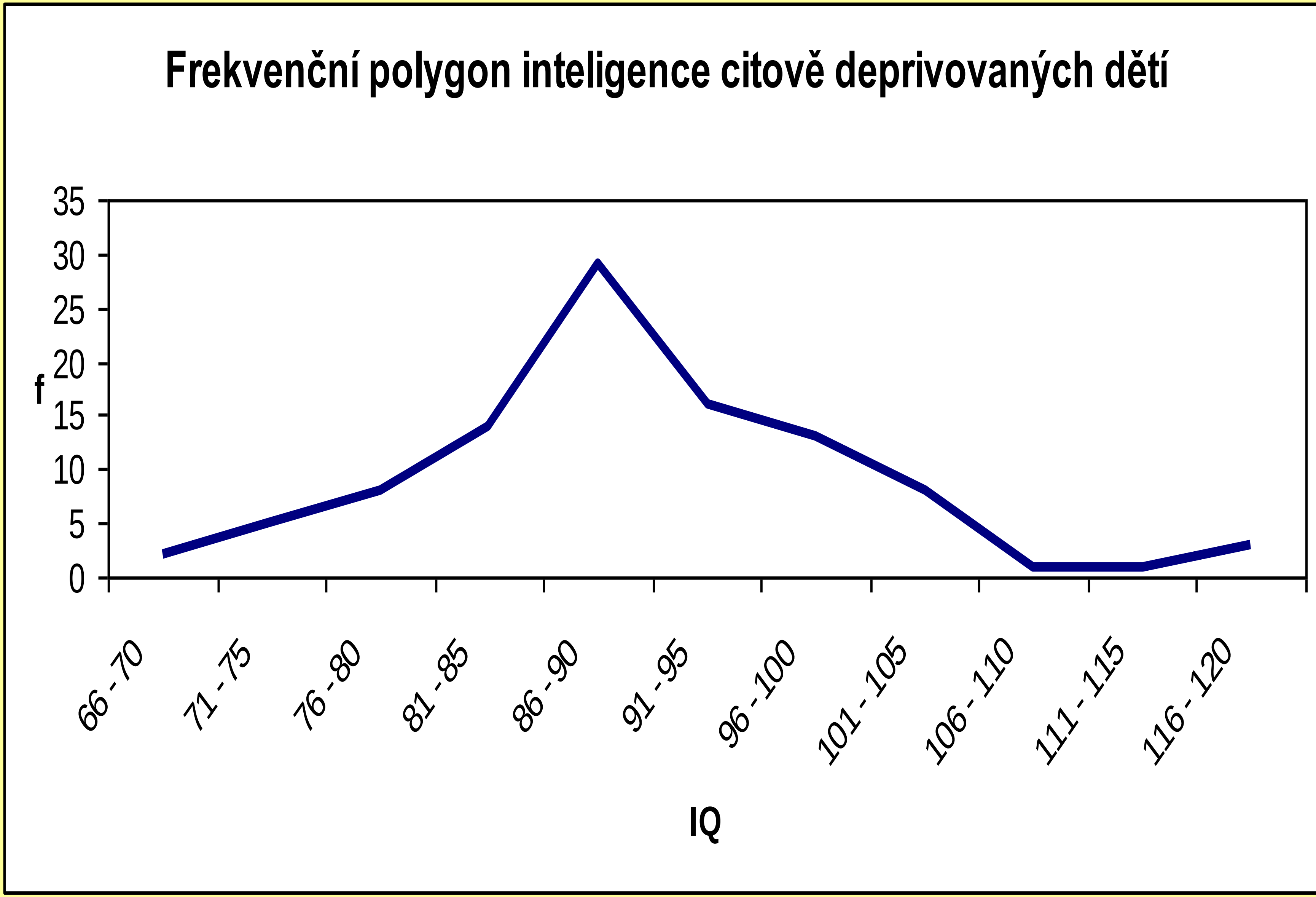
… jejich výška odpovídá *absolutní četnosti* sledovaného statistického znaku.

2) (FREKVENČNÍ) POLYGON

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

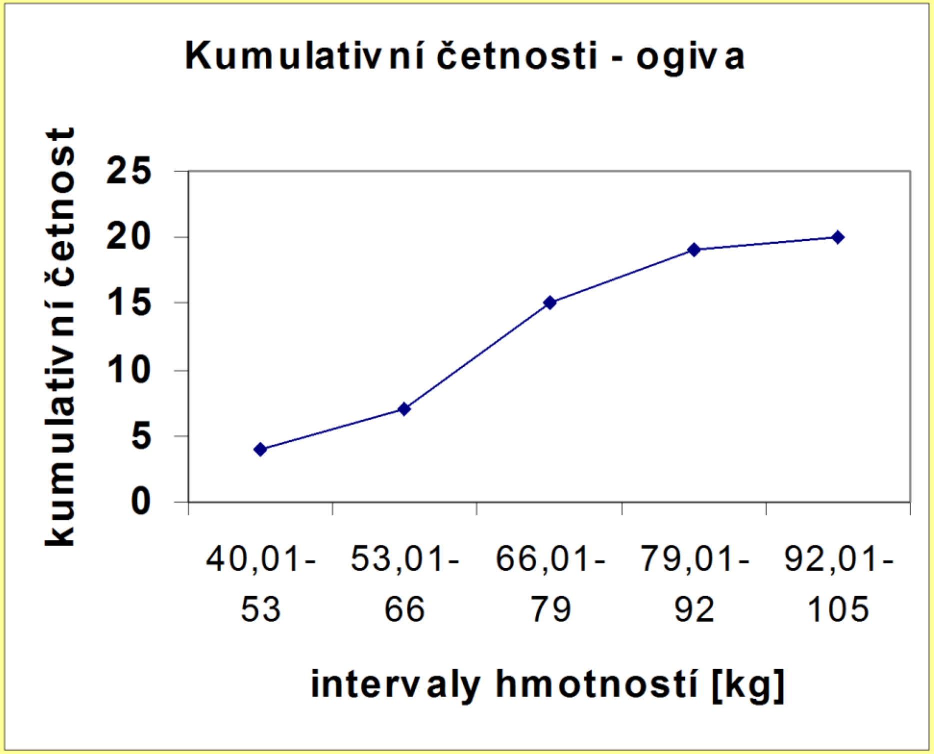
Tato lomená čára je *spojnice bodů* vytvořených v průsečících *středů intervalů* a *příslušných četností.*

2) (FREKVENČNÍ) POLYGON



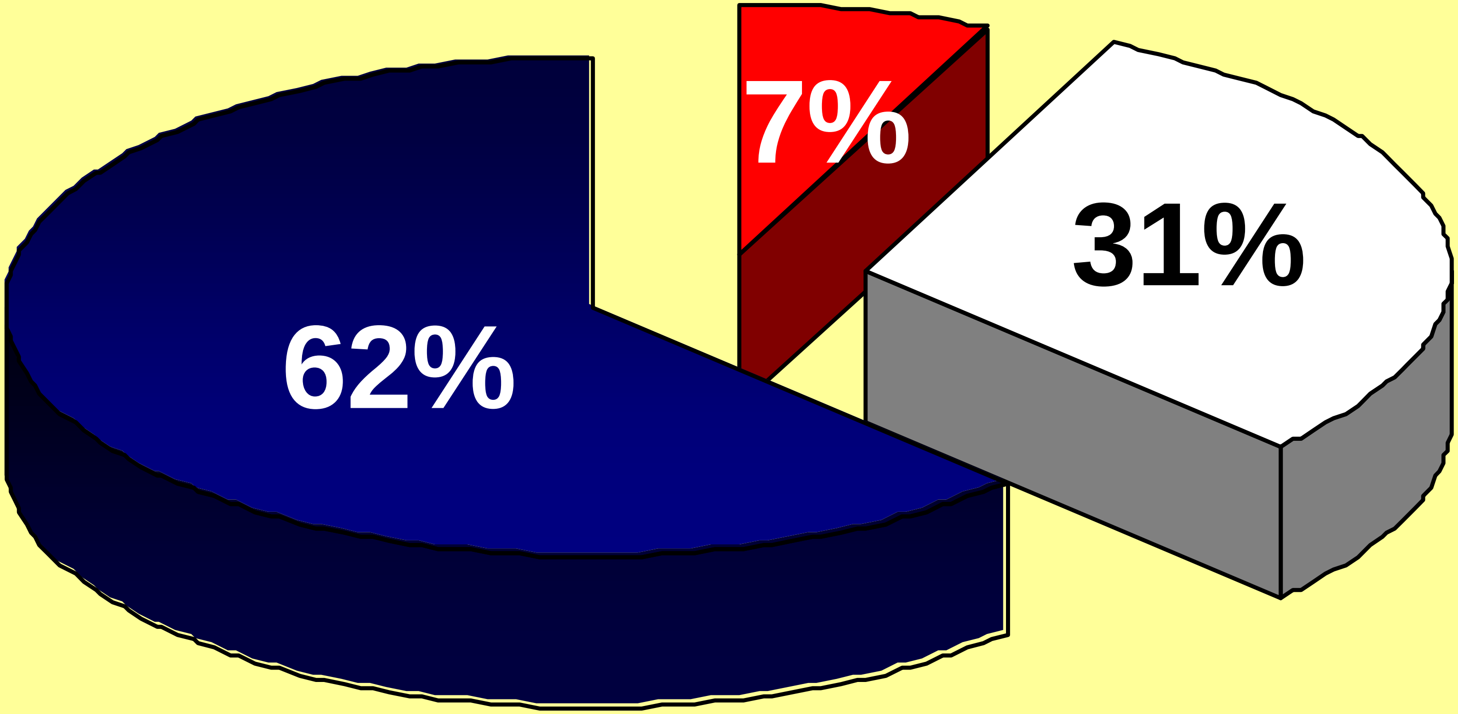
3) (GALTONOVA) OGIVA

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

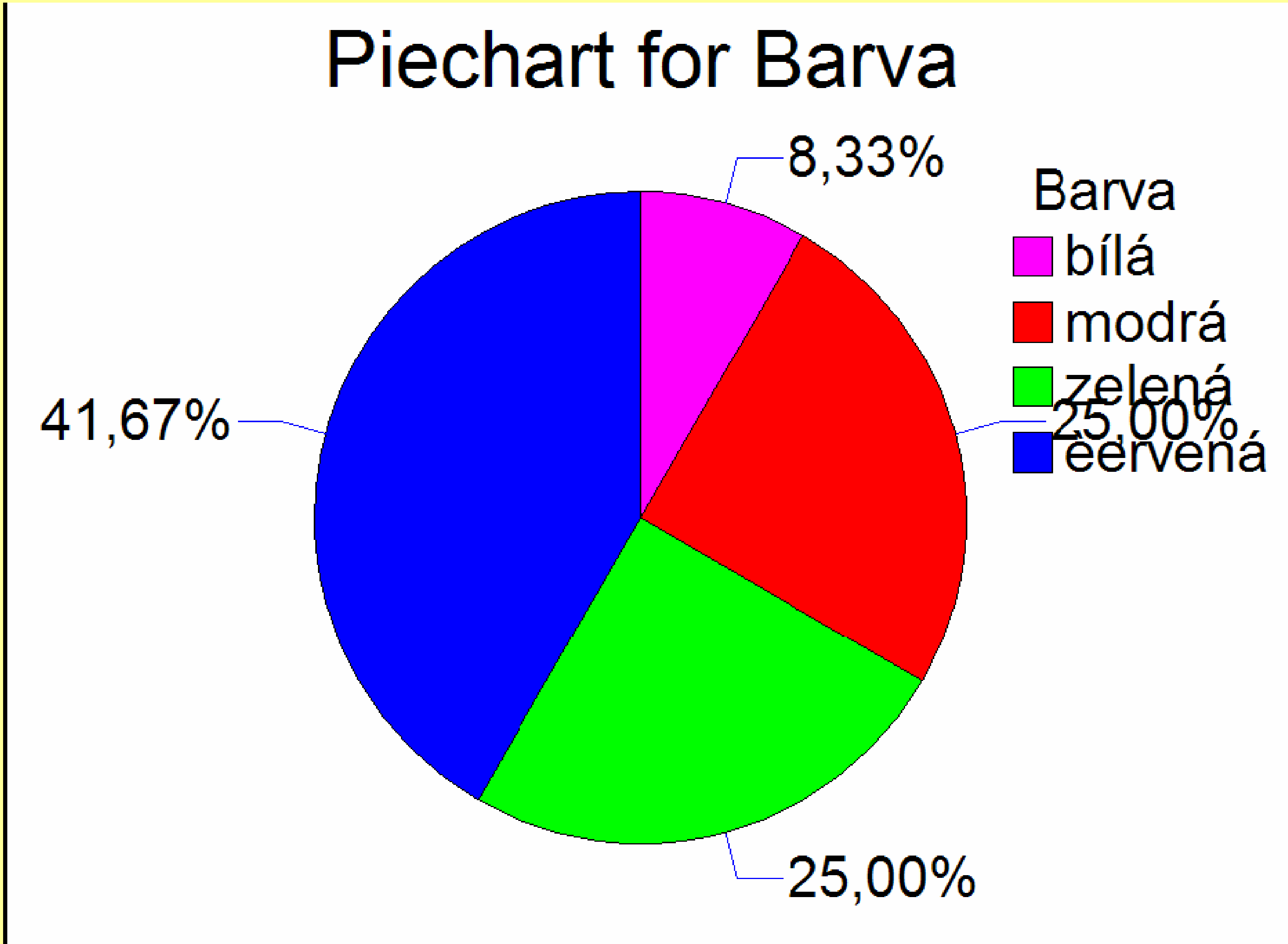


4) VÝSEČOVÝ (SEKTOROVÝ) GRAF

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



4) VÝSEČOVÝ (SEKTOROVÝ) GRAF

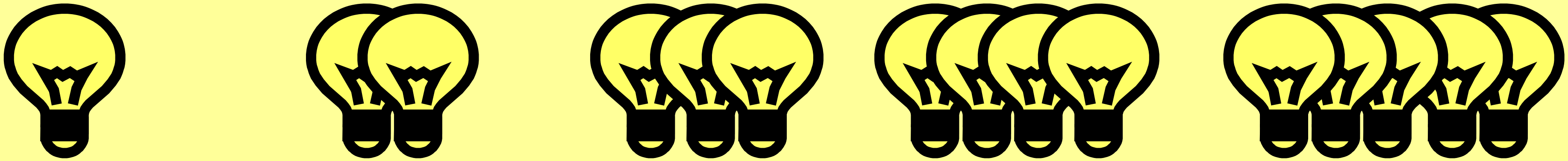


5) PIKTOGRAM

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).

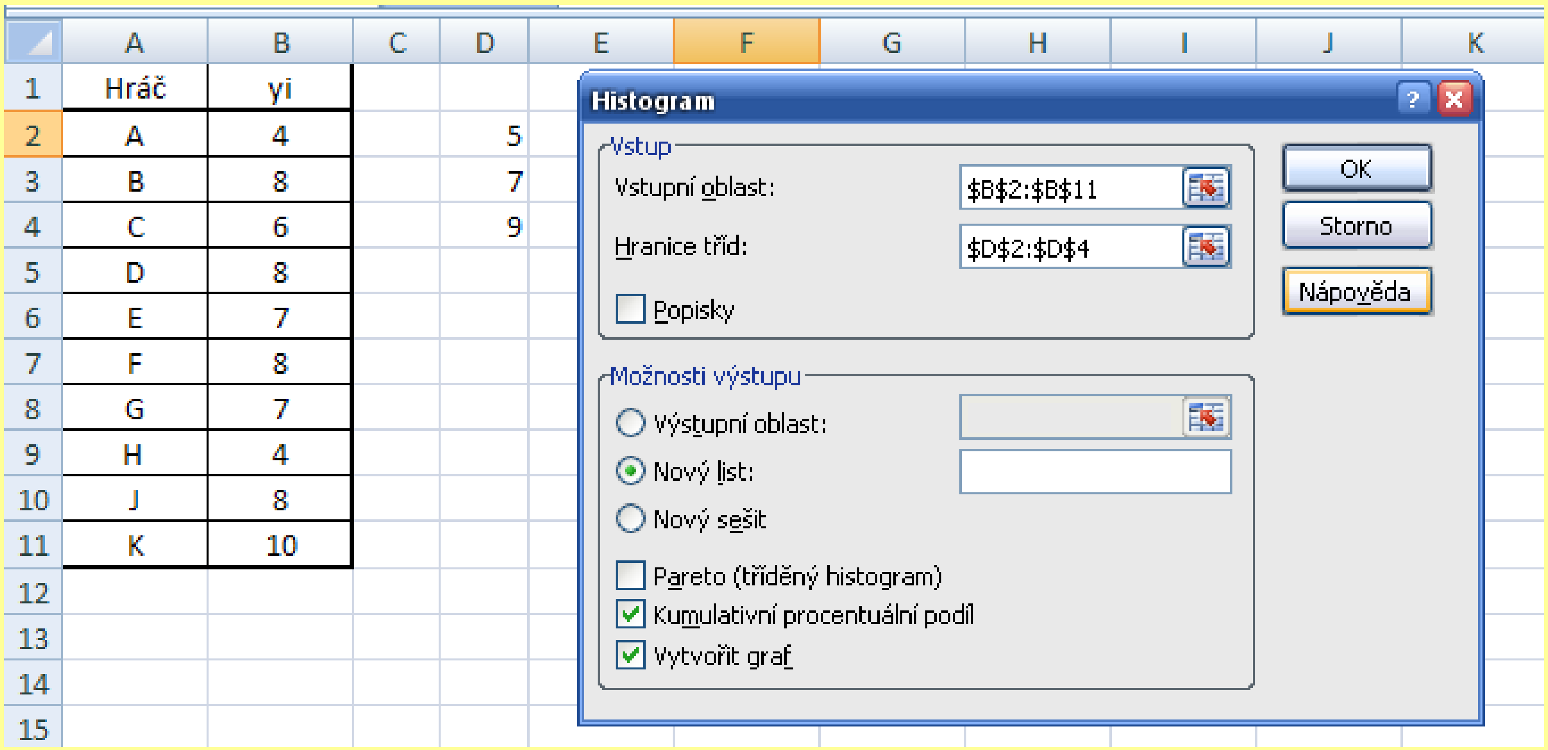
Spotřeba energie v městě X v letech

1960	1970	1980	1990	2000



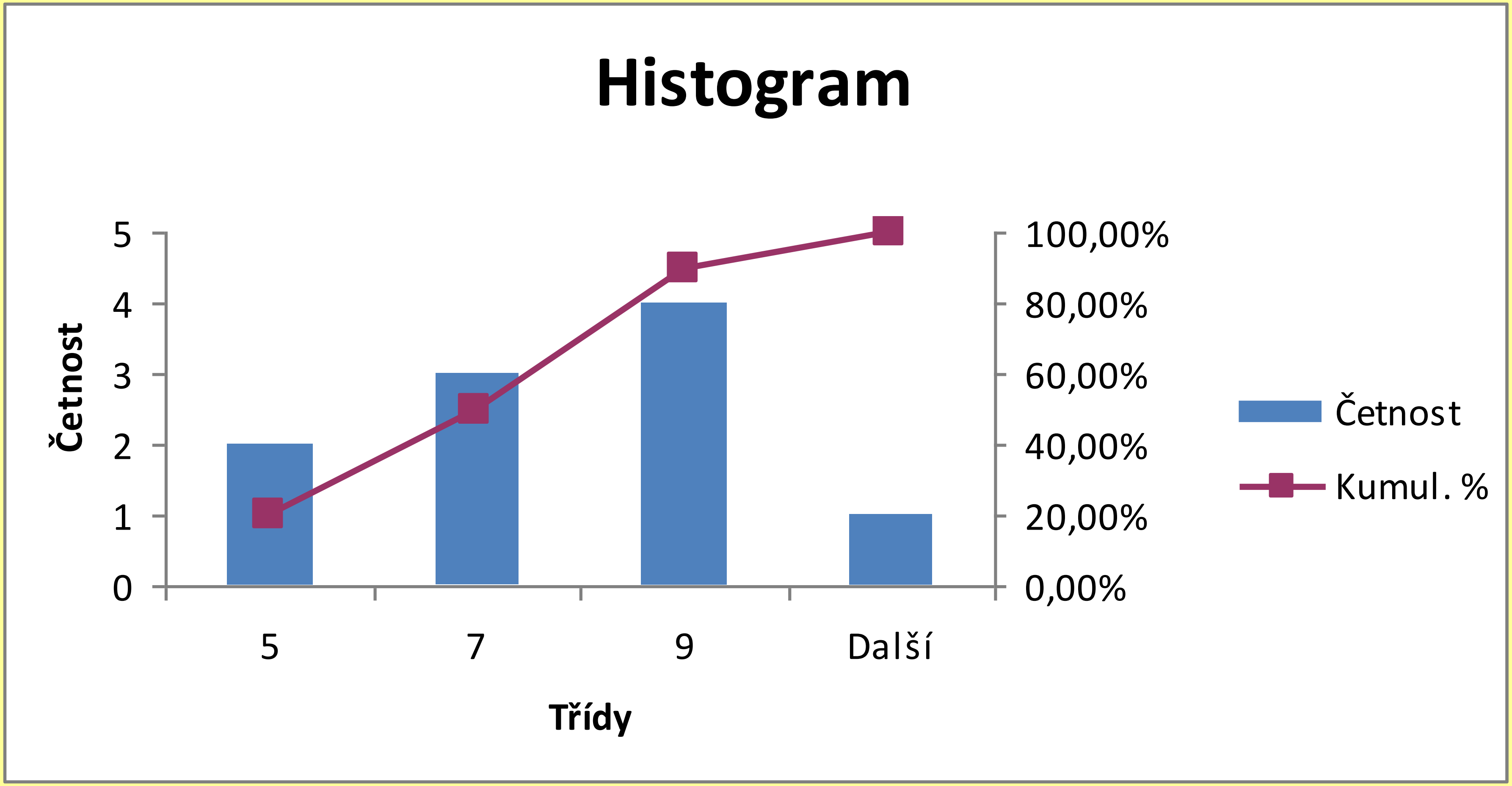
10 MW	22 MW	28 MW	43 MW	52 MW

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

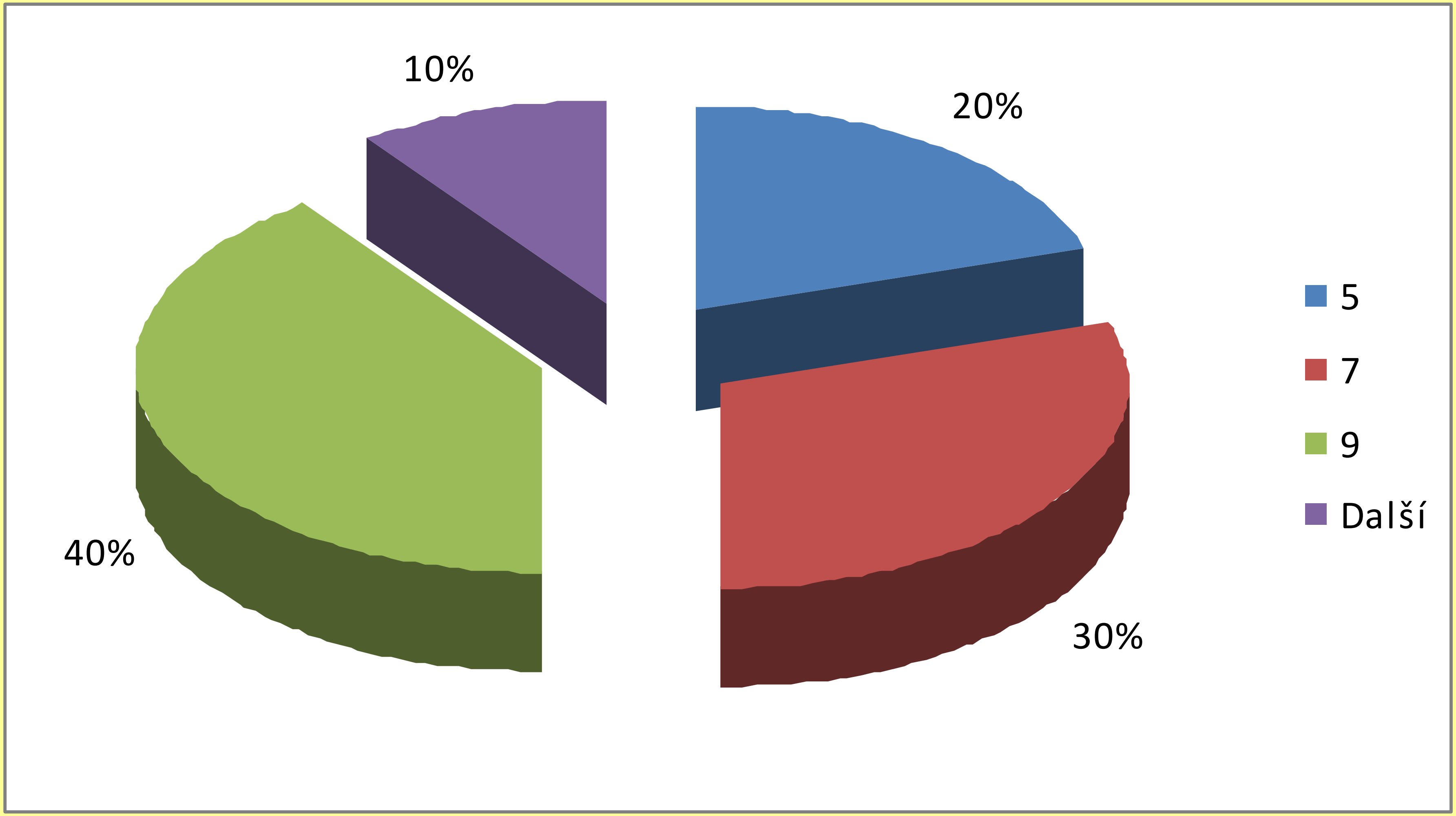


Třídy	*Četnost*	*Kumul. %*
5	2	20,00%
7	3	50,00%
9	4	90,00%
Další	1	100,00%

Histogram četností



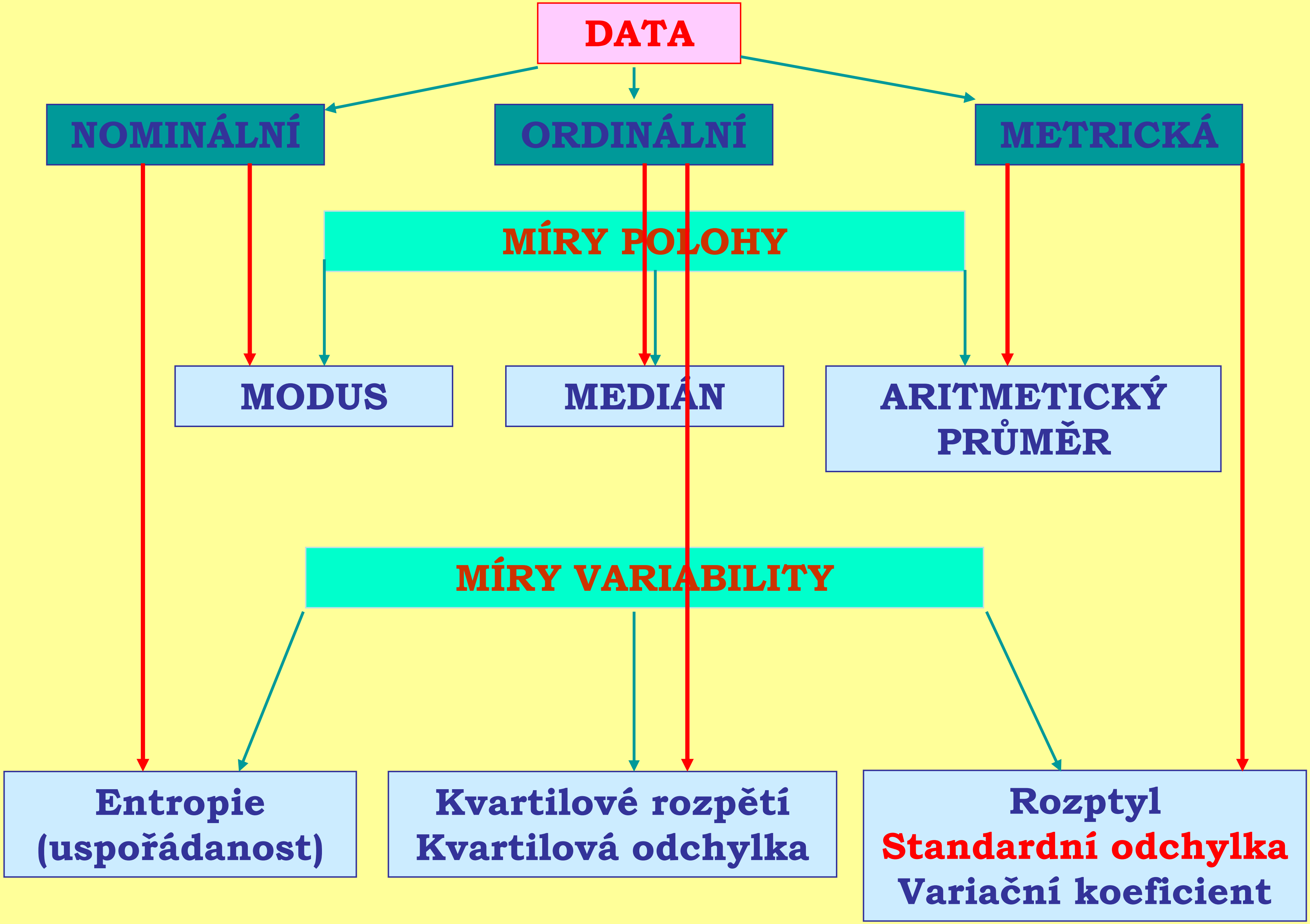
Výsečový (sektorový) graf



ZÁKLADNÍ STATISTICKÉ CHARAKTERISTIKY

PRO DATA ZÍSKANÁ NA ŠKÁLE:

NOMINÁLNÍ, ORDINÁLNÍ, METRICKÉ


DATA
NOMINÁLNÍ
ORDINÁLNÍ
METRICKÁ
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; 7; 8; 8; 8; 8; 9; 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.

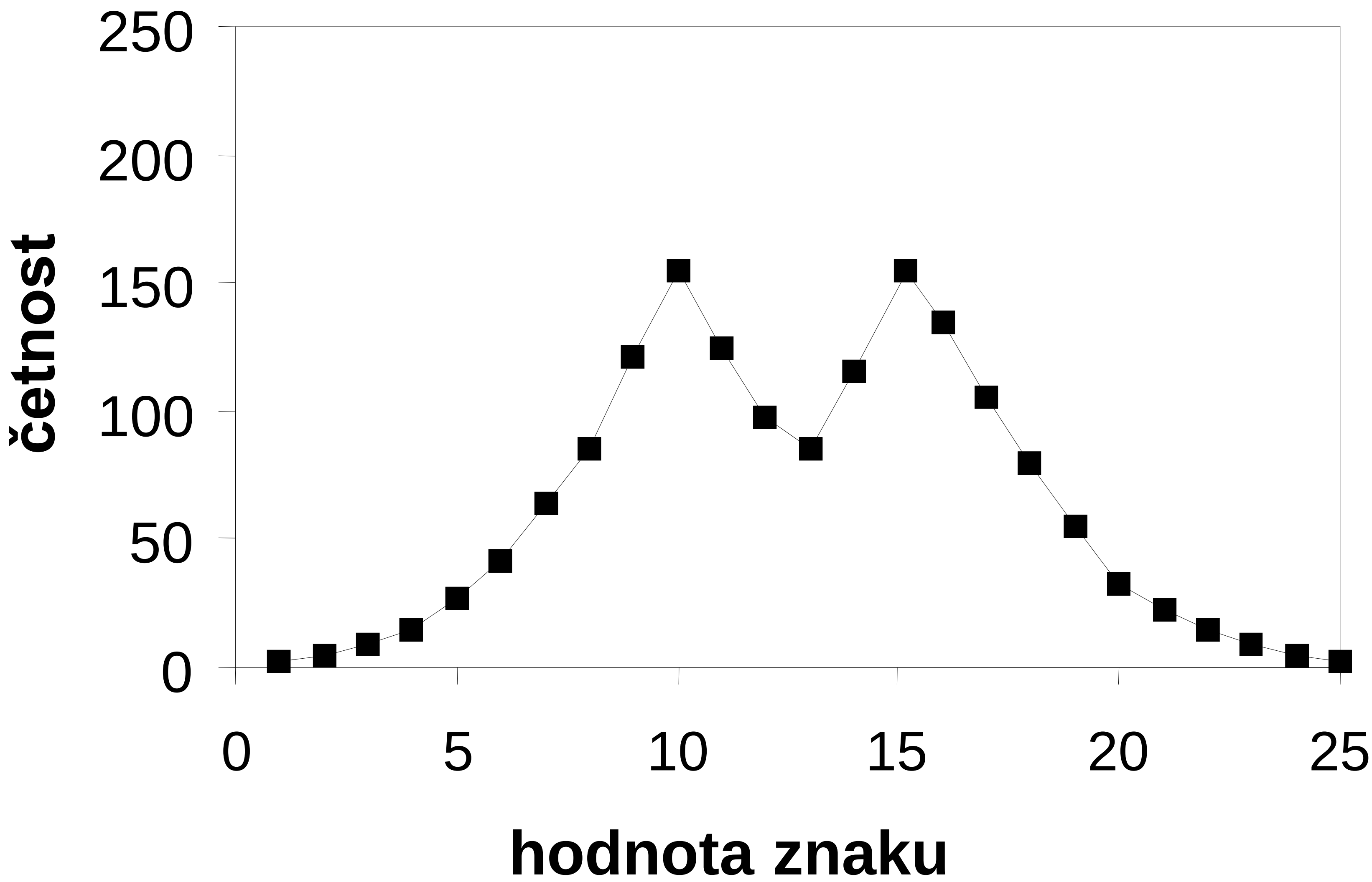
Soubor může mít jeden či více modů

(soubor bimodální, soubor trimodální).

Modus je použitelný pro nominální stupnice

(a všechny vyšší).

Rozdělení bimodální



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) **Mo = 8 Me = 8**

x_i : 6 7 7 8 8 8 8 9 9 10 10 (lichý počet)

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

$\cancel{\text{Ø}}$ … takto nikdy !

b) Vážený aritmetický průměr

- **užívá 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 FTK 1991 – 1993

Výsledky testu běh na 100m

1991 (n = 350)	AP = 13,0
1992 (n = 230)	AP = 12,5
1993 (n = 120)	AP = 12,0

Jaký je průměrný výkon v běhu na 100 m v letech 1991 – 1993?

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

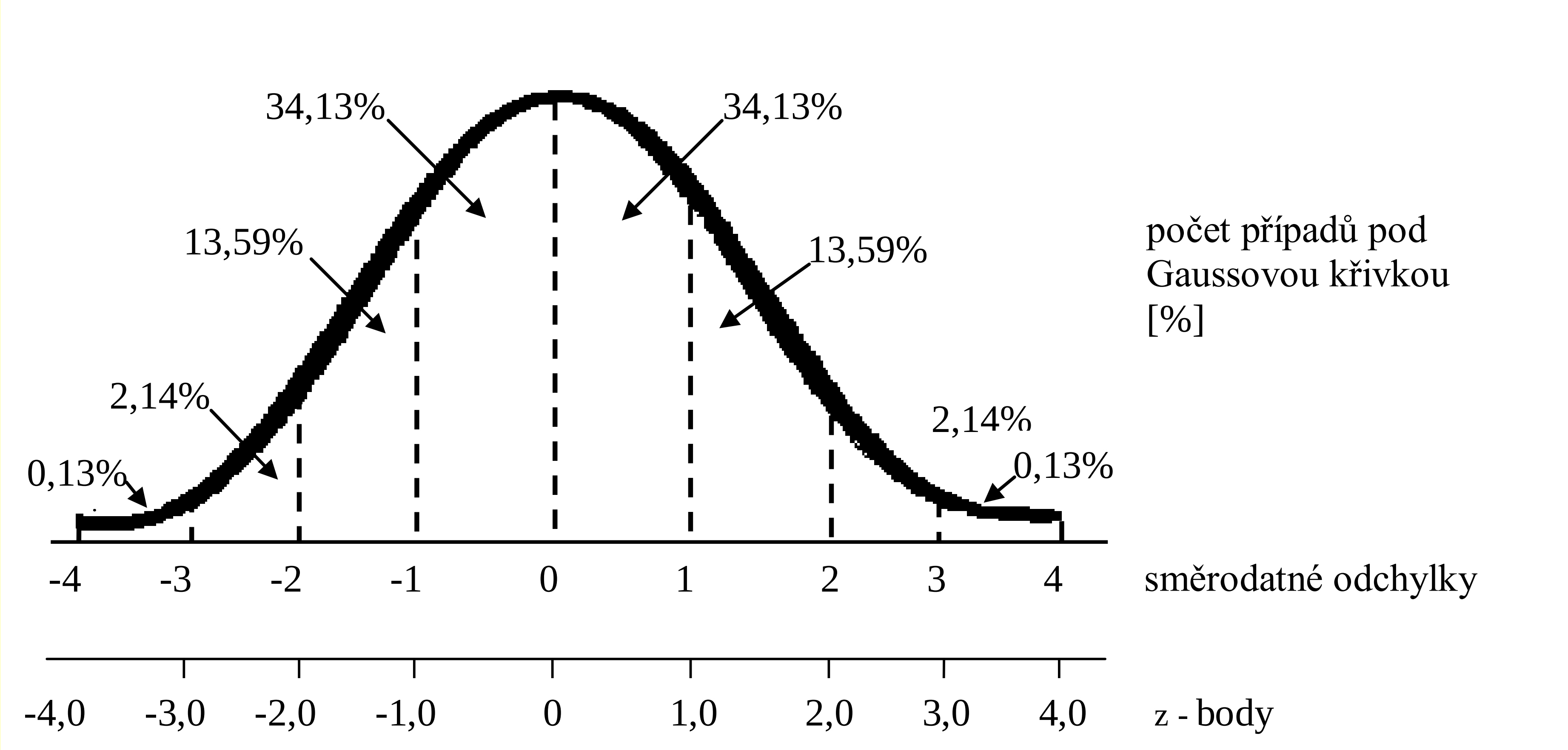
$$\bar{x} = \frac{x_1 w_1 + x_2 w_2 + \ldots + x_m w_m}{w_1 + w_2 + \ldots + w_m} = \frac{\sum_{i=1}^{m} x_i w_i}{\sum_{i=1}^{m} w_i}$$

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

Poznámky k rozložení četností a měr polohy

• **Při (Gaussově) normálním rozložení četností znaků jsou vypočítané střední hodnoty (aritmetický průměr, modus, medián) 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 = ?** **AP = ?** **Vážený AP = ?**

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

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

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

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

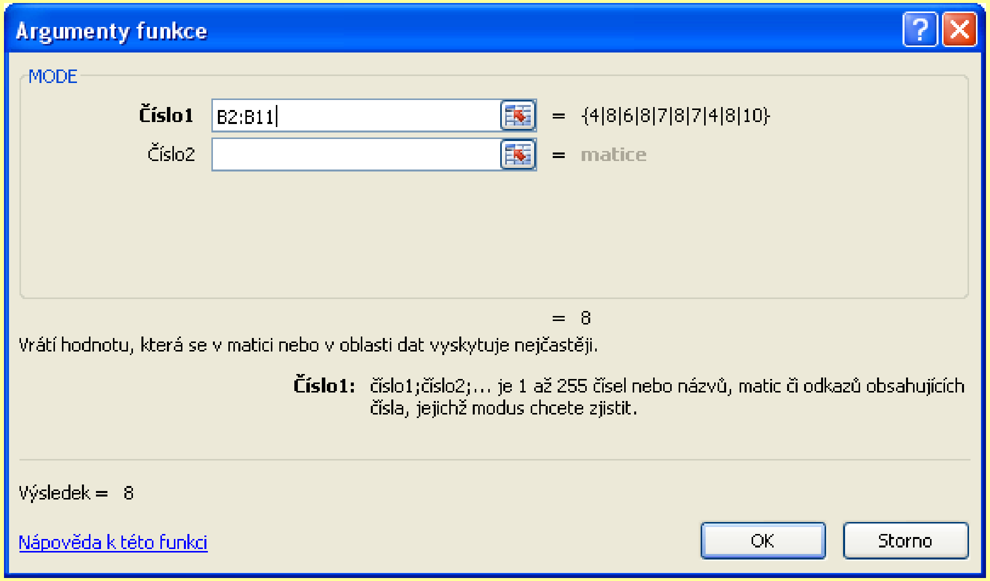
Pomocí Excelu – Statistické funkce
Výpočet: modus, medián, aritmetický průměr.

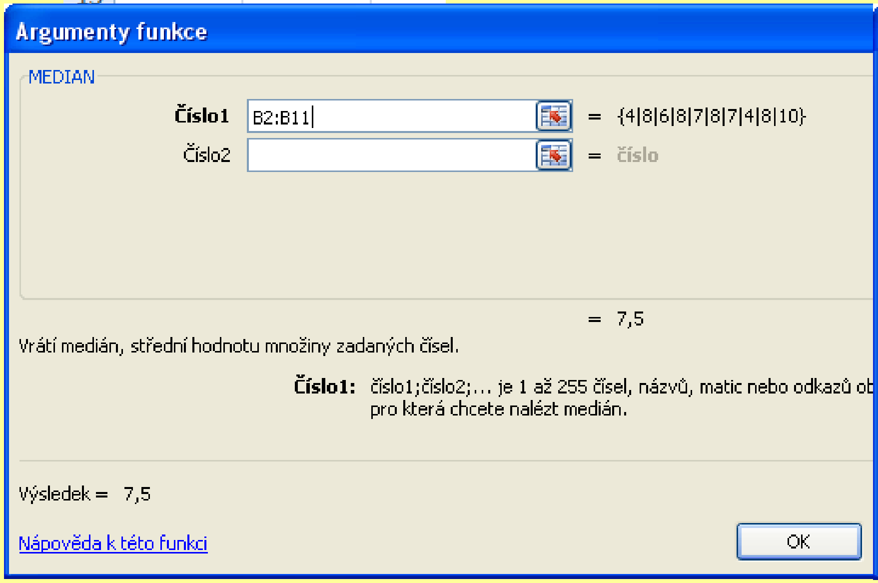
	A	B	C
1	Hráč	yi	
2	A	4	
3	B	8	
4	C	6	
5	D	8	
6	E	7	
7	F	8	
8	G	7	
9	H	4	
10	J	8	
11	K	10	
12			
13			
14			
15			

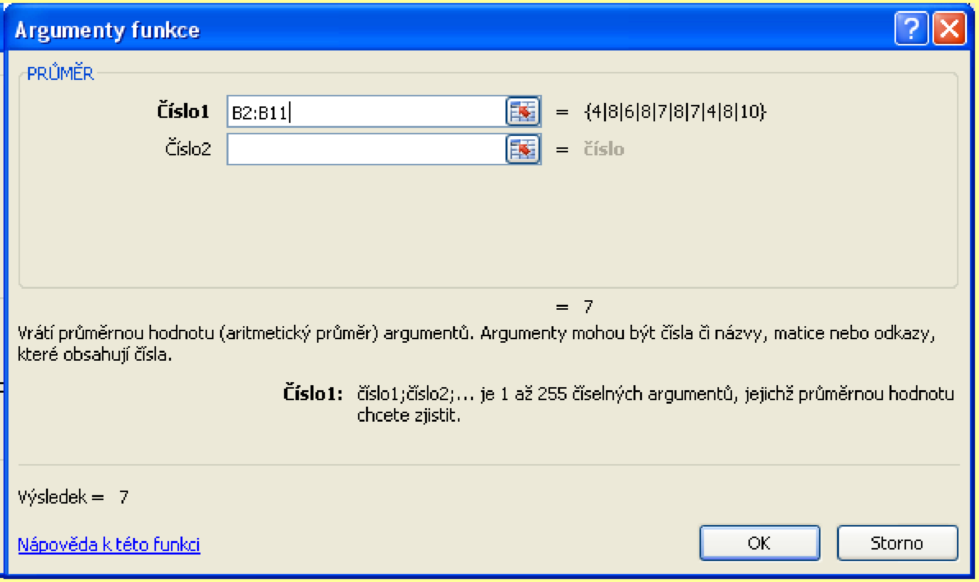
MODE

MEDIAN

PRŮMĚR



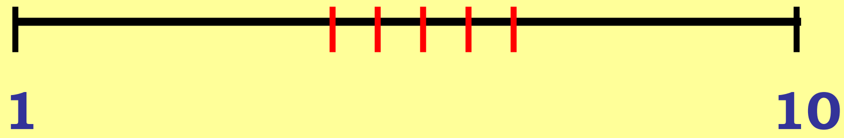


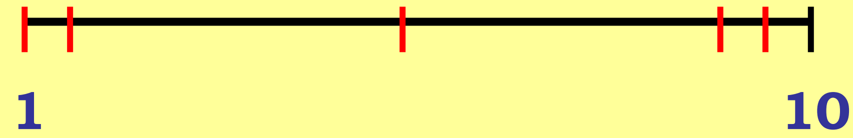


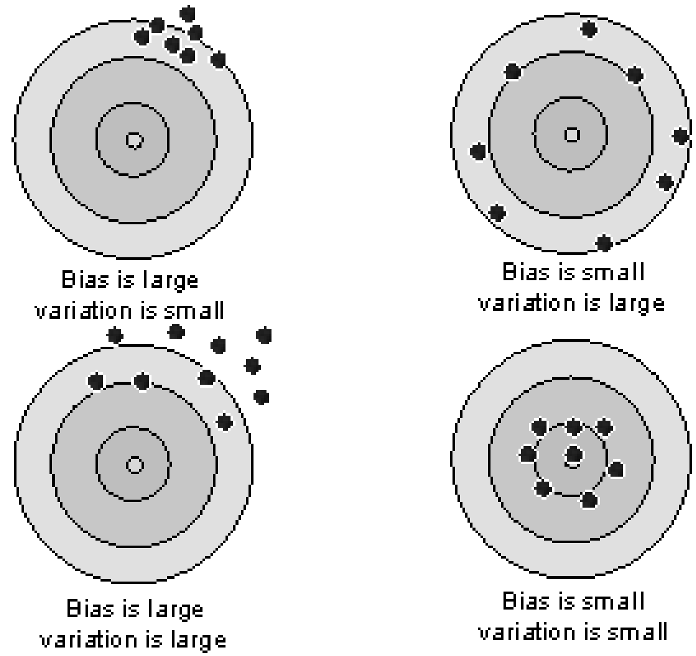
2. 3 MÍRY VARIABILITY

Popis statistického souboru pomocí měr polohy (určení 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)**







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

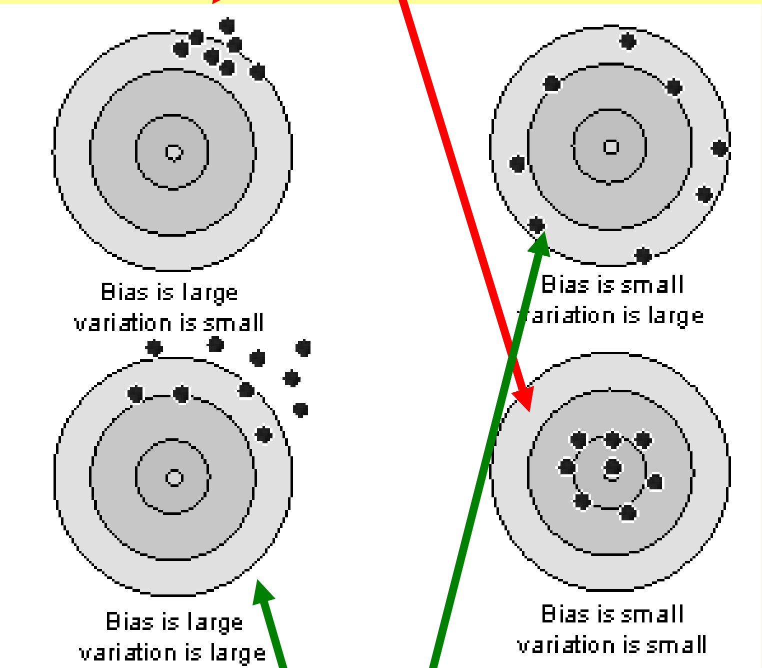
MÍRY VARIABILITY

V odborné literatuře jsou také označovány jako *míry variace, rozptýlení, měnlivosti.*

Míry variability charakterizují

- ✓ **vyrovnanost jednotek souboru,**
- ✓ **jak jsou hodnoty znaků souboru rozptýleny, jak se vzájemně odlišují,**
- ✓ **do jaké míry je sledovaný soubor homogenní (stejnorodý) resp. 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 (KMV)

KMV jsou využitelné pro stupnice ordinální a dále pro stupnice metrické v případech, kdy nelze prokázat normalitu rozložení četností dat (*proč ne pro nominální stupnice?*).

NEJČASTĚJI POUŽÍVANÉ KVM

VARIAČNÍ ŘADA = znaky statistického souboru seřazené podle velikosti.

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

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

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)

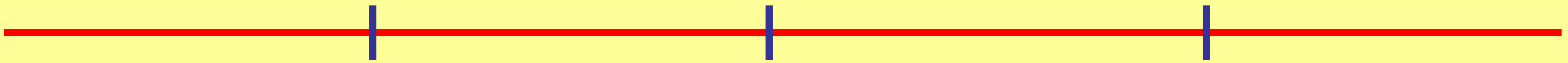
1. KVARTIL (Y) … **kvartily rozdělují variační řadu na čtvrtiny, na 4 skupiny.**

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 $n = 40$

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

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

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

$$Q = \frac{x_{90} - x_{10}}{2}$$

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

c) percentilová odchylka

$$Q = \frac{x_{99} - x_{1}}{2}$$

Je devadesáti osminou rozpětí krajních percentilů.

2.3.2 MOMENTOVÉ MÍRY VARIABILITY

Kvantilové míry variability udávají jen rozpětí, v němž se znaky pohybují.

MOMENTOVÉ MÍRY VARIABILITY umožňují výpočet číselných charakteristik, které udávají:

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

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

(pro rozsáhlé soubory)

AP…aritmetický průměr

x_i …hodnota znaku

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

Rozptyl „měří" variaci ve smyslu odlišnosti jednotlivých hodnot znaku *od průměru* i ve smyslu vzájemné odlišnosti *jednotlivých hodnot znaku.*

2. SMĚRODATNÁ (STANDARDNÍ) ODCHYLKA (s)

Symbolický tvar

$s = \sqrt{s^2}$ (var x)

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

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

(pro rozsáhlé soubory)

Směrodatná odchylka (s)

... je kvadratickým průměrem odchylek jednotlivých hodnot znaku od aritmetického průměru.

3. VARIAČNÍ KOEFICIENT

$$V = \frac{s}{|AP|}$$

$$V = 100 \times \frac{s}{|AP|}$$

Variační koeficient ...

- **umožňuje provést** ***srovnání variability dvou či více souborů,*** **jejichž znaky jsou měřeny v různých jednotkách (cm, kg, sekundy, ...),**

- **udává** ***poměr směrodatné odchylky k aritmetickému průměru,*** **přesněji řečeno udává,** ***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

AP=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 – směrodatná odchylka znaků y_i …

Variační koeficient V_1

$$V_1 = \frac{s}{AP} = \frac{1{,}09}{8} = \mathbf{0{,}14} \quad \text{resp.} \quad V_1 = \frac{s}{AP} \cdot 100 = \mathbf{14\ \%}$$

Sami - variační koeficient V_2 tj. znaků y_i …

$V_2 = 0{,}26$ resp. $V_2 = 26\ \%$ $\Rightarrow$ $V_1 < V_2$

Interpretace …

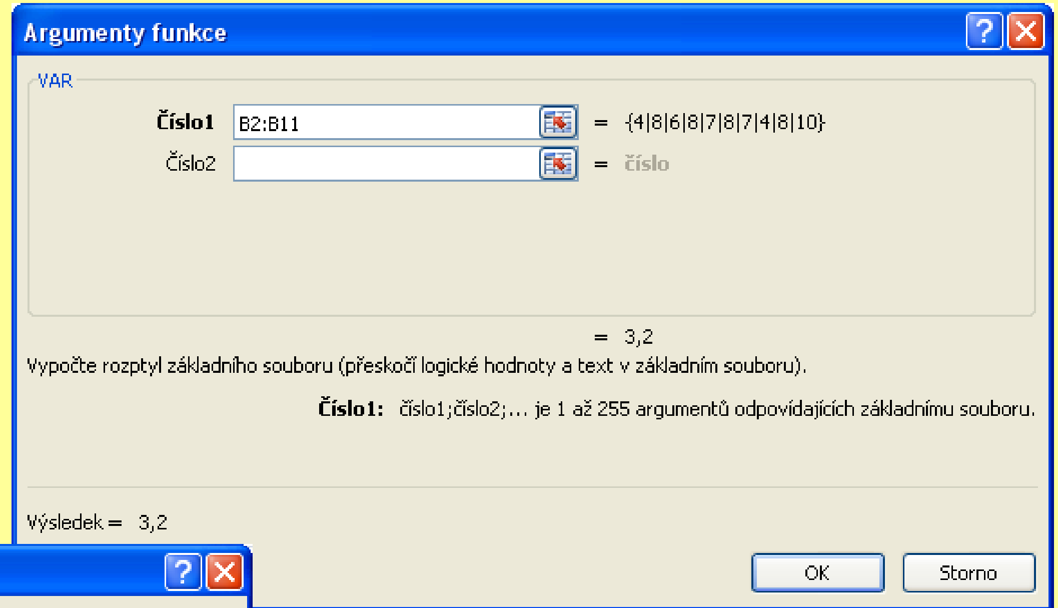
Pomocí Excelu – Statistické funkce

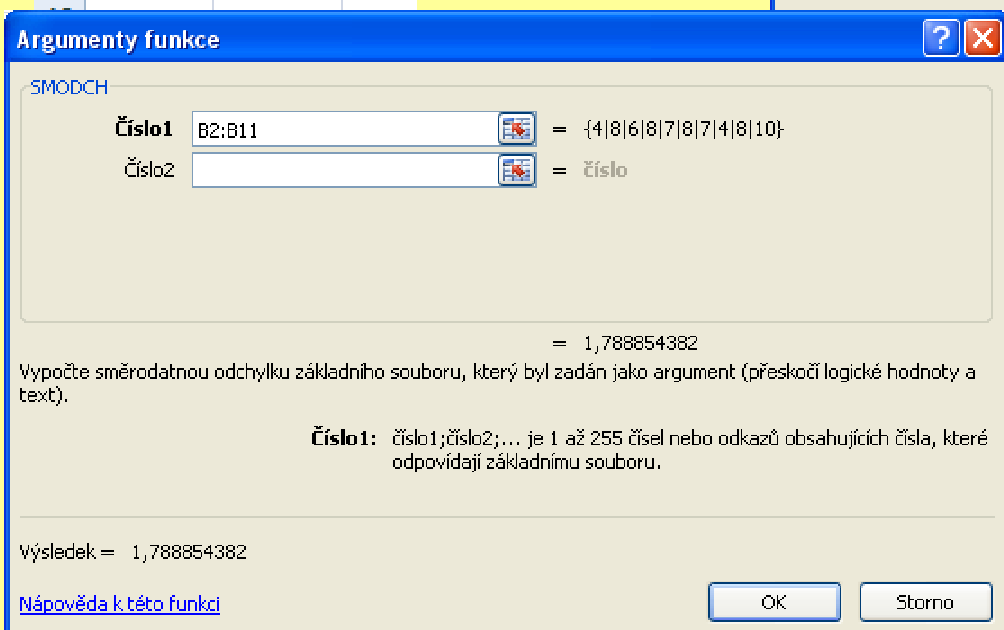
Výpočet: rozptyl, směrodatná odchylka, variační koeficient

	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



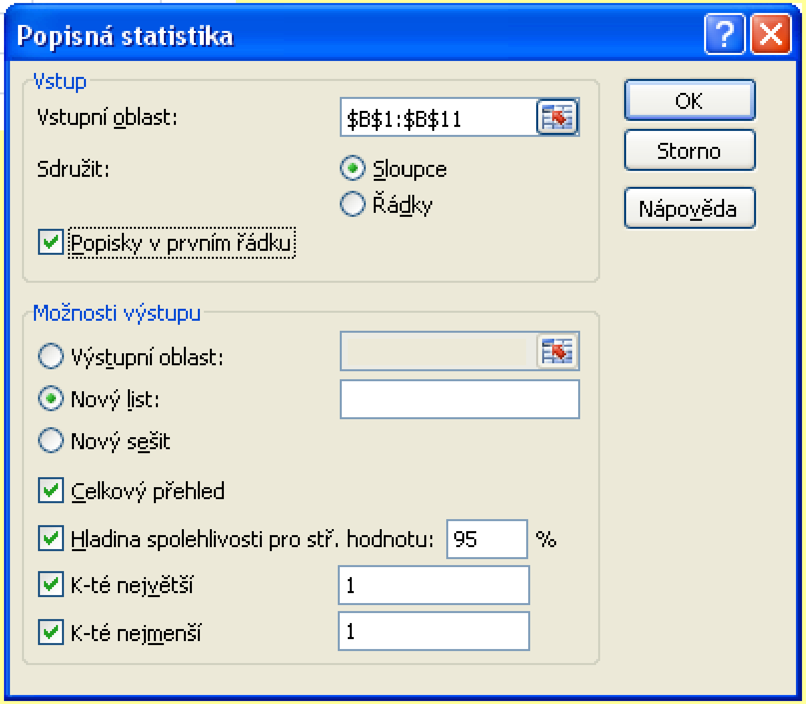


VAR.VÝBĚR
vypočte rozptyl výběru

SMODCH.VÝBĚR
vypočte směrodatnou odchylku výběru

Pomocí Excelu – Analýza dat – Popisná statistika

	A	B	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			



	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

Výsledky výpočtu pomocí Excelu:
Analýza dat – Popisná statistika

Tabulka 1. Základní statistické charakteristiky sledovaných proměnných

Sample	Female (n=65)			
Variables	M	SD	min	max
Age	10.20	0.60	9.0	10.9
Weight (kg)	36.76	6.10	25.8	53.0
Height (cm)	145.30	7.50	130.0	165.0
SR (kp)	18.90	4.82	11.0	36.6
SL (kp)	16.70	5.03	9.1	39.2

Vysvětlivky:

SR = max. strength right hand, SL = max. strength left hand,

M = mean, SD = standard deviation

ANALÝZA JEDNOROZMĚRNÉHO SOUBORU

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. VÝPOČET ZÁKLADNÍCH STATISTICKÝCH CHARAKTERISTIK

a) míry centrální tendence

b) míry variability

c) míry závislosti

2.5 MÍRY ZÁVISLOSTI

2.5.1 ZÁVISLOST PEVNÁ, VOLNÁ, STATISTICKÁ A KORELAČNÍ

Jednorozměrné soubory **- charakterizovány jednotlivými statistickými znaky (výkon ve skoku do dálky, v běhu na 100m, tělesná výška, tělesná hmotnost, ...**

Existence *souvislostí mezi znaky*:

- **úspěšnost střelby 1. a 2. pokus,**
- **tělesná výška x hmotnost,...**

2.5.1 ZÁVISLOST PEVNÁ, VOLNÁ, STATISTICKÁ A KORELAČNÍ

Míry závislosti se zabývají hledáním, zkoumáním a hodnocením souvislostí (závislostí, vztahů) mezi dvěma (či více) statistickými znaky.

Závislosti znaků, věcí a jevů mohou být velmi rozmanité:

- **nepodstatné (náhodné)**
- ***příčinné (kauzální) závislosti)*** **jsou výrazem určité vnitřní nutnosti (příčina vyvolává následek)**

Příčinná (kauzální) závislost **je závislost, kdy daný jev či několik jevů (příčina)** ***nutně vyvolává*** **za určitých podmínek jiný jev (následek, účinek).**

Nejjednodušší formy ***kauzálních závislostí*** **se vyskytují u přírodních jevů např. ....**

... při zahřívání tělesa za konstantních podmínek (elementární příčina) dochází ke zvětšování jeho objemu (elementární účinek) => tj. princip teploměru.

1. PEVNÁ ZÁVISLOST

Pevná závislost **= případ, kdy výskytu jednoho jevu**
NUTNĚ ODPOVÍDÁ výskyt druhého jevu.

Tedy *jedné hodnotě jedné proměnné odpovídá jen jedna hodnota jiné proměnné (funkční závislost).*

Např. Zahříváme-li těleso 5 min, vzroste teplota o 10 ° C.
Zahříváme-li těleso 10 min, vzroste teplota o 20 ° C, atd.
… (vědy o sportu?)

PEVNÁ ZÁVISLOST

Pevná závislost – charakteristika:

- **se opakuje ve všech jednotlivých případech (při dodržení standardních podmínek).**

- **může být tedy charakterizována jediným pozorováním** (větší počet pozorování slouží k ověření výsledků a vyloučení chyb).

- **setkáváme se s ní při formulování zákonitostí vztahů mezi proměnnými**

  (např. fyzikální zákony = Archimedův zákon).

2. VOLNÁ ZÁVISLOST

Volná závislost **= případ, kdy výskyt jednoho jevu** **OVLIVŇUJE výskyt druhého jevu (NE nutně odpovídá).**

Tedy každé hodnotě jedné proměnné odpovídají různé hodnoty jiné proměnné (statistická závislost).

Např. TV x TH

Volnou závislost lze zkoumat pouze na základě mnoha pozorování, jediné pozorování může přinést naprosto nahodilý výsledek.

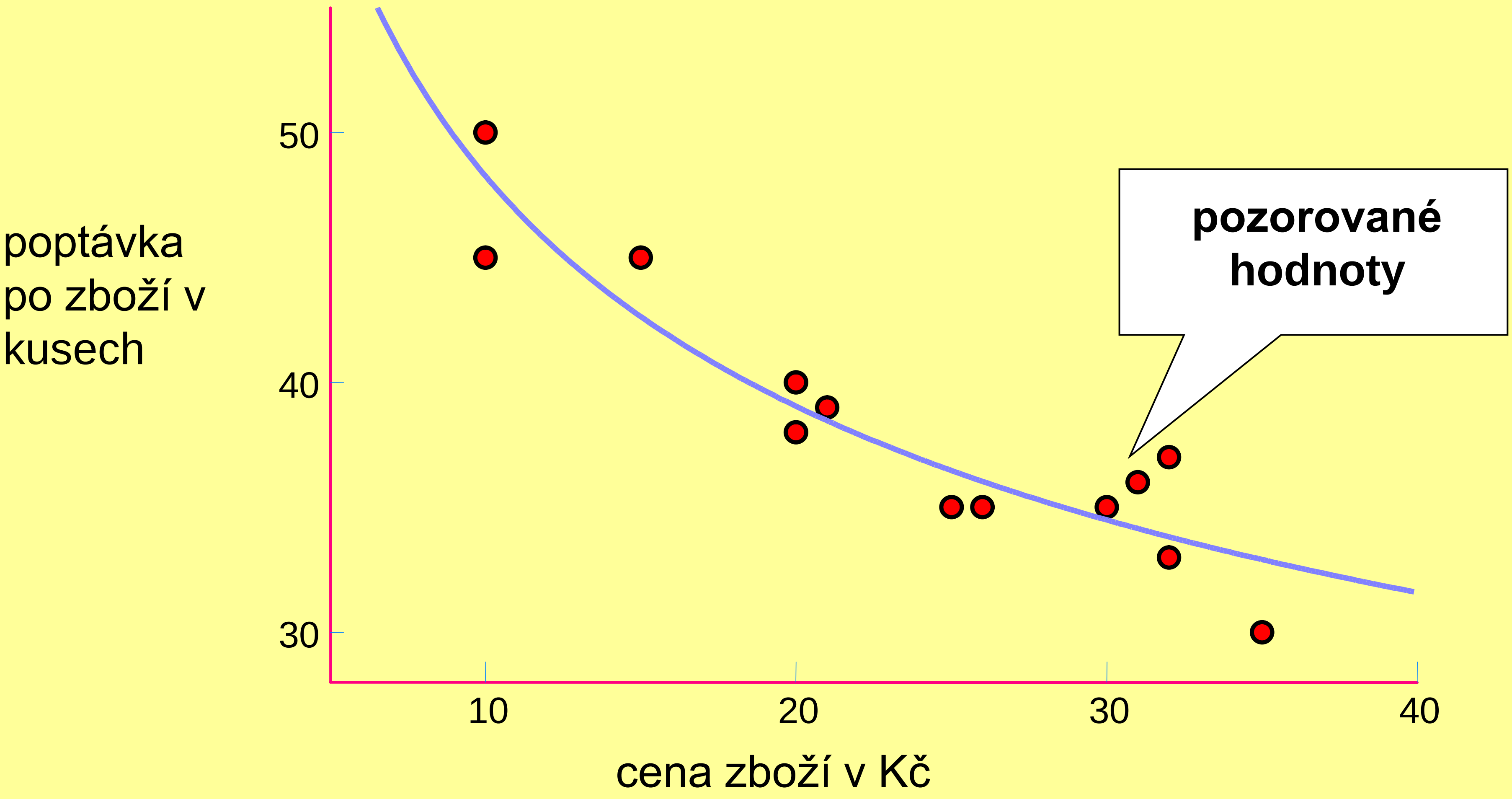
VOLNÁ ZÁVISLOST

- ***Nutný dostatečný rozsah souboru*** **(u malých souborů se může projevit vliv náhodných a vedlejších činitelů).**
- **Důležitou roli hraje volba *vhodných statistických znaků* (tedy takových, které postihují jevy co nejpřesněji),**
- **Při zkoumání společenských jevů se většinou nesetkáváme s pevnou závislostí ale s volnou, kdy určitá příčina vede k různým účinkům.**

Např. skok daleký: rychlost x délka skoku (volná z.)

VOLNÁ ZÁVISLOST

tržní poptávka



2.5.2 REGRESNÍ A KORELAČNÍ ANALÝZA

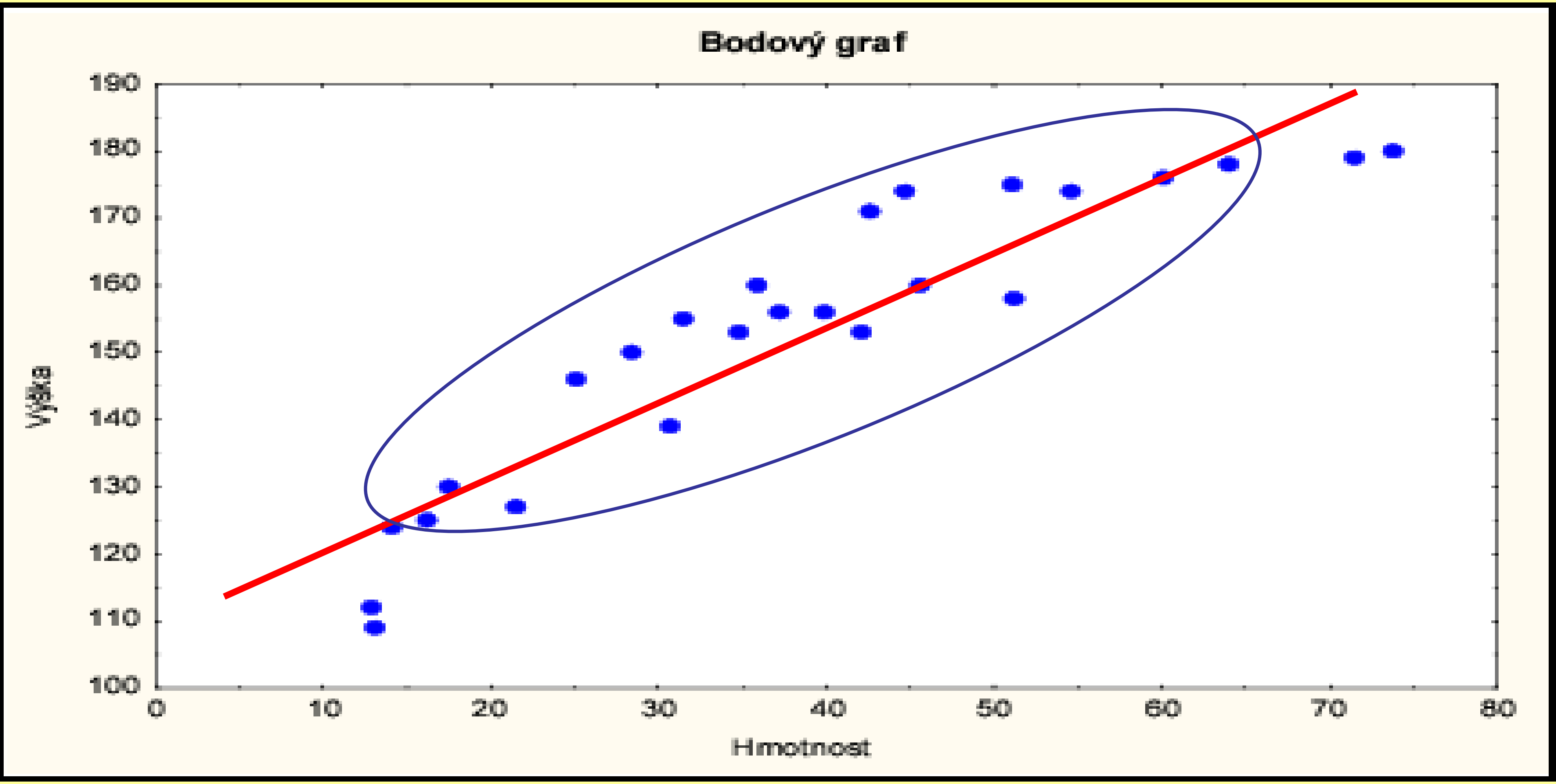
Metody regresní a korelační analýzy **slouží k poznání a matematickému popisu statistických závislostí; jsou souhrnně označované jako** ***korelační počet.***

Hlavní úkoly korelačního počtu:

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

Hlavní úkoly korelačního počtu:

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

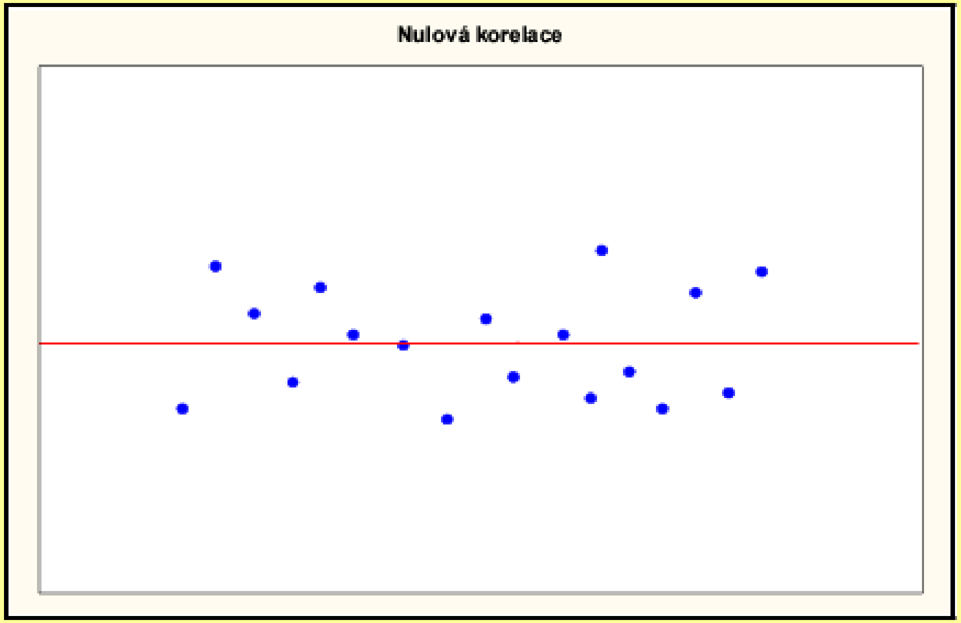


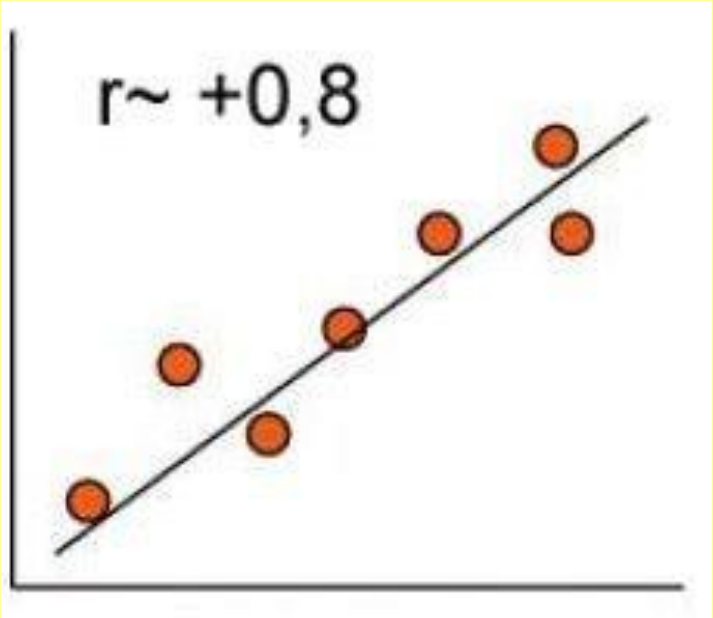
HLAVNÍ ÚKOLY KORELAČNÍHO POČTU

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

Povaha korelační závislosti je nejčastěji vyjadřována matematickou funkcí - hovoříme o *regresní funkci.*

Tento úkol korelačního počtu nazýváme *regresní analýza.*



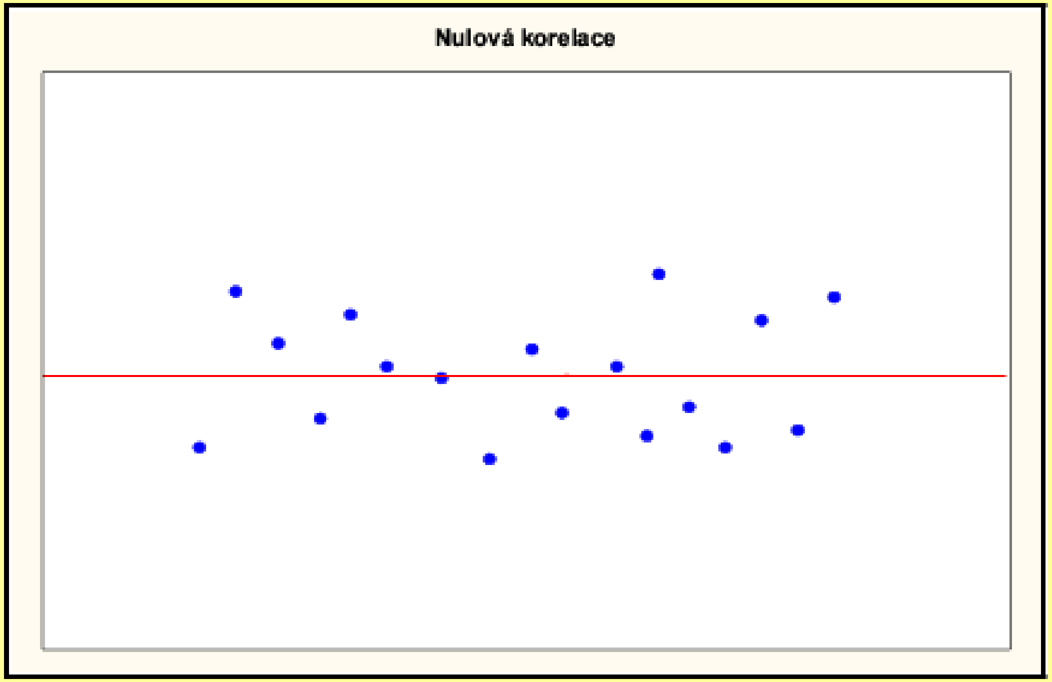


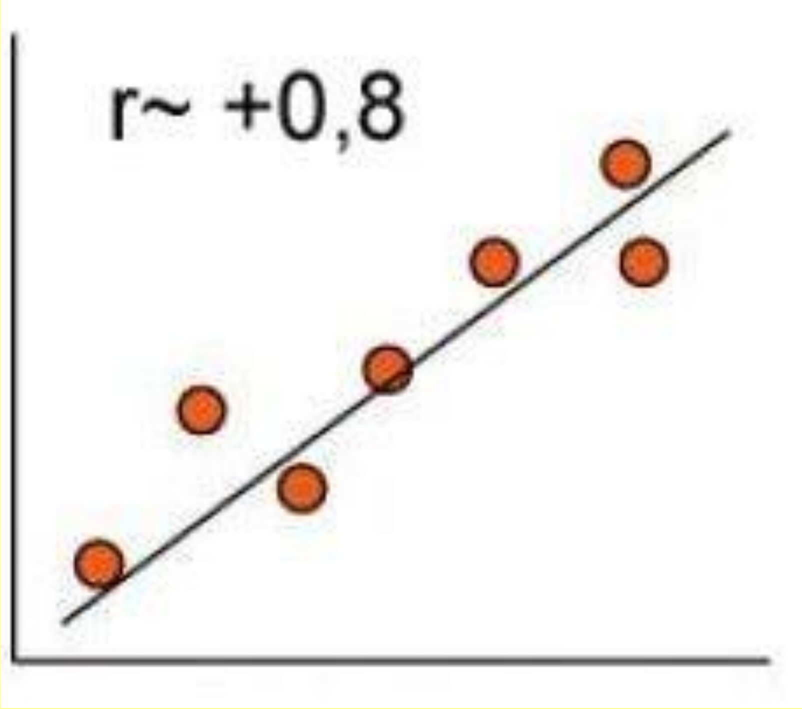
HLAVNÍ ÚKOLY KORELAČNÍHO POČTU

2. měření těsnosti **korelační závislosti umožňuje posuzovat** ***přesnost regresních odhadů*** **a** ***míru korelační závislosti*** **- hovoříme o** ***vlastní korelaci.***

Korelace je vyjadřována tzv. ***korelačním koeficientem.***

Tento úkol korelačního počtu nazýváme ***korelační analýza.***





2.5.3 REGRESNÍ ANALÝZA (LINEÁRNÍ)

1. ÚKOL „POSTIŽENÍ POVAHY KORELAČNÍ ZÁVISLOSTI“

Regresní analýza **umožňuje postihnout povahu závislosti pomocí** ***matematické funkce,*** **která by co nejlépe vyjadřovala charakter zkoumané závislosti.**

Hledaná matematická funkce se nazývá regresní funkce a je vyjádřena regresní rovnicí.

Regresní funkce **může nabývat mnoha typů:**

- ***lineární regrese***
- ***kubická regrese***
- ***hyperbolická regrese***
- ***kvadratická regrese***
- ***polynomická regrese***
- ***logaritmická regrese***

...a mnohé jiné...

Nejjednodušší z nich je LINEÁRNÍ REGRESNÍ FUNKCE, která má ve své empirické podobě tvar

$$\boxed{Y = a + b \, . \, x}$$

Pro vyjádření regresní funkce konkrétní závislost (např. tělesné výšky a hmotnosti) je třeba určit tzv. regresní koeficienty a, b, přičemž vycházíme z empirických údajů (měřených znaků) sledované závislosti.

Pro výpočet regresních koeficientů a, b je výhodné použít následující vzorce

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

2.5.4 KORELAČNÍ ANALÝZA (LINEÁRNÍ)

2. ÚKOL „MĚŘENÍ TĚSNOSTI KORELAČNÍ ZÁVISLOSTI“

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

Korelace je nejobecněji definována jako…

…volná kvantitativní závislost dvou či více jevů.

Korelace vyjadřuje míru (těsnost, stupeň) závislosti a je charakterizována číslem, tzv. korelačním koeficientem (r), který..

…„měří“ těsnost závislosti popsané lineární regresní funkcí.

VZORCE PRO VÝPOČET KORELAČNÍHO KOEFICIENTU

Symbolická podoba vzorce pro výpočet 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 tedy matematicky *podíl kovariance a součinu obou směrodatných odchylek* (viz dále).

Metrická data $\Rightarrow$

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 součinové korelace - výpočtový tvar

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

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

*Pro **metrická data** užíváme k výpočtu míry těsnosti korelační závislosti tzv. Pearsonův koeficient součinové korelace.*

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

*Pro **ordinální data** užíváme k výpočtu míry těsnosti korelační závislosti tzv. Spearmanův koeficient pořadové korelace.*

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

VLASTNOSTI KORELACE

1. VELIKOST KORELACE

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

Význam hodnot 0, 1, -1

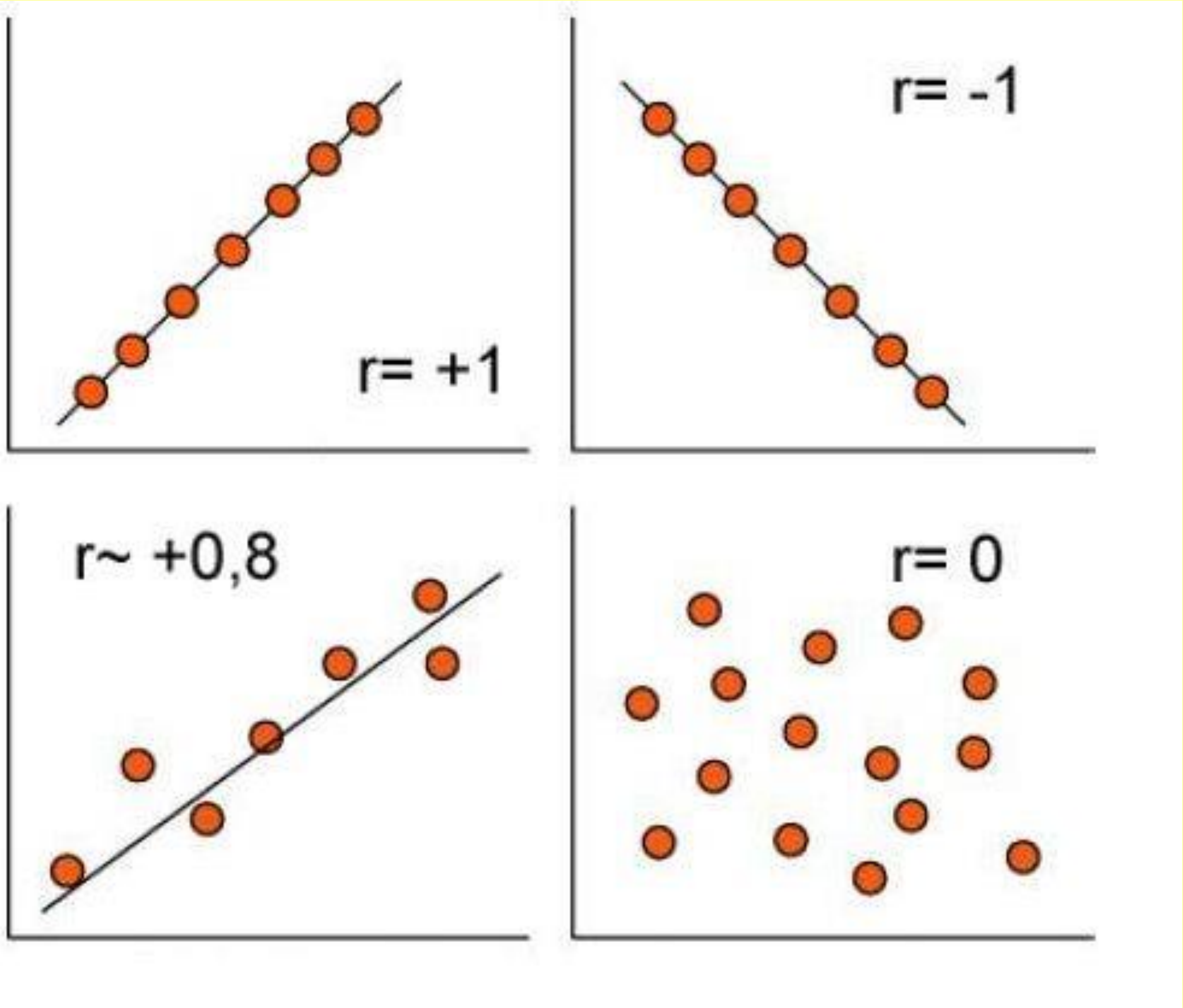
r = 0 ⇐ lineární nezávislost proměnných

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

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

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

čím více se r blíží *hodnotě 0*, tím považujeme *závislost za slabší.*

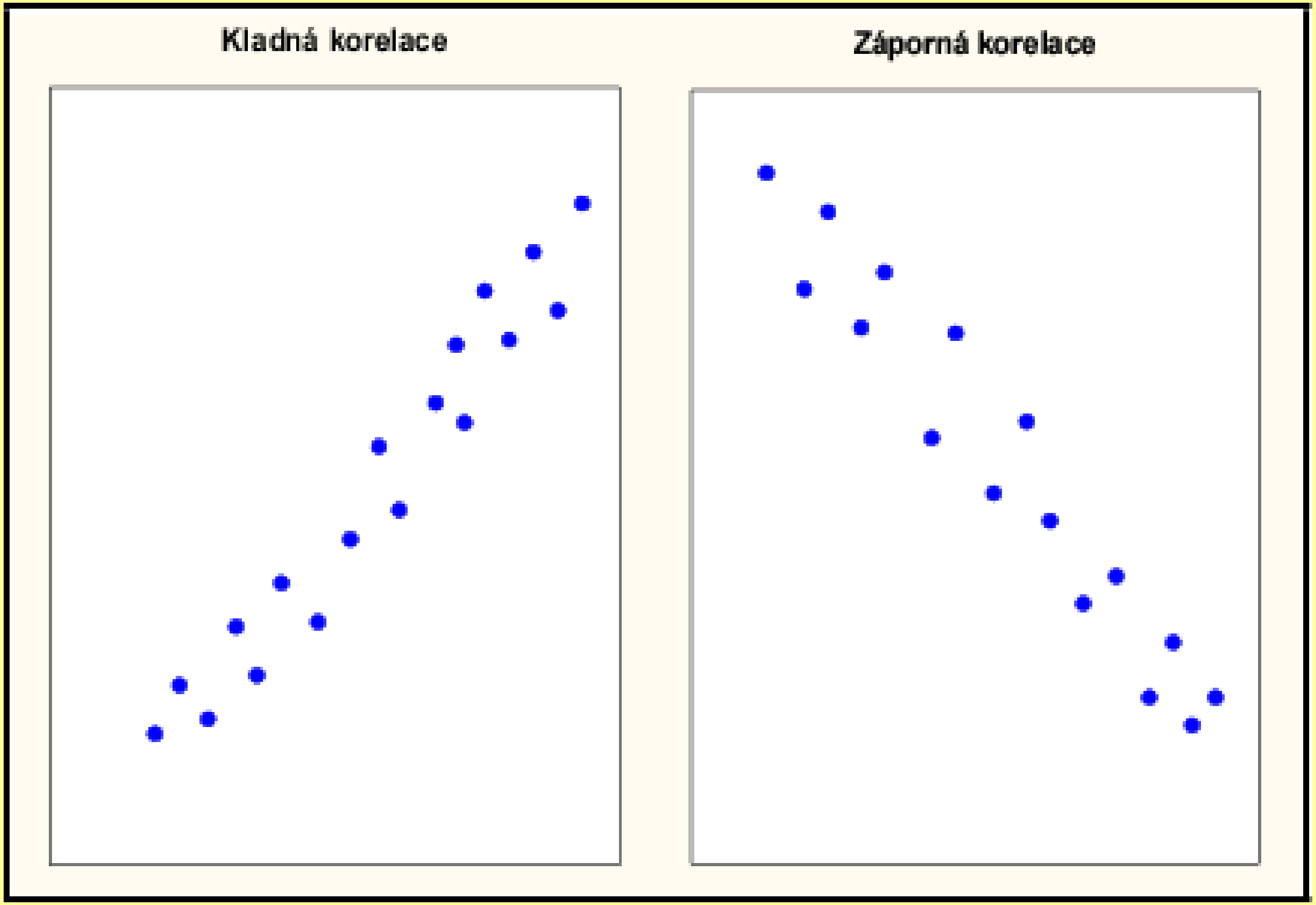


r= +1
r= -1
r~ +0,8
r= 0

2. SMĚR KORELACE

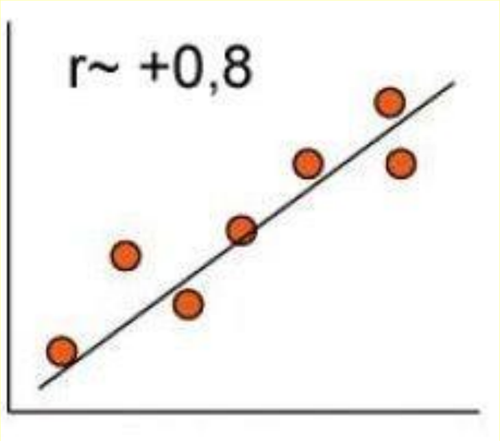
a) **kladná (pozitivní) <0;1>**

b) **záporná (negativní) <-1;0>**

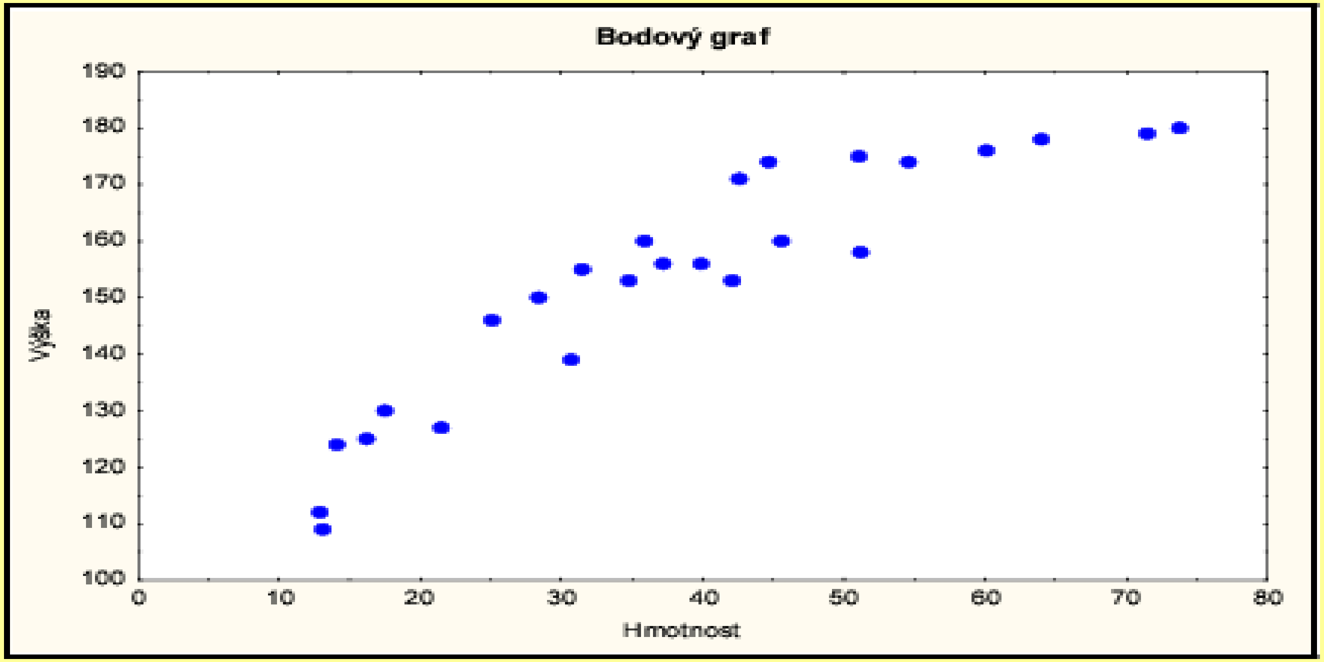


3. TVAR KORELACE

a) lineární (lze dosti dobře proložit přímku)



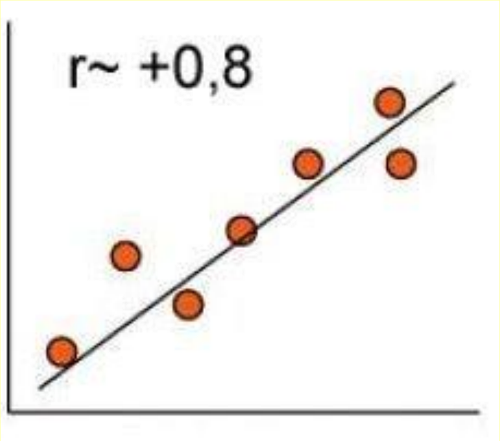
b) nelineární (nelze proložit přímku)



POZNÁMKY KE KORELACÍM

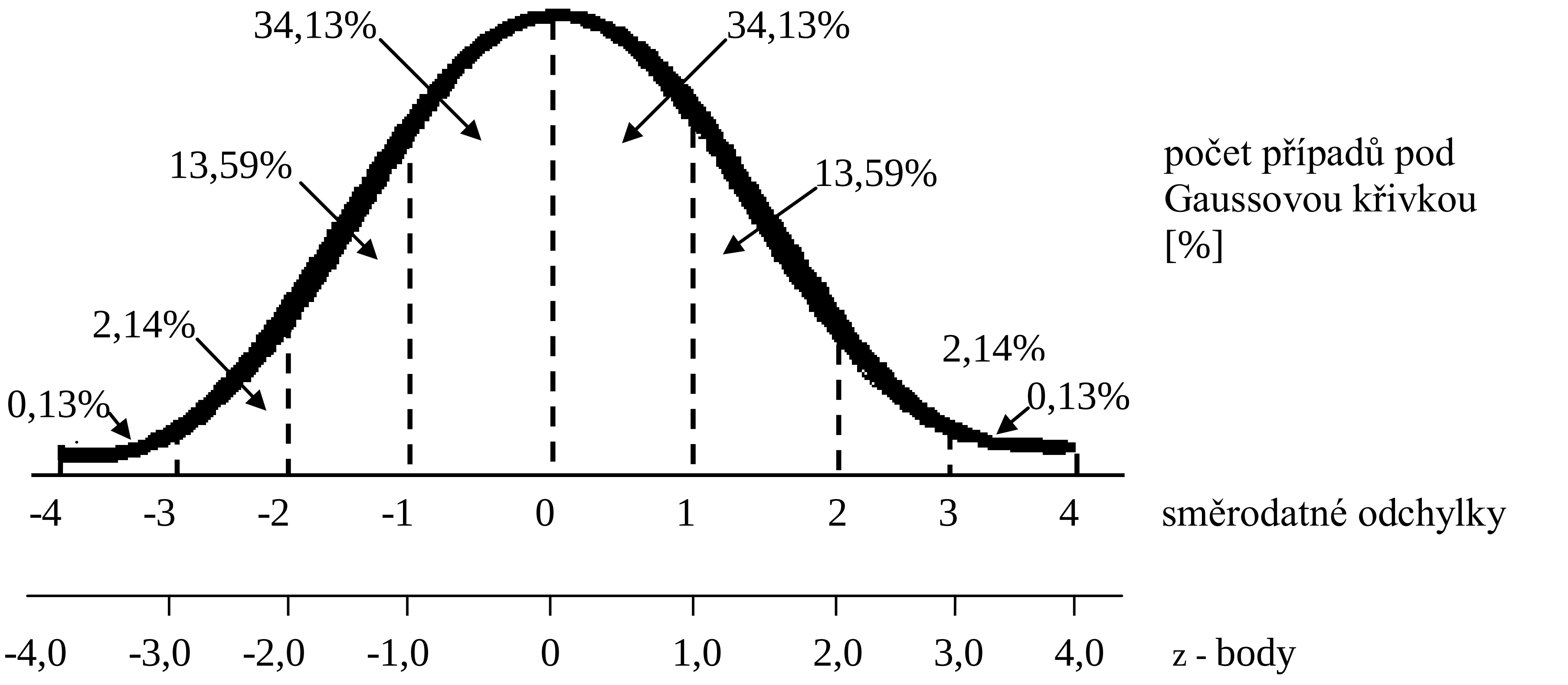
1. Matematicko-statistické předpoklady výpočtu korelačního koeficientu:

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 FTK (n=185) v běhu na 100m …

… a ve skoku dalekém je

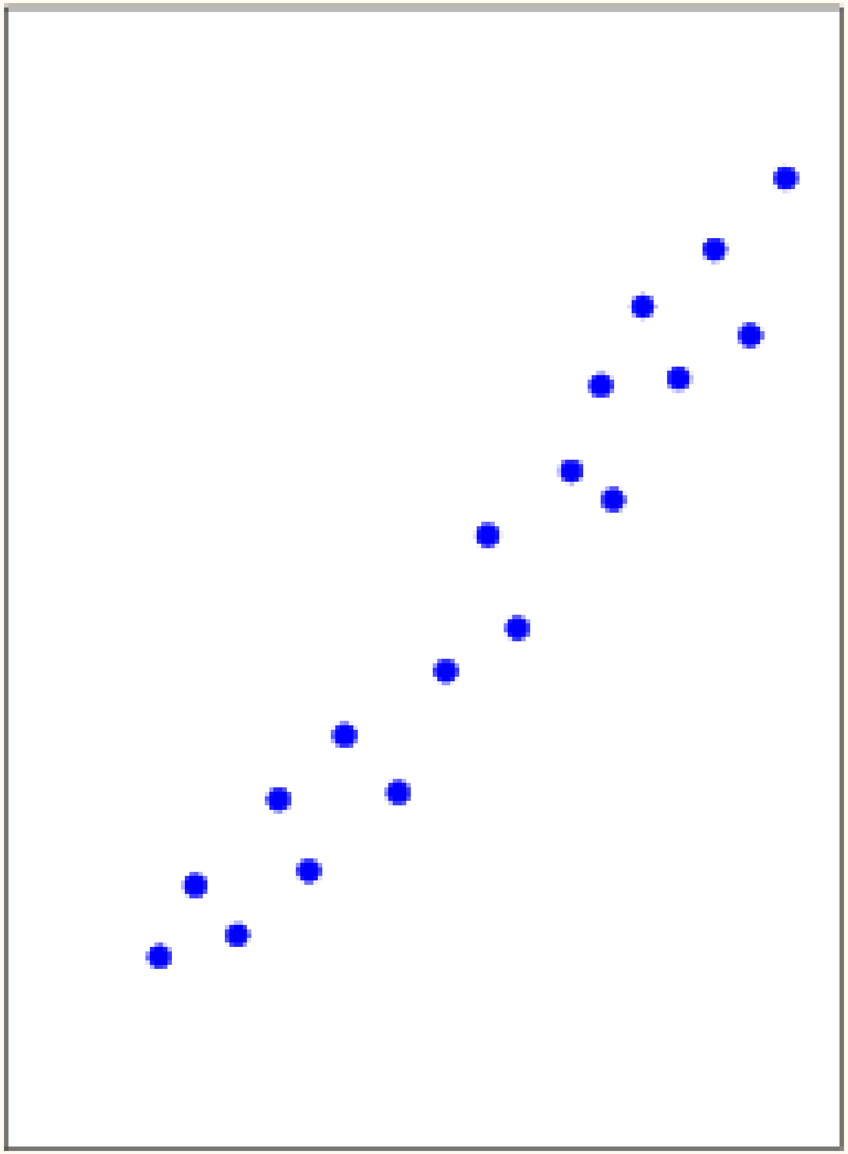
$$r = -0{,}8$$

<-1 ; 1>

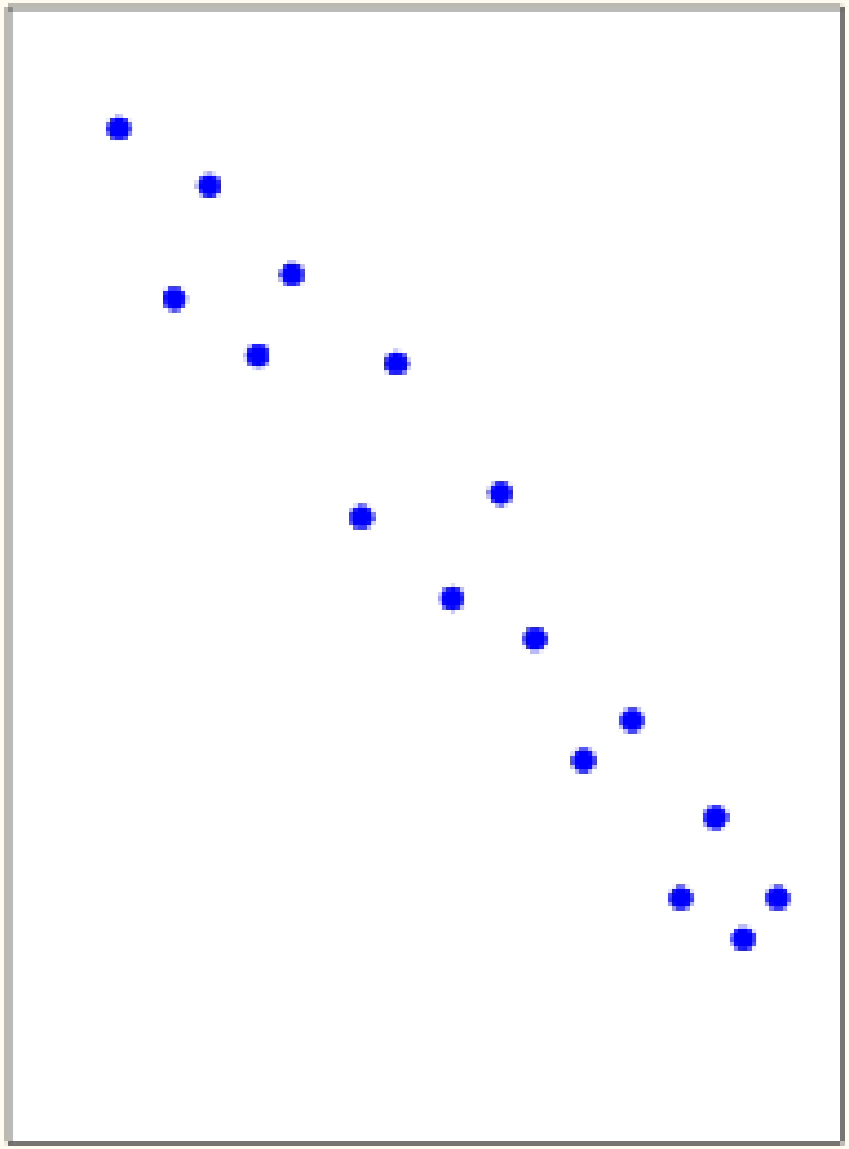
Co to znamená z hlediska interpretace?

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

Kladná korelace



Záporná korelace



To by ovšem znamenalo, že kdo je rychlejší v běhu na 100m, ten dosahuje horších 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…**

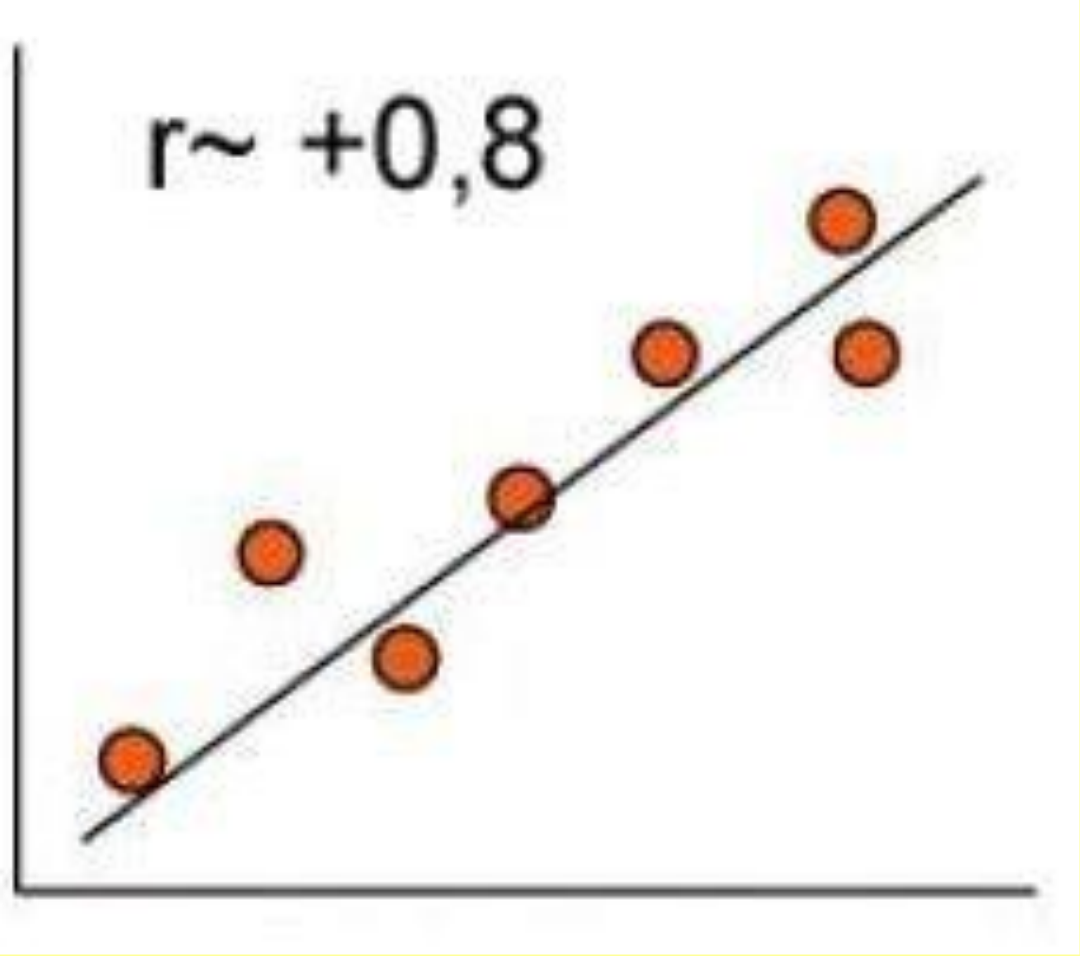
Koeficient determinace ***r^2 = 0, 64 (64 %).***

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	$X_i \cdot Y_i$
$\sum$	80	70	652	522	569

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

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

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

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

Konstrukce regresní přímky za pomocí regresní rovnice

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

Pozn. x...nezávisle proměnná y...závisle proměnná

Volba x ⇒

X	8	10

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

Výpočet y ⇒

Y	7	8,5

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

2) $X = a + b \, . \, Y$ 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

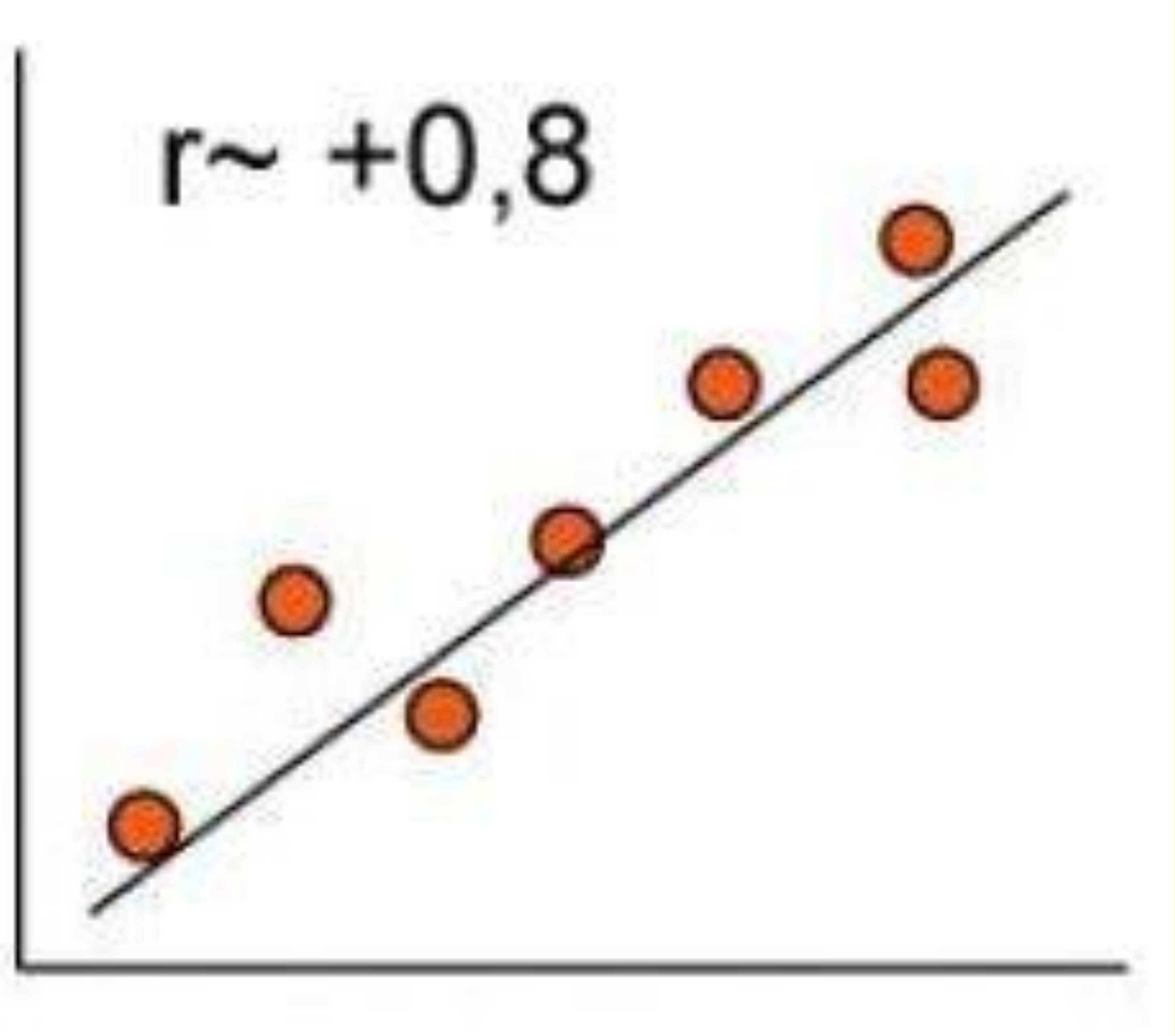
$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum y_i^2 - (\sum y_i)^2} \qquad a = \frac{\sum x_i - b . \sum y_i}{n}$$

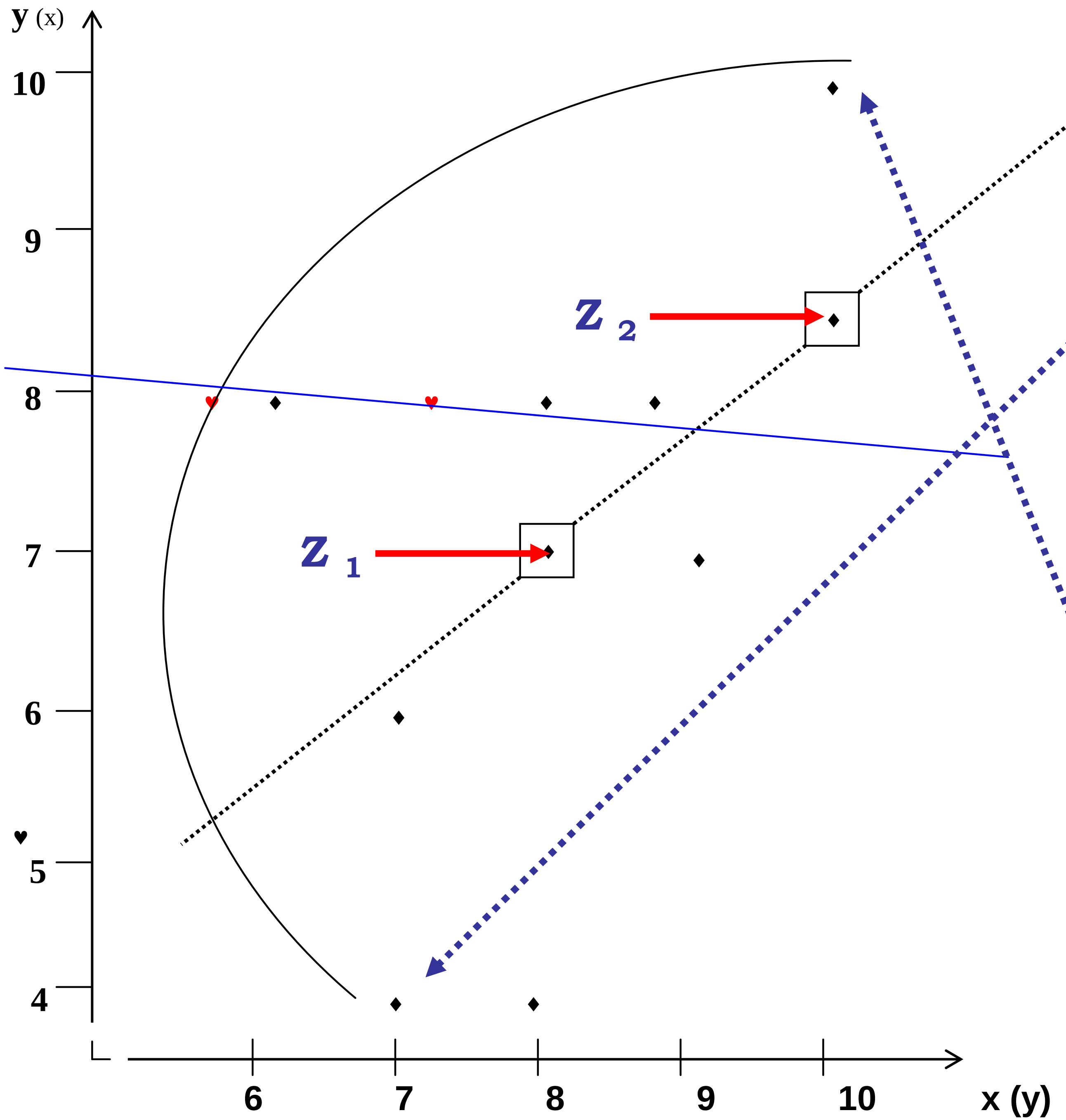
$$b = \frac{10.569 - 80.70}{10.522 - (70)^2} = \frac{5690 - 5600}{5220 - 4900} = \frac{90}{320} = 0,28$$

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

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

Graf korelační závislosti (= korelogram) - konstrukce



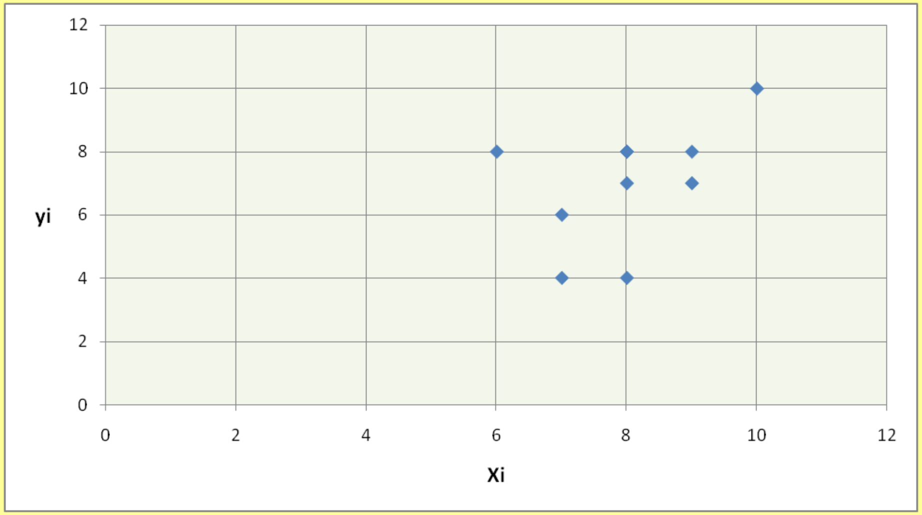


	x_i	y_i
A	7	4
B	6	8
C	7	6
D	8	8
E	9	7
F	8	8
G	8	7
H	8	4
J	9	8
K	10	10

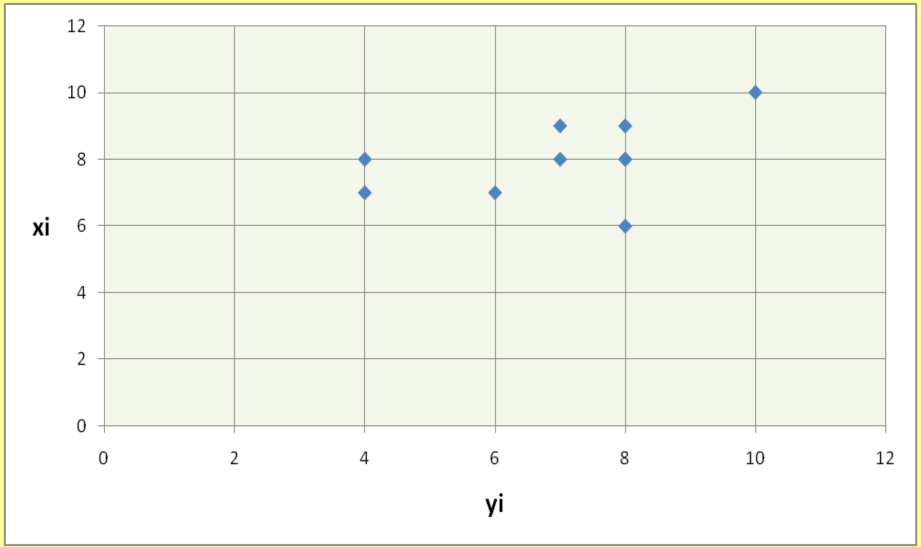
Pomocí Excelu – Statistické funkce
Výpočet koeficientů regresní přímky

	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				

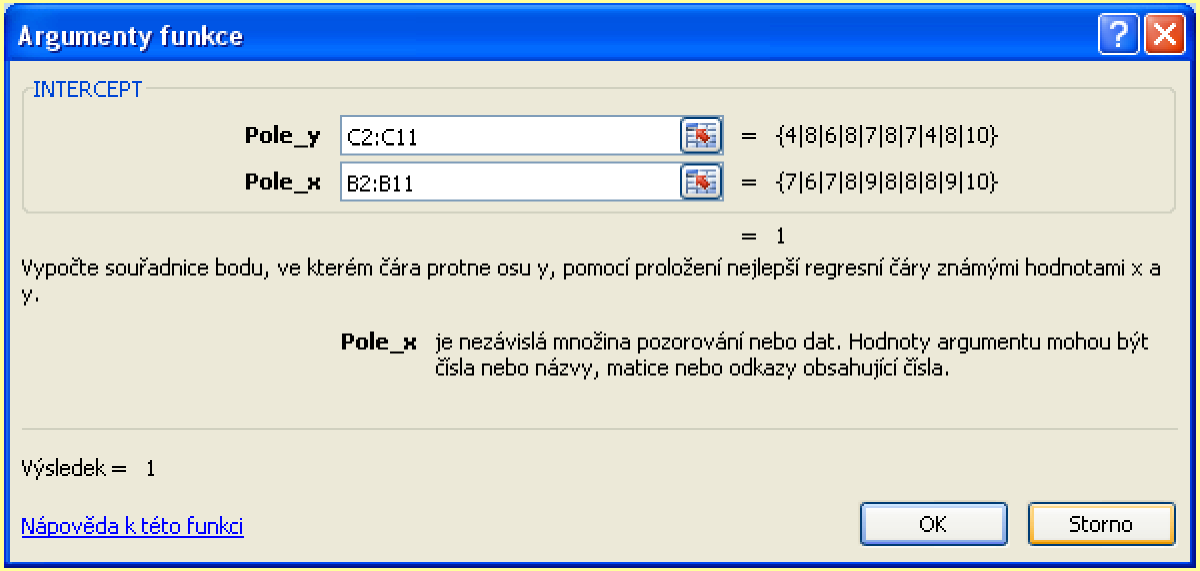
1) Y = a + b . X

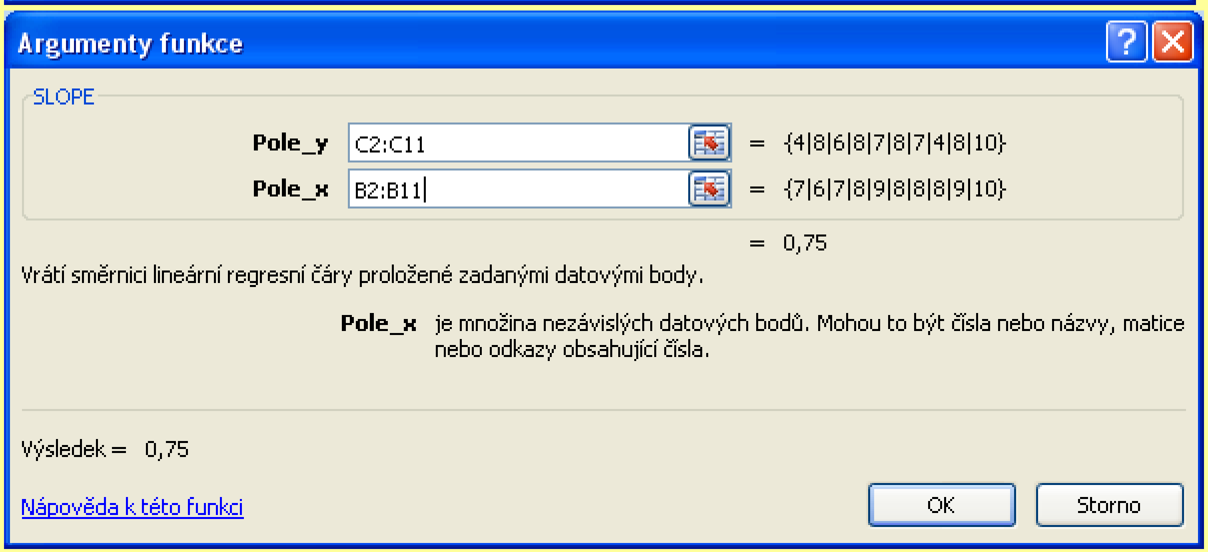


2) X = a + b . Y



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) $X = a + b \cdot Y$

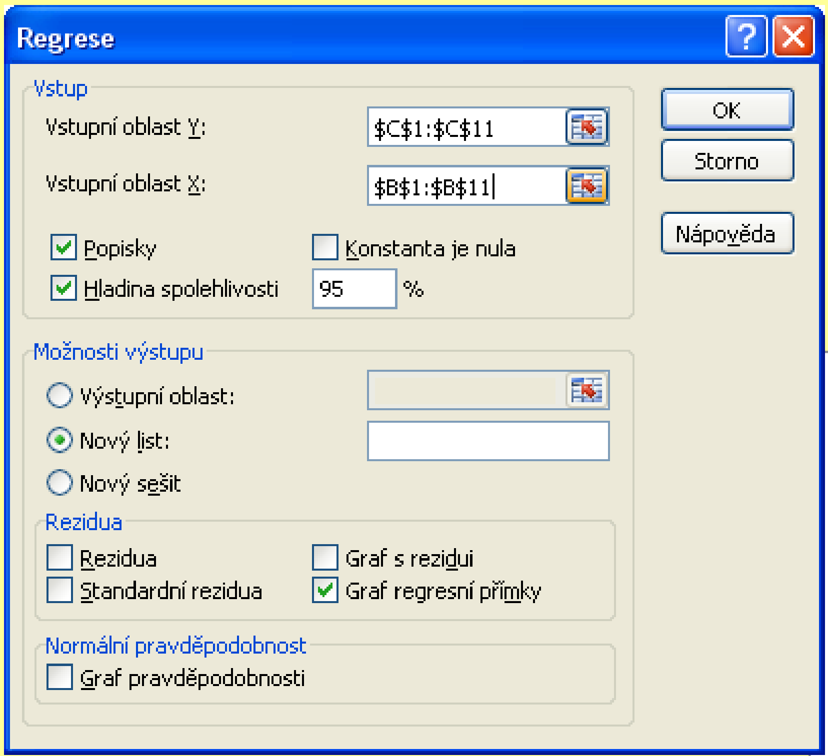
$X = 6 + 0{,}28 \cdot Y$

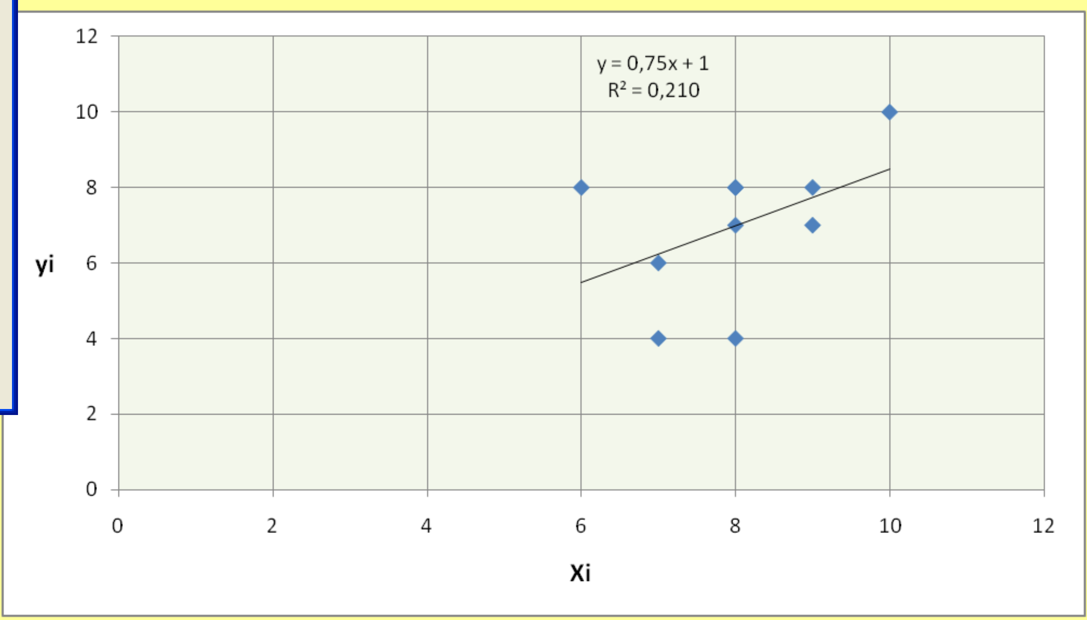
Pomocí Excelu – Analýza dat – Regrese
Výpočet koeficientů regresní přímky

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

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

	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

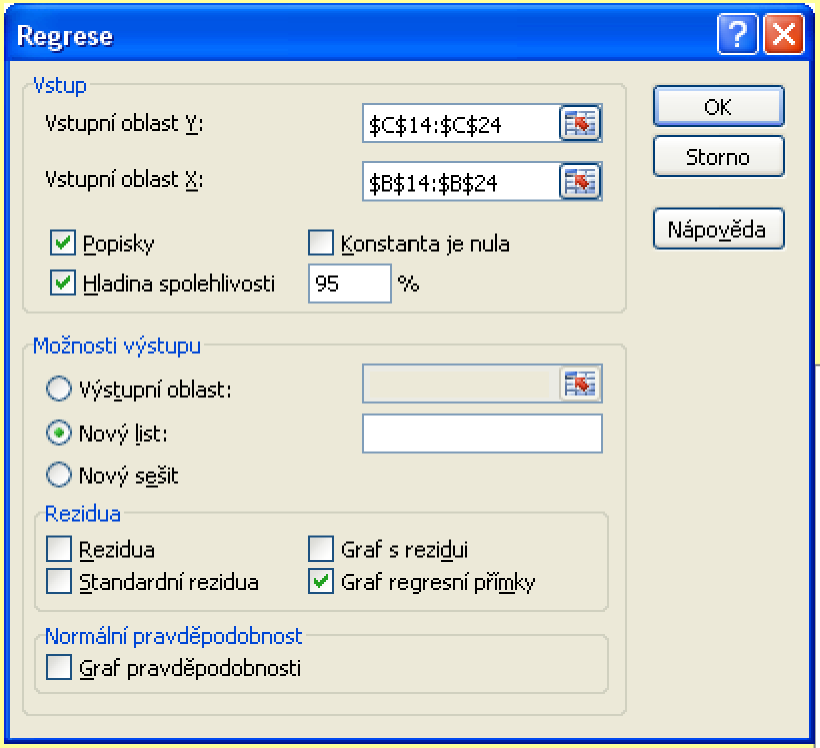
	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	6,75	6,75	2,138613861	0,181775314
Rezidua	8	25,25	3,15625		
Celkem	9	32			

		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

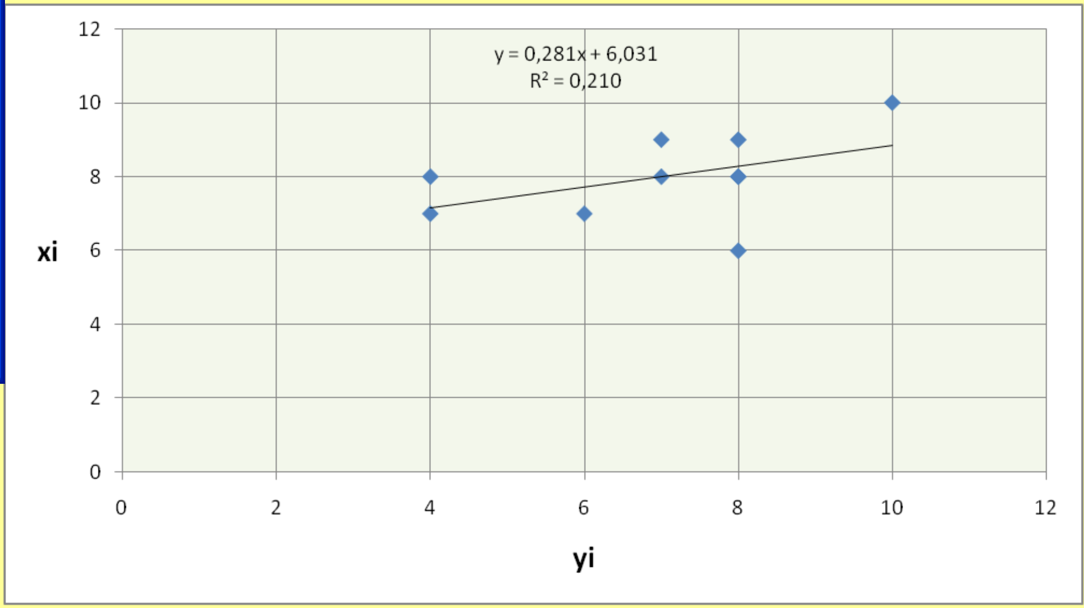
Pomocí Excelu – Analýza dat – Regrese
Výpočet koeficientů regresní přímky

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

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



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

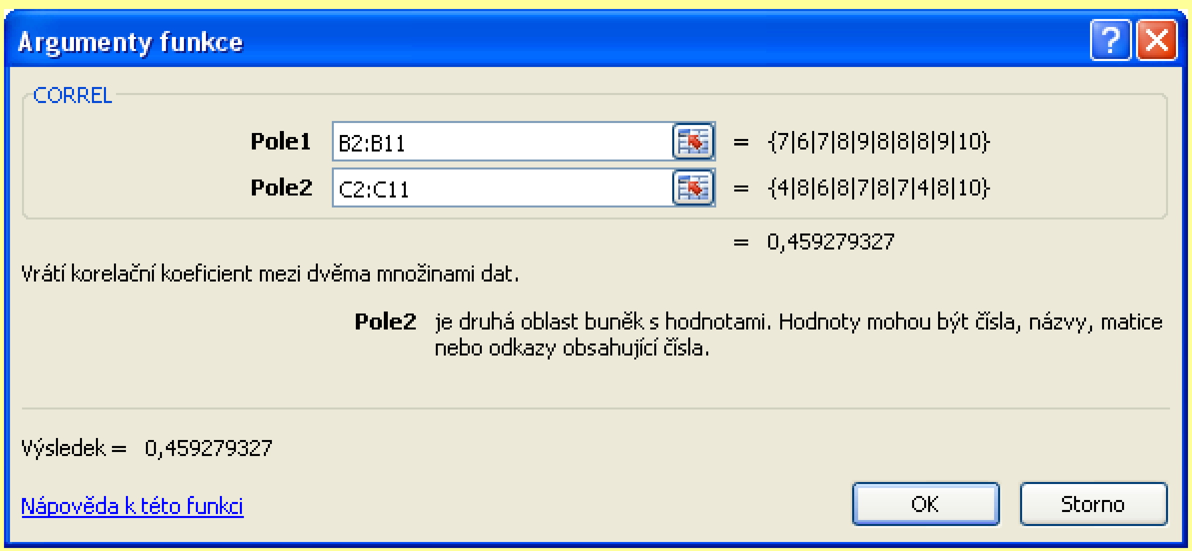
$$r_{x,y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]}}$$

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

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



r = 0,459

POSOUZENÍ A INTERPRETACE KORELAČNÍ ZÁVISLOSTI

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

Vzhledem k intervalu <0;1> resp. <-1;0> se jedná o ***střední míru závislosti (asociace).***

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 závislosti „r“ na základní soubor (populaci), musíme ověřit (testovat) hypotézu o statistické významnosti korelačního koeficientu.

POSOUZENÍ A INTERPRETACE KORELAČNÍ ZÁVISLOSTI

3. Při testování hypotézy a statistické významnosti „r“ (resp. 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)*

PRO NÁŠ PŘÍKLAD, kdy *r = 0,46; n = 10*

...zjistíme v tabulce kritických hodnot koeficientu součinové korelace, ...

Tabulka kritických hodnot

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

... že „náš" korelační koeficient r = 0,46 je pro obě hladiny významnosti menší, než tzv. kritická hodnota, je tedy

STATISTICKY NEVÝZNAMNÝ.

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

Ale pro n=30?

Příklad. Výpočet (Pearsonova) korelačního koeficientu

Testujte hypotézu, zda <u>výběrový korelační koeficient</u> je statisticky významný (s ohledem na rozsah souboru).

Test1	Test2
8	7
5	5
4	4
6	4
7	5
6	4
5	5
7	6

Tab. XVIII.5. Kritické hodnoty $r(\alpha)$ korelačního koeficientu r; $P\{|r| \geqq r(\alpha)\} = \alpha$

Rozsah výběru n	α 0,05	α 0,01
3	0,996 9	0,999 9
4	0,950 0	0,990 0
5	0,878 3	0,958 7
6	0,811 4	0,917 2
7	0,754 5	0,874 5
8	0,706 7	0,834 3
9	0,666 4	0,797 7
10	0,631 9	0,764 6
11	0,602 1	0,734 8
12	0,576 0	0,707 9
13	0,552 9	0,683 5
14	0,532 4	0,661 4
15	0,514 0	0,641 1
16	0,497 3	0,622 6
17	0,482 2	0,605 5
18	0,468 3	0,589 7
19	0,455 5	0,575 1
20	0,443 8	0,561 4
21	0,432 9	0,548 7
22	0,422 7	0,536 8
23	0,413 2	0,525 6
24	0,404 4	0,515 1
25	0,396 1	0,505 2

Příklad. Výpočet (Pearsonova) korelačního koeficientu

Pořadí osob	X_i	Y_i	X_i^2	Y_i^2	$X_i . Y_i$
$\sum$	48	40	300	208	247

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

$$r = \frac{8.\,247 - 48.40}{\sqrt{(8.\,300 - 2304).(8.\,208 - 1600)}} = \frac{56}{78} = 0{,}71$$

Rozsah výběru n	α 0,05	α 0,01
3	0,996 9	0,999 9
4	0,950 0	0,990 0
5	0,878 3	0,958 7
6	0,811 4	0,917 2
7	0,754 5	0,874 5
8	0,706 7	0,834 3
9	0,666 4	0,797 7
10	0,631 9	0,764 6

$r = 0{,}71 > 0{,}7067$ pro $\alpha = 0{,}05$

Na hladině $\alpha = 0{,}05$ zamítáme nulovou hypotézu. Koeficient je statisticky významný.

SPEARMANŮV KOEFICIENT POŘADOVÉ KORELACE

Spearmanův koeficient pořadové korelace se používá pro výpočet těsnosti závislosti:

- **u znaků získaných na ordinální stupnici (ordinálních znaků)**
- **u souborů o nevelkém rozsahu (n menší než 20)**
- **jestliže znaky nemají (či nelze prokázat) normální rozložení četností**

Vzorec pro výpočet Spearmanova koeficientu pořadové korelace:

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

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

$\boxed{r = 0{,}48}$

Pearsonův koeficient součinové korelace

$\boxed{r = 0{,}46}$

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; $P\{|r_S| > r_S(\alpha)\} \leqq \alpha$

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

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

Kritické hodnoty z tabulek $\alpha = 0{,}05$ 0,6364

$\alpha = 0{,}01$................ 0,7818

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

ANALÝZA JEDNOROZMĚRNÉHO SOUBORU

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. VÝPOČET ZÁKLADNÍCH STATISTICKÝCH CHARAKTERISTIK

a) míry centrální tendence

b) míry variability

c) míry závislosti

ZÁKLADY STATISTIKY

3. ANALYTICKÁ STATISTIKA

3.1 Základní soubor, výběrový soubor, náhodný výběr, závislé a nezávislé soubory

3.2 Hypotézy (stručné opakování)

3.3 Věcná a statistická významnost

3.4 Testování statistických hypotéz

(MATEMATICKÁ) STATISTIKA

DESKRIPTIVNÍ

(popisná)

ANALYTICKÁ

(inferentní, induktivní)

DESKRIPTIVNÍ STATISTIKA
(zpracováním a popis dat).

ANALYTICKÁ STATISTIKA
(analyzovat a vyhodnocení dat)

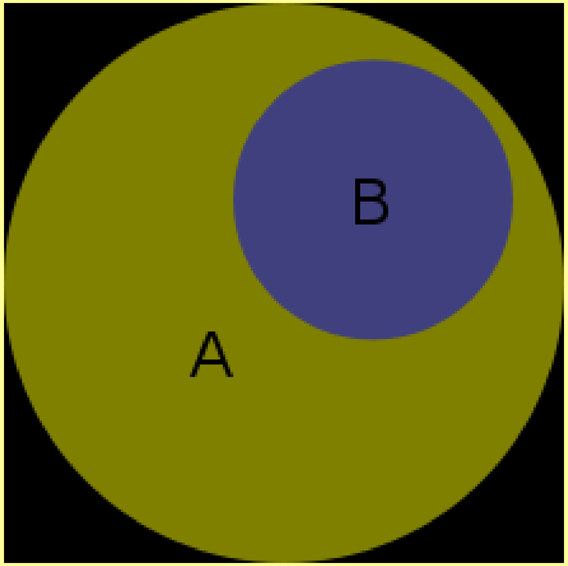
(1) stanovit, zda výsledky testů dvou tréninkových skupin vykazují významný rozdíl mezi středními hodnotami (vliv tréninkové metody),

(2) vyhodnotit léčebný účinek u 2 souborů pacientů.

3.1 ZÁKLADNÍ SOUBOR

(generální soubor, population, Grundgesamtkeit) **je soubor všech jedinců, u kterých bychom teoreticky měli šetření provádět.**

Základní soubor obvykle není dostupný, musíme se proto spokojit s omezeným počtem jedinců (objektů), takovýto soubor potom nazýváme **výběrovým souborem** (náhodný výběr, sample, Stichprobe).



VÝBĚROVÝ SOUBOR **je náhodnou podmnožinou prvků základního souboru, je získaný náhodným, resp. záměrným výběrem.**

Z poznatků zjištěných u výběrového souboru, můžeme (při splnění určitých statistických požadavků) činit závěry platné pro základní soubor.

ZÁVISLÉ SOUBORY

(test hod na koš, družstvo A 1., 2. pokusy)

NEZÁVISLÉ SOUBORY

(test hod na koš, družstvo A, družstvo B)

3.2 HYPOTÉZA

je podmíněný výrok o vztahu mezi dvěma nebo více proměnnými (Kerlinger, 1972).

Hypotézy jsou důležité a nepostradatelné prostředky vědeckého výzkumu, jsou pracovními nástroji teorie.

Kritéria dobrých hypotéz

1. hypotézy jsou *výroky o vztazích* mezi proměnnými

2. hypotézy obsahují *jasné implikace* (např. jestliže ..., pak ...) pro ověřování předpokládaných vztahů.

Hypotéza formuluje předpokládaný vztah mezi proměnnými, který se zamítá nebo nelze zamítnout.

Druhy hypotéz (Röthig, 1992)

1. ***Pracovní hypotéza*** **- subjektivní domněnky o předmětu výzkumného problému.**
   Pracovní hypotéza je formulována všeobecně, je základem pro realizaci předvýzkumu.

2. ***Výzkumná (věcná) hypotéza*** **– zdůvodněný předpoklad o existenci vztahu mezi dvěma či více proměnnými.**
   Zpřesněná formulace, ověřujeme testováním statistických hypotéz.

3. Statistická hypotéza **- hypotetické tvrzení vyjádřené ve statistických termínech o relacích, vyvozených z předpokládaných vztahů ve věcné H.**

H_0: $\mu = \mu_0$ H_A: $\mu \neq \mu_0$; H_A: $\mu > \mu_0$; H_A: $\mu < \mu_0$

Stupeň obecnosti ověřovaného tvrzení (hypotézy) klesá (od pracovní H –> ke statistické H).

Stupeň přesnosti ověřovaného tvrzení (hypotézy) vzrůstá (od pracovní H –> ke statistické H).

Hypotéza je testována pomocí tzv. testovacích metod (testů) a zamítá se, je-li zjištěn výsledek, který je při platnosti nulové hypotézy nepravděpodobný.

Co je považováno za nepravděpodobný výsledek, má být stanoveno předem (*např. tělesná výška mužů a žen je stejná*).

Výsledky testování hypotéz jsou posuzovány na tzv. hladině významnosti (p, α), která vyjadřuje pravděpodobnost chyby I. druhu (tedy chybné zamítnutí testované hypotézy).

Úroveň hladiny významnosti $\alpha = 0{,}05$ znamená, že nulová hypotéza se zamítá, když je pravděpodobnost, platnosti nulové hypotézy menší než 5% (obdobná interpretace platí pro $\alpha = 0{,}01$).

HYPOTÉZA NULOVÁ

Základním typem úvahy při statistickém testování tzv. *nulová hypotéza* (H_O). Př. Tělesná výška x věk

Podstatou *nulové hypotézy* je odůvodněný předpoklad, že mezi dvěma jevy <u>není statisticky významný rozdíl</u> (rozdíl je nulový, resp. malý).

Jako *nulová hypotéza* se označuje domněnka, že dva statistické soubory se shodují v určitých statistických parametrech (např. M, r).

H_0: $\mu = \mu_0$ H_A: $\mu \neq \mu_0$; H_A: $\mu > \mu_0$; H_A: $\mu < \mu_0$

HYPOTÉZA ALTERNATIVNÍ

Předpokládáme-li, že mezi dvěma jevy existuje významný rozdíl, formulujeme tzv. *alternativní hypotézu H_A* *(oboustranná, resp. jednostranná).*

K rozhodnutí, zda hypotézu (nulovou či alternativní) **zamítáme, či nezamítáme používáme tzv. testovací metody** *(viz dále).*

Co je považováno za výsledek pravděpodobný (TV M ≠ Ž, H_1), resp. nepravděpodobný (TV M = Ž, H_0) musí být tedy stanoveno předem.

H_0: $\mu = \mu_0$ H_A: $\mu \neq \mu_0$; H_A: $\mu > \mu_0$; H_A: $\mu < \mu_0$

3.3 VĚCNÁ A STATISTICKÁ VÝZNAMNOST

(1) STATISTICKÁ VÝZNAMNOST

Smysluplné použití **posuzování výsledků výzkumu** pomocí **statistické významnosti** je omezeno jen na soubory pořízené metodami **náhodného výběru,** resp. u **randomizovaných experimentů** (často nerespektováno).

Hlavní nevýhoda testování H pomocí statistické významnosti je její **vazba na rozsah souboru** (n):
- u **velkých výběrů** jsou i nepatrné rozdíly, resp. asociace (korelace) statisticky významné,
- u **malých výběrů** jsou i velké rozdíly či velká asociace (korelace) statisticky nevýznamné.

Tab. XVIII.5. Kritické hodnoty $r(\alpha)$ korelačního koeficientu r; $P\{|r| \geqq r(\alpha)\} = \alpha$

Rozsah výběru n	α 0,05	α 0,01
3	0,996 9	0,999 9
4	0,950 0	0,990 0
5	0,878 3	0,958 7
6	0,811 4	0,917 2
7	0,754 5	0,874 5
8	0,706 7	0,834 3
16	0,497 3	0,622 6
17	0,482 2	0,605 5
18	0,468 3	0,589 7
19	0,455 5	0,575 1
20	0,443 8	0,561 4
21	0,432 9	0,548 7
22	0,422 7	0,536 8
23	0,413 2	0,525 6
24	0,404 4	0,515 1
25	0,396 1	0,505 2

Výsledky testování hypotéz jsou posuzovány na tzv. hladině významnosti. Interpretace hladiny významnosti α = 0,05 znamená, že nulová hypotéza se zamítá s 5% pravděpodobností omylu.

VĚCNÁ A STATISTICKÁ VÝZNAMNOST

(2) VĚCNÁ VÝZNAMNOST

U **nenáhodných výběrů** se proto doporučuje posuzovat významnost **rozdílů** či **vztahů** pomocí tzv. **věcné významnosti** („size of effect“, „effect size“ neboli „velikost efektu“, např. pomocí ES indexů (Cohen, 1988).

Hlavní výhoda použití teorie věcné významnosti je malá závislost na rozsahu souboru (n).

http://www.socscistatistics.com/effectsize/Default3.aspx

https://www.statskingdom.com/index.html

https://stats.libretexts.org/Learning_Objects/02%3A_Interactive_Statistics

POSUZOVÁNÍ VĚCNÉ VÝZNAMNOSTI

(1) Cohen (1988, 1992). **Indexy velikosti efektu** (hodnoty pro malé, střední a velké efekty).

Test	**Effect size**		
	small	medium	large
d	.20	.50	.80
r	.10	.30	.50
Chi^2	.10	.30	.50

Vysvětlivky:
d = pro diference středních hodnot
R = pro korelace
Chi^2 = pro chí kvadrát

POSUZOVÁNÍ VĚCNÉ VÝZNAMNOSTI

(2) Soukup (2013). Effect size po úpravě do intervalů

Test	small	medium	large
d	0,2-0, 49	0,5-0,79	větší než 0,8
r	0,1-0,29	0,3-0,49	větší než 0,5
Chi2	0,1-0,29	0,3-0,49	větší než 0,5

3.3.1 TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ POMOCÍ VĚCNÉ VÝZNAMNOSTI

Postup při hodnocení výsledků výzkumu:

1. nejprve zhodnotit **věcnou významnost** jak absolutně (v jednotkách měření), tak i relativně k podílu vlivu ostatních faktorů.

Pouze a jen, jde-li o randomizovaný výzkum, pak

2. použít výpočet **statistické významnosti,** jakožto kritérium pro posouzení rizika zobecnění.

Formulace: nulová hypotéza (H_0)

Příklad 1

H_{01}: *intersexuální rozdíly* *somatických a motorických předpokladů mezi tenisty (n=221) a tenistkami (n=193) ve věkové kategorii* ***11 -12 let jsou nevýznamné*.**

Soubor/SC H	Tenisté		Tenistky		Cohen´s d, hodnocení efektu
	M	SD	M	SD	
Výška (cm)	155,10	7,62	154,60	6,94	0,07 (žádný)
Hmotnost (kg)	43,50	6,68	43,49	7,17	0,00 (žádný)
MS (kp)	25,14	4,60	23,08	4,61	0,45 (malý)
RS	0,58	0,09	0,53	0,09	0,56 (střední)

Formulace: alternativní hypotéza (H_A, H_1)

Příklad 2

H_{A1}: *intersexuální rozdíly* *somatických a motorických předpokladů mezi tenisty (n=157) a tenistkami (n=163) ve věkové kategorii* ***13 -14 let jsou významné***.

Category	M (male)	SD	M (female)	SD	Cohen´s d
Height (cm)	169.79	9.27	164.93	5.80	0.63 (med)
Weight (kg)	57.05	9.26	53.57	6.31	0.44 (small)
MHSL (kp)	34.64	7.53	29.09	3.84	0.94 (large)
RHSL	0.61	0.10	0.55	0.06	0.73 (med)

VĚCNÁ VÝZNAMNOST – LITERATURA

Blahuš, P. (2000). Statistická významnost proti vědecké průkaznosti výsledků výzkumu. *Česká kinantropologie, 4*(2), 53-72.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 1*(112), 155-159. doi:10.1037/0033-2909.112.1.155

Soukup (2013). Věcná významnost výsledků a její možnosti měření. *Data a výzkum - SDA Info, 7*(2), 125-148.
http://dx.doi.org/10.13060/23362391.2013.127.2.41

Soukup, P. (2010). Nesprávná užívání statistické významnosti a jejich možná řešení. *Data a výzkum - SDA Info, 4*(2), 77-104.
http://dav.soc.cas.cz/uploads/27e65d18f9df9bee6df1af9649f82b267f9cccda_DaV10_2_s77_104.pdf

3.3.2 TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ POMOCÍ STATISTICKÉ VÝZNAMNOSTI

Výsledky testování (statistická významnost) jsou posuzovány na tzv. hladině významnosti.

Úroveň hladiny významnosti α=0,05 znamená, že nulová hypotéza se zamítá, když je pravděpodobnost, že nastane nulová hypotéza, menší než 5% (α = 0,01).

V tomto případě se přikláníme k platnosti *alternativní hypotézy*.

Nejčastěji srovnáváme střední hodnoty dvou výběrových souborů (rozsahu n_1, n_2), resp. závislosti.

1. NOMINÁLNÍ DATA - STATISTICKÉ TESTOVACÍ METODY

1. Lyžaři

2. Lyžaři

Znak - kouření

PŘEDPOKLAD	PROBLÉM	TESTOVACÍ METODA
Dva nezávislé soubory (znaky nabývají právě dvou hodnot)	Zkouška významnosti rozdílů souborů	X^2-čtyřpolní test (Fischerův test, čtyřpolní tabulka)
Dva nezávislé soubory (znaky nabývají více hodnot)	Zkouška významnosti rozdílů souborů	X^2-vícepolní test (kontingenční tabulka)
Dva závislé soubory (znaky nabývají právě dvou hodnot)	Zkouška významnosti změn	X^2-Mc Nemarův test
Dva závislé soubory	Hodnocení závislosti	Koef. kontingence C

2. ORDINÁLNÍ DATA - STATISTICKÉ TESTOVACÍ METODY

1. Gymnasté A

2. Gymnasté B

Znak - body

PŘEDPOKLAD	PROBLÉM	TESTOVACÍ METODA
Dva nezávislé soubory	Test rovnosti centrálních tendencí	Medianový test (jednoduchý), U-test Mann-Whitneyho, Kolmogorov-Smirnovův test, Marshallův test
Dva závislé soubory	Test rovnosti centrálních tendencí	Znaménkový test, Wilcoxonův test
Více nezávislých souborů	Test rovnosti centrálních tendencí	Medianový test (rozšířený), H-test Kruskal-Wallisův (analýza rozptylu)
Dva závislé soubory	Hodnocení míry závislosti	Spearmanův resp. Kendallův koeficient korelace
Více závislých souborů	Hodnocení míry závislosti	Friedmanova analýza rozptylu

3. METRICKÁ DATA - STATISTICKÉ TESTOVACÍ METODY I

Tenisté

Tenistky

Znak: TV

PŘEDPOKLAD	PROBLÉM	TESTOVACÍ METODA
Dva nezávislé soubory	Zkouška rovnosti rozptylů (homogenita)	F-test
Dva nezávislé soubory	Zkouška rovnosti středních hodnot	t-test
Dva nezávislé soubory	Zkouška nezávislosti korelací	Korelační test
Dva závislé soubory	Zkouška rovnosti rozptylů (homogenita)	F-test

3. METRICKÁ DATA - STATISTICKÉ TESTOVACÍ METODY II

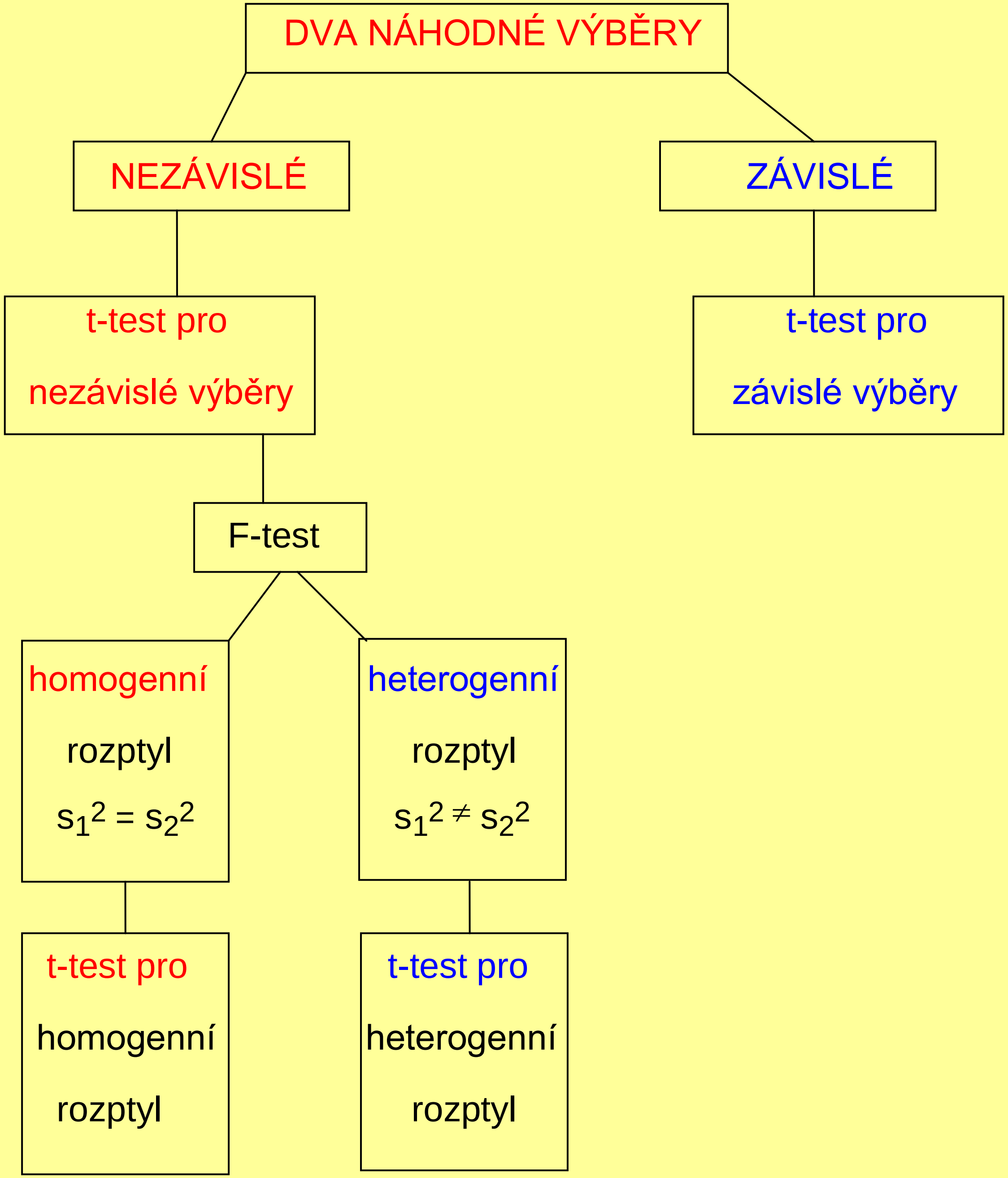
Tenisté

Tenistky

Znak: TV

PŘEDPOKLAD	PROBLÉM	TESTOVACÍ METODA
Dva závislé soubory	Zkouška rovnosti středních hodnot	Diferenční t-test (párový)
Dva závislé soubory	Hodnocení závislosti	Koef. součinové korelace a regrese
Více nezávislých souborů	Zkouška rovnosti průměrů	Analýza rozptylu, Duncanův test pořadí, Bartlettův test
Více nezávislých souborů	Zkouška rovnosti korelačních koeficientů	Test homogenity

ROZHODOVACÍ DIAGRAM PRO UŽITÍ t-TESTU



STATISTICKÉ TESTOVACÍ METODY

Párový t - test

- dva závislé soubory
- zkouška rovnosti středních hodnot

PŘÍKLAD – Zjistěte, zda se na automobilu určité značky sjíždějí obě přední pneumatiky stejně rychle

číslo automobilu	1	2	3	4	5	6
pravá pneumatika	1,8	1	2,2	0,9	1,5	1,6
leva pneumatika	1,5	1,1	2	1,1	1,4	1,4
rozdíl	0,3	-0,1	0,2	-0,2	0,1	0,2

$H_0 : \mu = \mu_1 - \mu_2 = 0$ $\quad$ $H_A : \mu = \mu_1 - \mu_2 \neq 0$

$$T = \frac{\left|\bar{X} - \mu\right|}{s}\sqrt{n} \qquad T < t_{n-1;1-\frac{\alpha}{2}} \quad \Rightarrow$$

hypotézu nelze zamítnou

STATISTICKÉ TESTOVACÍ METODY

Párový t - test

číslo automobilu	1	2	3	4	5	6
pravá pneumatika	1,8	1	2,2	0,9	1,5	1,6
leva pneumatika	1,5	1,1	2	1,1	1,4	1,4
rozdíl	0,3	-0,1	0,2	-0,2	0,1	0,2

$$\overline{X} = \frac{1}{n}\sum_{n=1}^{n} X_i = \frac{1}{6}(0{,}3 - 0{,}1 + 0{,}2 - 0{,}2 + 0{,}1 + 0{,}2) = \frac{0{,}5}{6} = \underline{\underline{0{,}0833}}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{0{,}2167^2 + (-0{,}1833)^2 + 0{,}1167^2 + (-0{,}2833)^2 + 0{,}0167^2 + 0{,}1167^2}{5} =$$

$$= \frac{0{,}18833}{5} = 0{,}0377$$

$$s = \sqrt{s^2} = \sqrt{0{,}0377} = \underline{\underline{0{,}1941}}$$

STATISTICKÉ TESTOVACÍ METODY

Párový t - test

$$t_{n-1;1-\frac{\alpha}{2}} = t_{6-1;1-\frac{0,05}{2}} = t_{5;0,975} = 2,571$$

= > z tabulek

$$T = \frac{\left|\overline{X} - \mu\right|}{s}\sqrt{n} = \frac{0,0833-0}{0,1941}\sqrt{6} = 1,0518 < 2,571$$

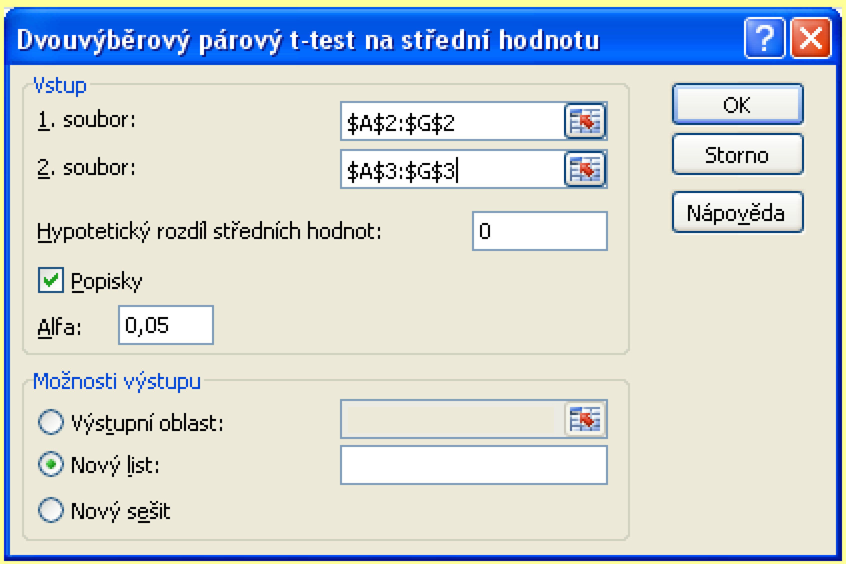
Protože 1,0518 < 2,571, nelze na základě získaných dat zamítnout hypotézu, že se obě přední pneumatiky sjíždějí stejně rychle.

STATISTICKÉ TESTOVACÍ METODY

Párový t - test

Pomocí Excelu – Analýza dat – Dvouvýběrový párový t-test na střední hodnotu

	A	B	C	D	E	F	G	H
1	číslo automobilu	1	2	3	4	5	6	
2	pravá pneumatika	1,8	1	2,2	0,9	1,5	1,6	
3	leva pneumatika	1,5	1,1	2	1,1	1,4	1,4	
4	rozdíl	0,3	-0,1	0,2	-0,2	0,1	0,2	
5								



Dvouvýběrový párový t-test na střední hodnotu

	pravá pneumatika	*leva pneumatika*
Stř. hodnota	1,5	1,416666667
Rozptyl	0,24	0,109666667
Pozorování	6	6
Pears. korelace	0,961571662	
Hyp. rozdíl stř. hodnot	0	
Rozdíl	5	
t Stat	1,051757905	
P(T<=t) (1)	0,17053101	
t krit (1)	2,015048372	
P(T<=t) (2)	0,34106202	
t krit (2)	2,570581835	

STATISTICKÉ TESTOVACÍ METODY

Dvouvýběrový t - test

- dva nezávislé soubory
- test rovnosti středních hodnot

PŘÍKLAD – U studentů rozdělených do dvou skupin byl zaznamenán počet leh-sedů za 1 minutu. Jsou obě skupiny stejně výkonné?

1. skupina	62	54	55	60	53	58
2. skupina	52	56	49	50	51	

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)s_X^2 + (m-1)s_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}$$

$H_0: \mu_1 = \mu_2$

$H_A: \mu_1 \neq \mu_2$

$|T| < t_{n+m-2;1-\frac{\alpha}{2}} \Rightarrow$ hypotézu nelze zamítnou

STATISTICKÉ TESTOVACÍ METODY

Dvouvýběrový t - test

1. skupina	62	54	55	60	53	58
2. skupina	52	56	49	50	51	

n1=6 n2=5 AP$_X$=57 AP$_Y$=51,6 s$_X^2$ =12,8 s$_Y^2$ =7,3

$$T=\frac{\overline{X}-\overline{Y}}{\sqrt{(n-1)s_X^2+(m-1)s_Y^2}}\sqrt{\frac{nm(n+m-2)}{n+m}}=$$

$$=\frac{57-51{,}6}{\sqrt{(6-1)12{,}8+(5-1)7{,}3}}\sqrt{\frac{6.5.(6+5-2)}{6+5}}=$$

$$=\frac{5{,}4}{\sqrt{62{,}5+29{,}2}}\sqrt{24{,}55}=2{,}79$$

STATISTICKÉ TESTOVACÍ METODY

Dvouvýběrový t -test

$$t_{m+m-2;1-\frac{\alpha}{2}} = t_{6+5-2;1-\frac{0,05}{2}} = t_{9;0,975} = 2,262$$ **= > z tabulek**

$$|T| = 2,79 \geq 2,262$$

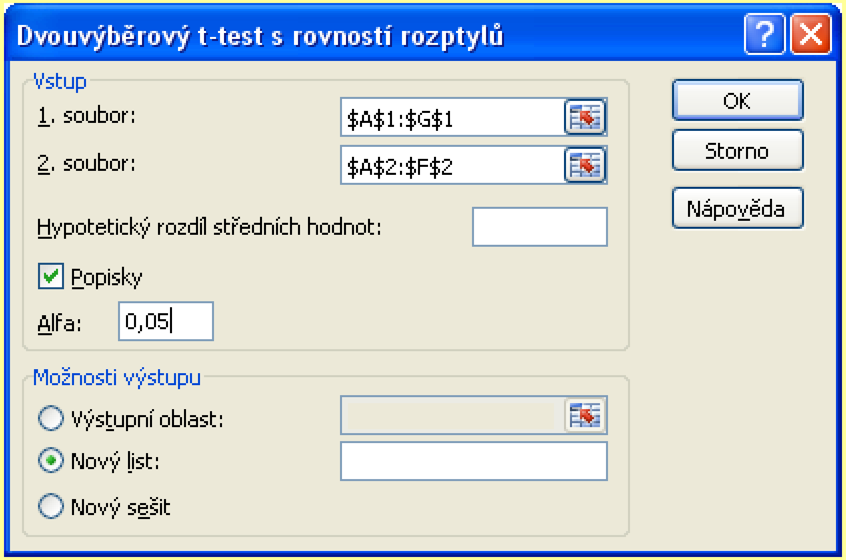
Protože 2,79 ≥ 2,262 zamítáme hypotézu, že se obě skupiny studentů jsou stejně výkonné.

STATISTICKÉ TESTOVACÍ METODY

Dvouvýběrový t - test

Pomocí Excelu – Analýza dat – Dvouvýběrový t-test s rovností rozptylů

	A	B	C	D	E	F	G	H
1	1. skupina	62	54	55	60	53	58	
2	2. skupina	52	56	49	50	51		



Dvouvýběrový t-test s rovností rozptylů

	1. skupina	2. skupina
Stř. hodnota	57	51,6
Rozptyl	12,8	7,3
Pozorování	6	5
Společný rozptyl	10,35555556	
Hyp. rozdíl stř. hodnot	0	
Rozdíl	9	
t Stat	2,77122216	
P(T<=t) (1)	0,010855041	
t krit (1)	1,833112923	
P(T<=t) (2)	0,021710083	
t krit (2)	2,262157158	

STATISTICKÉ TESTOVACÍ METODY

F - test

- dva nezávislé soubory
- zkouška rovnosti rozptylů

PŘÍKLAD – **Na základě dat uvedených v předchozím příkladě rozhodněte, zda oba základní soubory mají stejné rozptyly.**

1. skupina	62	54	55	60	53	58
2. skupina	52	56	49	50	51	

$$Z = \frac{s_X^2}{s_Y^2} \qquad volím \quad tak, aby \quad Z > 1$$

$$Z < F_{n-1, m-1; 1-\frac{\alpha}{2}} \qquad \Rightarrow$$ hypotézu nelze zamítnou

$H_0: \quad \sigma_X^2 = \sigma_Y^2$

$H_A: \quad \sigma_X^2 \neq \sigma_Y^2$

STATISTICKÉ TESTOVACÍ METODY

F - test

1. skupina	62	54	55	60	53	58
2. skupina	52	56	49	50	51	

n=6 m=5 s_X^2 =12,8 s_Y^2 =7,3

$$Z = \frac{s_X^2}{s_Y^2} = \frac{12,8}{7,3} = 1,753$$

$$F_{n-1,m-1;1-\frac{\alpha}{2}} = F_{6-1,5-1;1-\frac{0,05}{2}} = F_{5,4;0,975} = 9,36$$ **= > z tabulek**

$$Z = 1,753 < 9,36$$

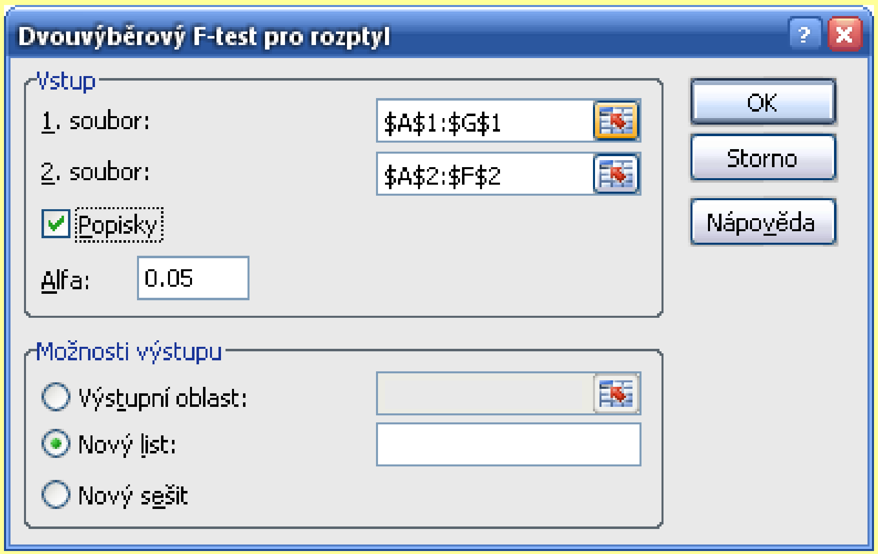
Protože 1,753 < 9,36 nelze zamítnout hypotézu o shodnosti rozptylů.

STATISTICKÉ TESTOVACÍ METODY

F - test

Pomocí Excelu – Analýza dat – Dvouvýběrový F-test pro rozptyl

	A	B	C	D	E	F	G	H
1	1. skupina	62	54	55	60	53	58	
2	2. skupina	52	56	49	50	51		



Dvouvýběrový F-test pro rozptyl

	1. skupina	*2. skupina*
Stř. hodnota	57	51.6
Rozptyl	12.8	7.3
Pozorování	6	5
Rozdíl	5	4
F	1.753424658	
P(F<=f) (1)	0.303172533	
F krit (1)	6.256056502	

Děkuji za pozornost