

4. Zpracování číselných dat

4.1 Jednoduché hodnocení dat

4.2 Začlenění dat do písemné práce

Zásady zpracování vědecké práce pro obory BOZO, PÚPN,
LS 2011

4.1 Hodnocení číselných dat

Popisná data: střední hodnota

Průměr

Příklad:

- Průměrná výška patnáctiletých hochů v okrese Jindřichův Hradec v r. 2010
- Průměrný výnos sena v 1 .seči na studovaném lučním komplexu
- Průměrná spotřeba automobilu určité značky na 100 km

Popisná data: rozptyl

- Rozptyl = variabilita hodnot
tj. jejich rozložení kolem střední hodnoty
- Směrodatná odchylna = míra rozptylu
vypovídá o tom, jak moc se od sebe navzájem liší
typické případy v souboru zkoumaných čísel.
- Variační koeficient
Podíl směrodatné odchylny a průměru vyjádřený v
procetech

Základní soubor a náhodný výběr

- Základní soubor: obsahuje všechny případy dané kategorie. Jen výjimečně je mohu změřit všechny (příklad: patnáctiletí chlapci v ČR v r. 2010)
- Náhodný výběr: obsahuje pouze měřené případy z rozsáhlejšího souboru
- Statistické soubory většinou představují náhodné výběry

Charakteristiky základního souboru a náhodného výběru

- Základní soubor i náhodný výběr mají střední hodnotu (průměr) i indikaci rozptylu (směrodatnou odchylku).
- Vypočítat (pomocí vzorců) můžeme jen hodnoty pro náhodný výběr.
- Z charakteristik náhodného výběru můžeme usuzovat na charakteristiky základního výběru (s určitou pravděpodobností).

Střední chyba průměru

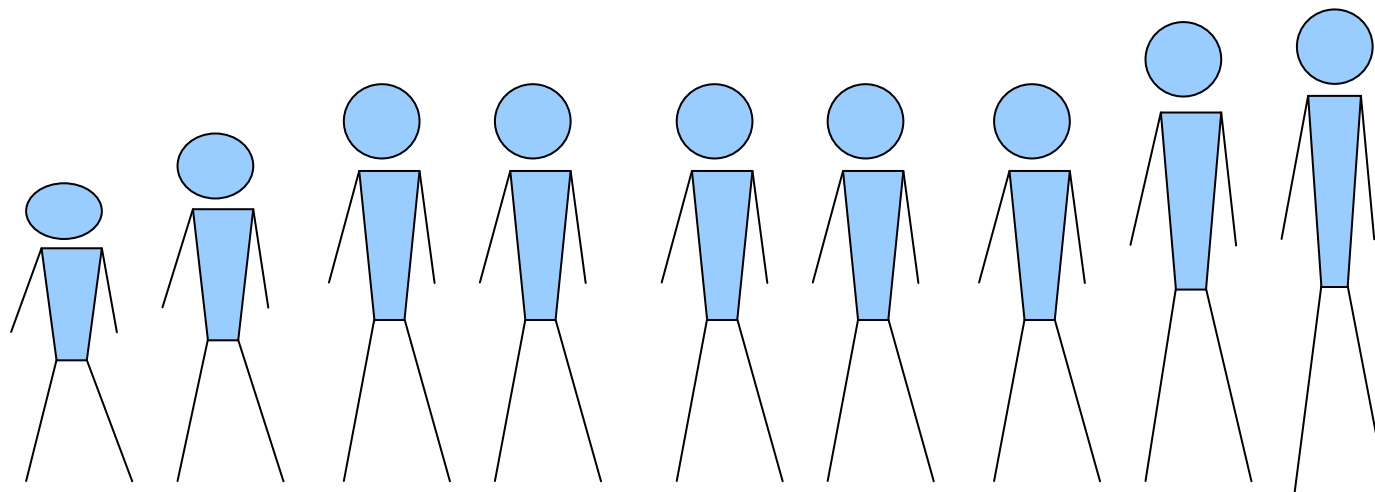
Spočtený průměr náhodného výběru se používá jako bodový odhad průměru celého základního souboru. Musíme ale počítat s tím, že tento odhad se od skutečného průměru základního souboru více nebo méně odchyluje. Přesnost odhadu můžeme vyjádřit tak, že výsledek doplníme o velikost možné odchylky. Běžně používanou mírou je střední (směrodatná) chyba průměru.

- Střední chyba průměru udává chybu odhadu průměru základního souboru.
- Konfidenční interval udává meze, v nichž s určitou pravděpodobností (95%) leží průměr základního souboru.

Podmínka využití

- Průměr, směrodatná odchylka a střední chyba průměru vypovídají o datech dobře (smysluplně) jen tehdy, mají-li data normální rozdělení
- Normalitu rozdělení hodnot v datových souborech lze testovat

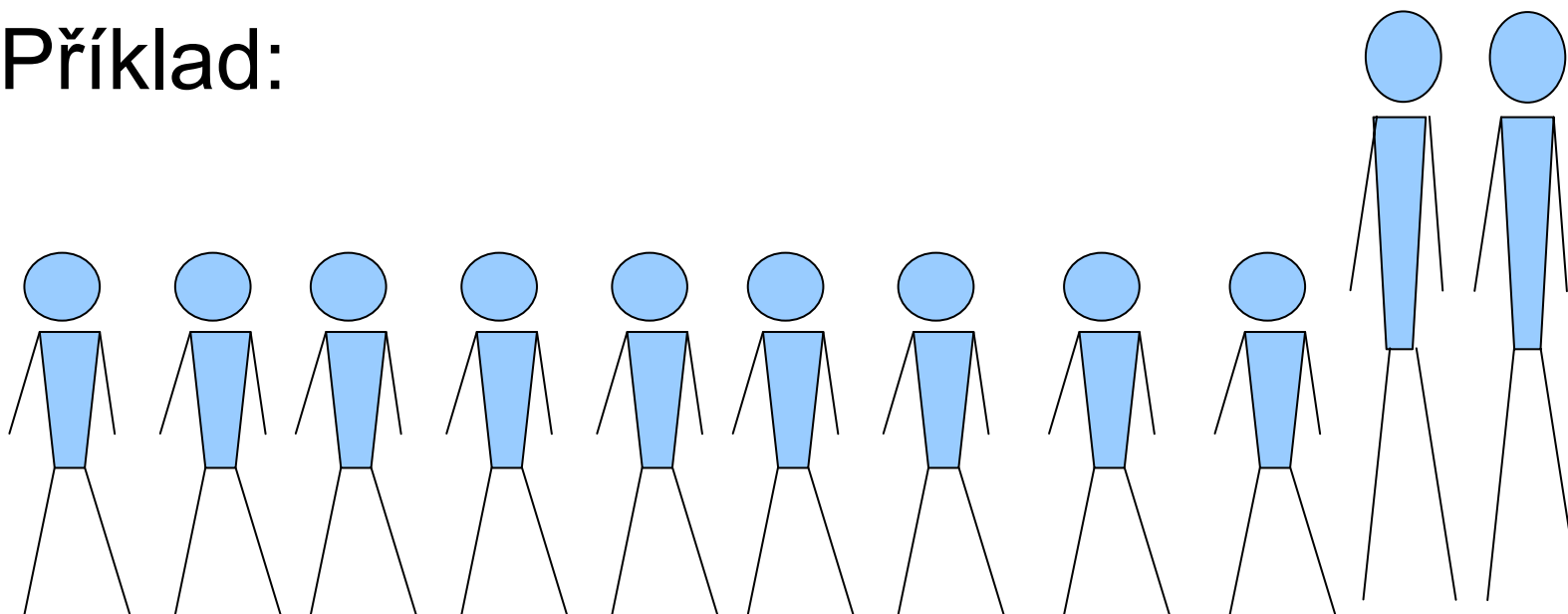
Normální rozdělení hodnot



Nejvíce hodnot leží blízko průměru, hodnot vzdálených od průměru je málo. Počet hodnot menších než průměr je podobný jako počet hodnot větších než průměr

Jiná rozdělení hodnot

Příklad:



Malých hodnot je výrazně více než velkých.

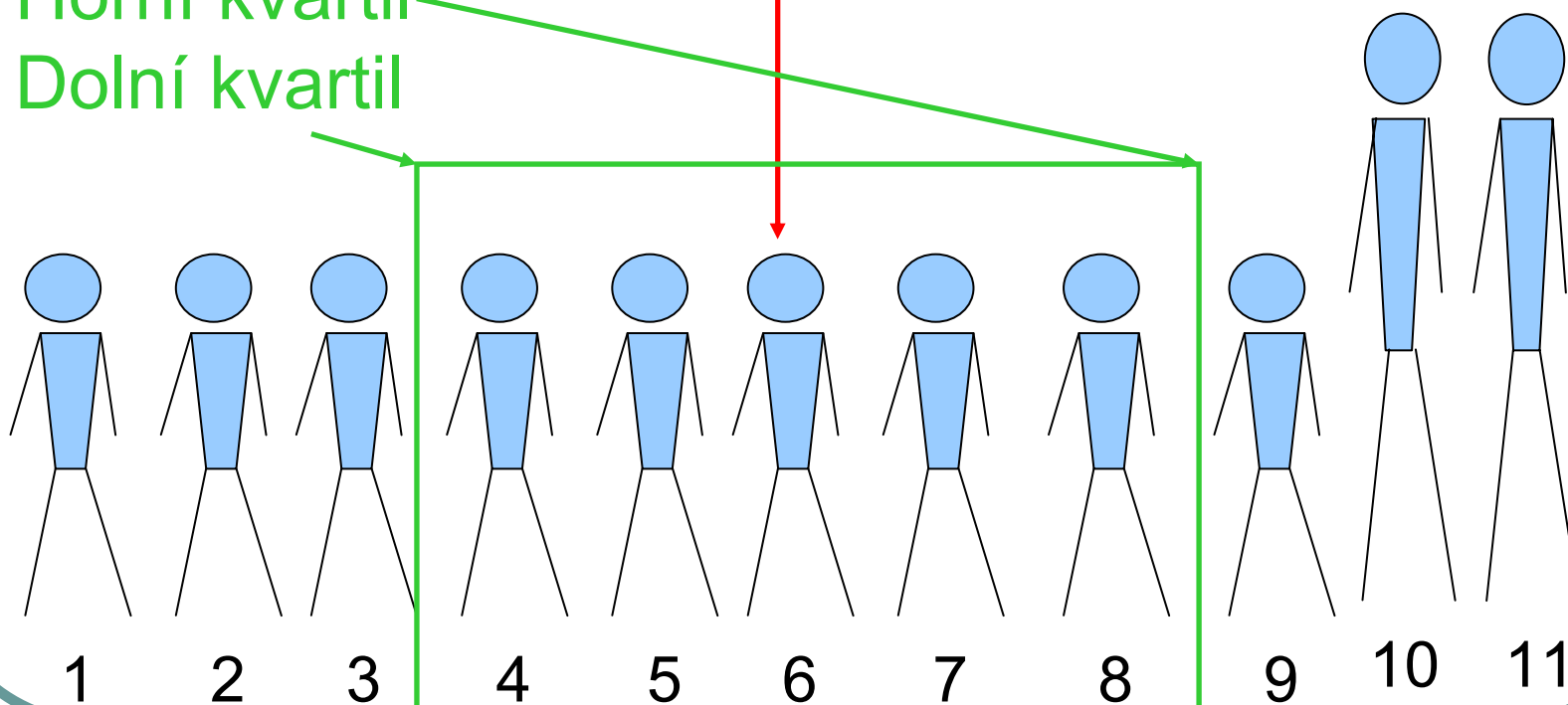
Popis souborů, které nemají normální rozdělení

Střední hodnota: medián

Rozptyl hodnot:

Horní kvartil

Dolní kvartil



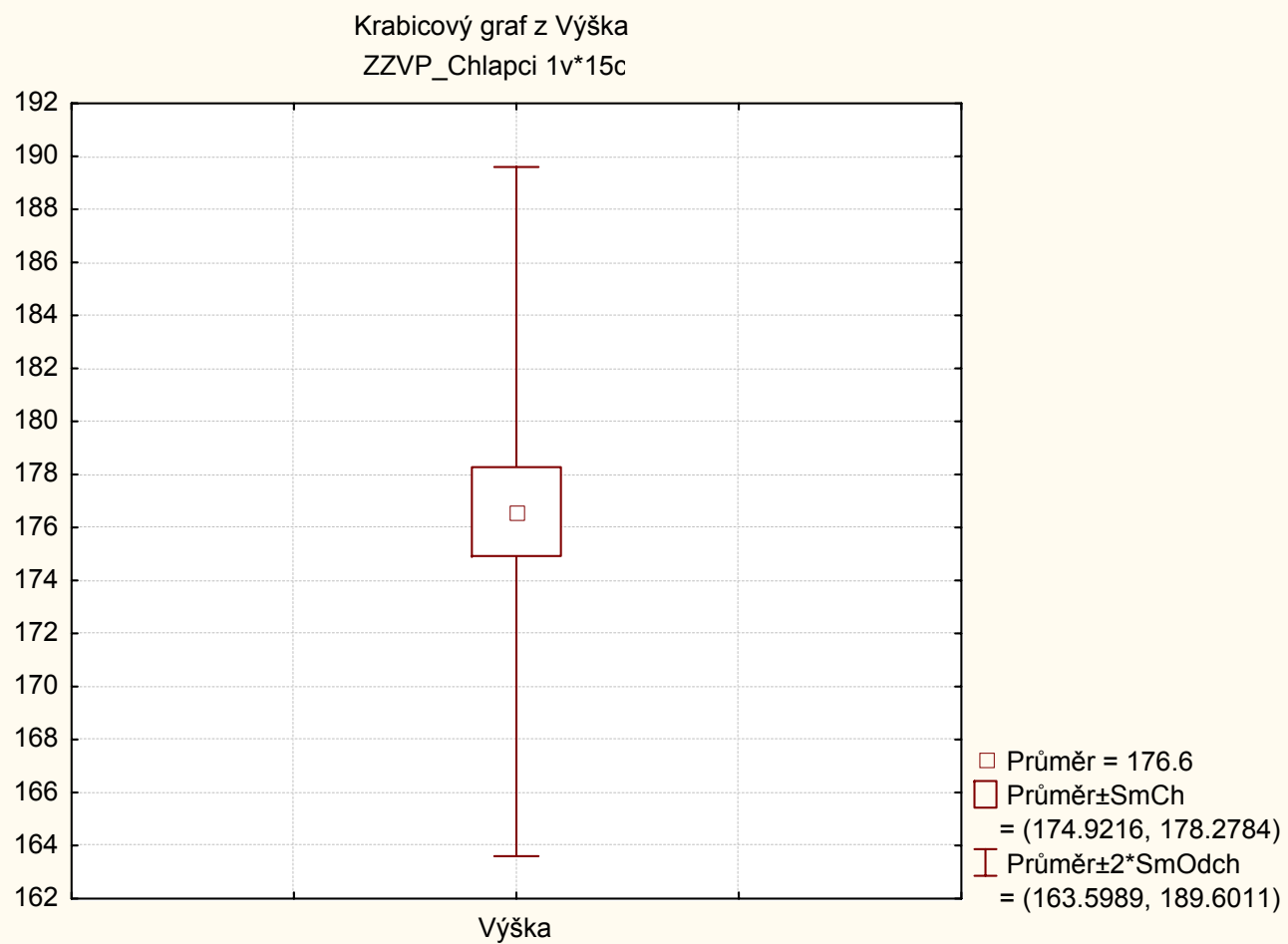
Popisné statistiky v programu Statistica

Střední hodnota a ukazatel rozptylu

Krabicový graf:

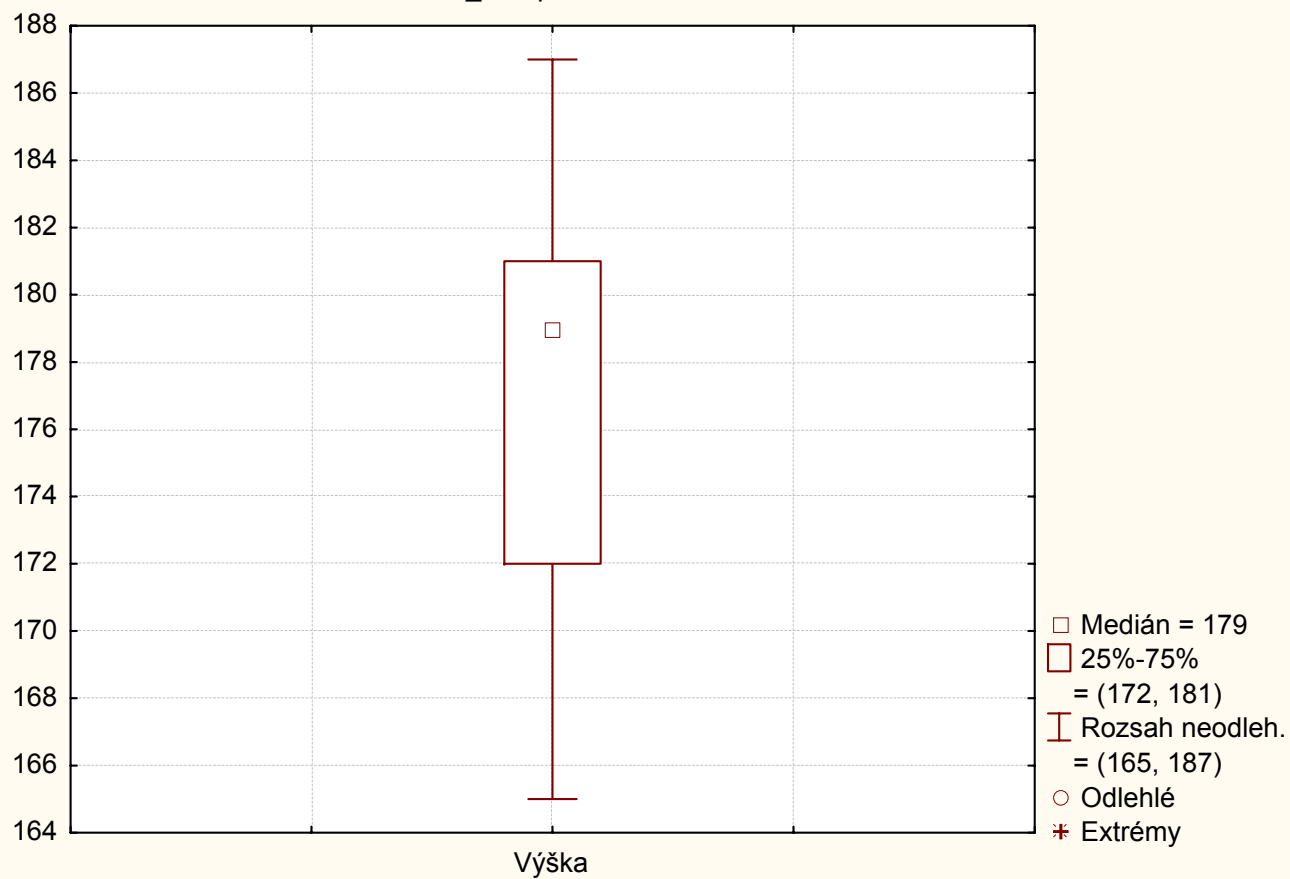
1. Otevřít datový soubor (příklad: Chlapci.xls)
2. Grafy, 2D grafy, Krabicové grafy
3. Proměnné: Závislá proměnná
4. Details: lze zvolit typ střední hodnoty (průměr nebo medián) a indikace rozptylu (směrodatná odchylka nebo kvantily)

Krabicové grafy 1



Krabicové grafy 2

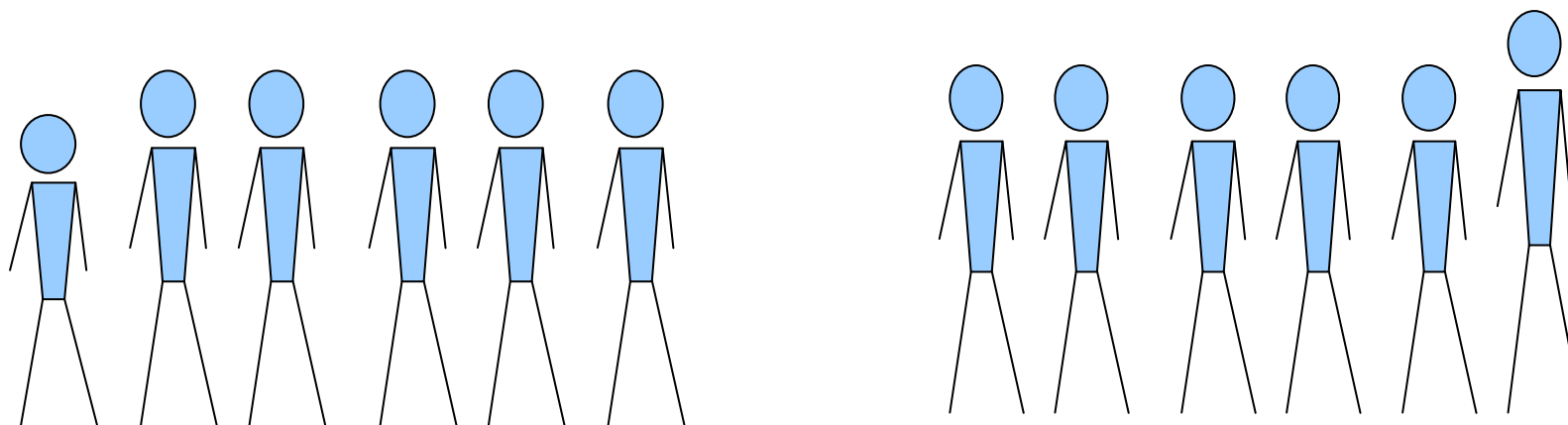
Krabicový graf z Výška
ZZVP_Chlapci 1v*15c



Testy odlišnosti dvou souborů

Odlišnost dvou datových souborů může být náhodná!

Proto nestačí pouze porovnat průměry, ale je třeba otestovat odlišnost souborů statisticky!



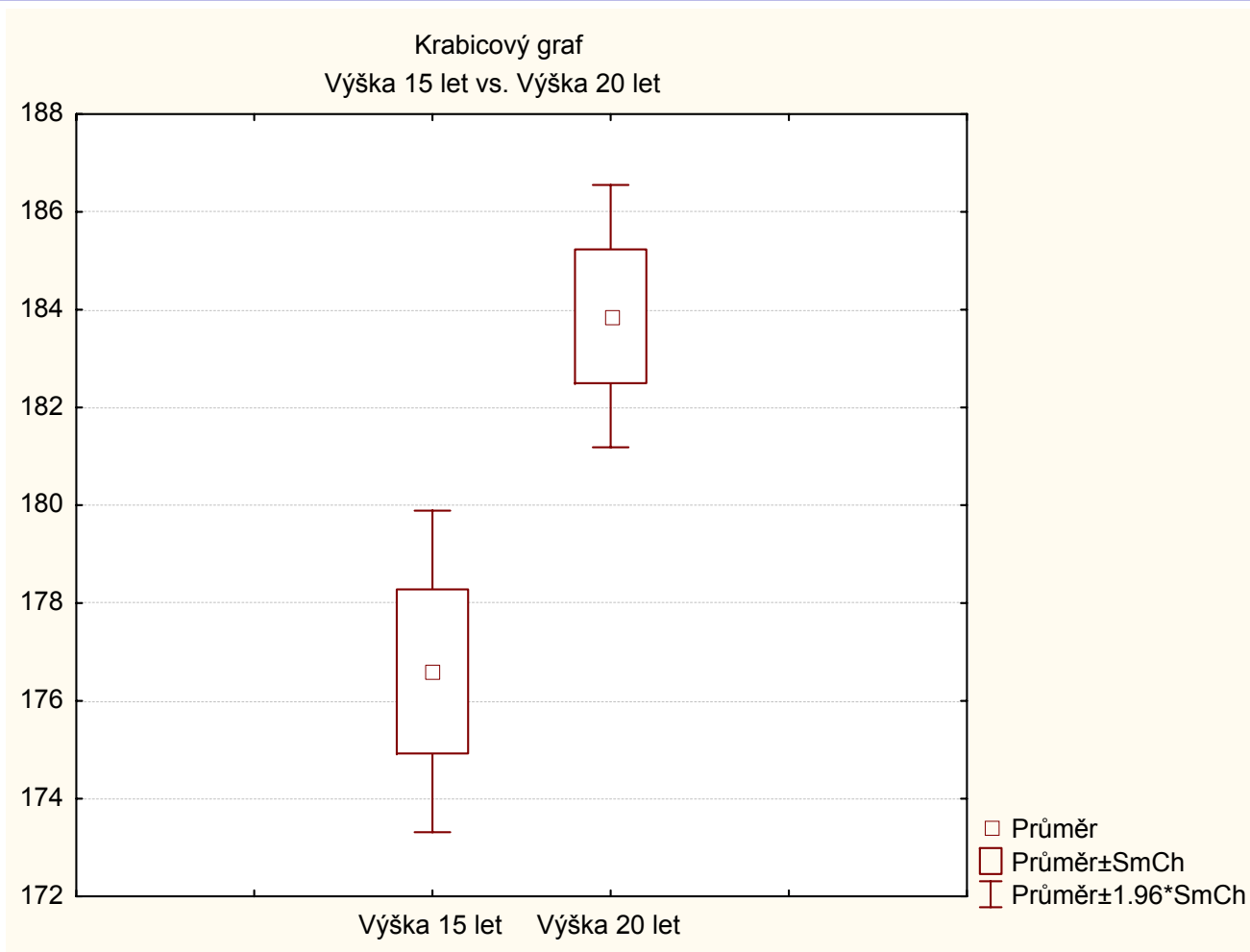
T-test v programu Statistica

- Příklad: dvě různé skupiny chlapců.
Testujeme, zda se liší jejich výška
 1. Statistica, otevřít datový soubor (Chlapci.xls)
 2. Menu: Statistika, Základní statistiky/tabulky
 3. T-test, nezávislé dle proměnných, OK
 4. Proměnné: vybrat proměnnou pro 1. seznam a pro 2. seznam, OK
 5. Výpočet
 6. Znovu zvolit analýzu (kliknout na minimalizované okno v levém dolním rohu)
 7. Krabicové grafy

T-test v programu Statistica

Skup. 1 vs. skup. 2	T-test pro nezávislé vzorky (ZZVF) Pozn.: Proměnné byly brány jako		
	Průměr skup. 1	Průměr skup. 2	Hodnota t
Výška 15 let vs. Výška 20 let	176.6000	183.8667	-3.35470

T-test v programu Statistica

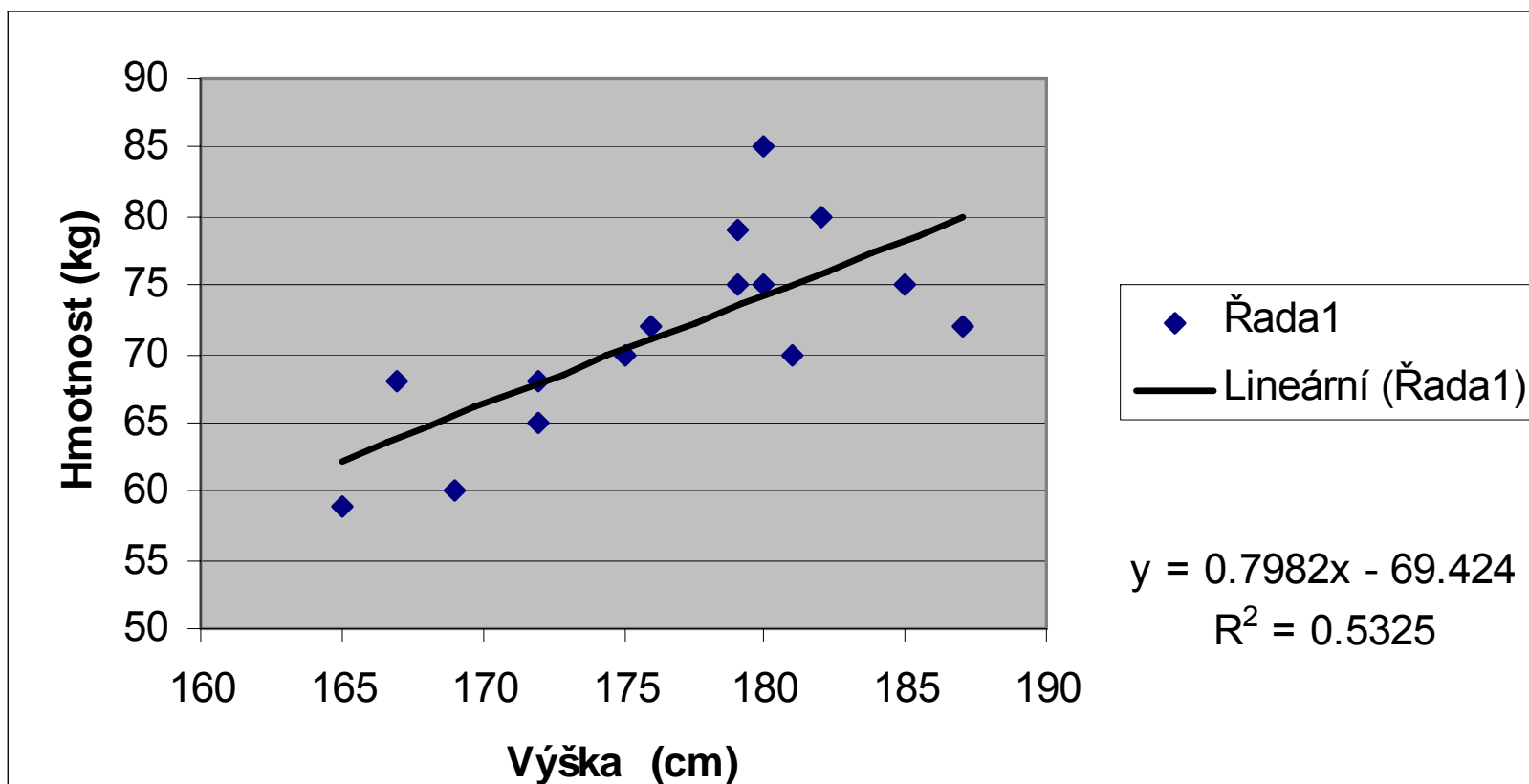


Vztah dvou proměnných

Nejjednodušší v programu MS Excel

1. Otevřít datový soubor (Chlapci.xls)
2. Sestrojit graf x-y bodový
3. Zvolit Graf, Přidat spojnici trendu, typ: lineární
4. Graf, Přidat spojnici trendu, Možnosti, Zobrazit rovnici regrese, Zobrazit hodnotu spolehlivosti R

Vztah dvou proměnných



R² = determinační koeficient, udává procento variability vysvětlené danou závislostí

4.2 Začlenění číselných dat do textu

Jak napsat kapitolu Výsledky

1. Uspořádání primárních dat do tabulek

- Doporučeny jsou tabulky do velikosti A4. Větší tabulky je vhodné rozdělit na několik menších.
- Tyto tabulky lze po formální úpravě zahrnout do datových příloh BP.

2. Výběr způsobu (statistického) zpracování primárních dat

Pro každý soubor primárních dat je třeba najít vhodný způsob souhrnného zhodnocení a prezentace.

Příklad 1 (tabulkové srovnání více hodnot):

- Ze změřených dat spočítat průměry a směrodatné odchylky.
- Získané hodnoty průměrů a odchylek uspořádat do souhrnné tabulky.
- Tuto tabulku umístit do kapitoly Výsledky.

2. Výběr způsobu (statistického) zpracování primárních dat

Pro každý soubor primárních dat je třeba najít vhodný způsob souhrnného zhodnocení a prezentace.

Příklad 2 (grafické srovnání více hodnot):

- Ze změřených dat spočítat průměry a směrodatné odchylky.
- Získané hodnoty průměrů a odchylek vynést do sloupcového grafu s odchylkami nebo do krabicového grafu.
- Tento graf umístit do kapitoly Výsledky.

2. Výběr způsobu (statistického) zpracování primárních dat

Pro každý soubor primárních dat je třeba najít vhodný způsob souhrnného zhodnocení a prezentace.

Příklad 3 (časová závislost):

- Ze změřených dat spočítat průměry. směrodatné odchylky.
- Získané hodnoty průměrů vynést do grafu (x-y bodový) v závislosti na čase.
- Tento graf umístit do kapitoly Výsledky.

3. Logické uspořádání souhrnných grafů a tabulek

- Souhrnné tabulky a grafy je vhodné vytisknout a seřadit za sebou tak, aby na sebe logicky dobře navazovaly.

Příklady uspořádání:

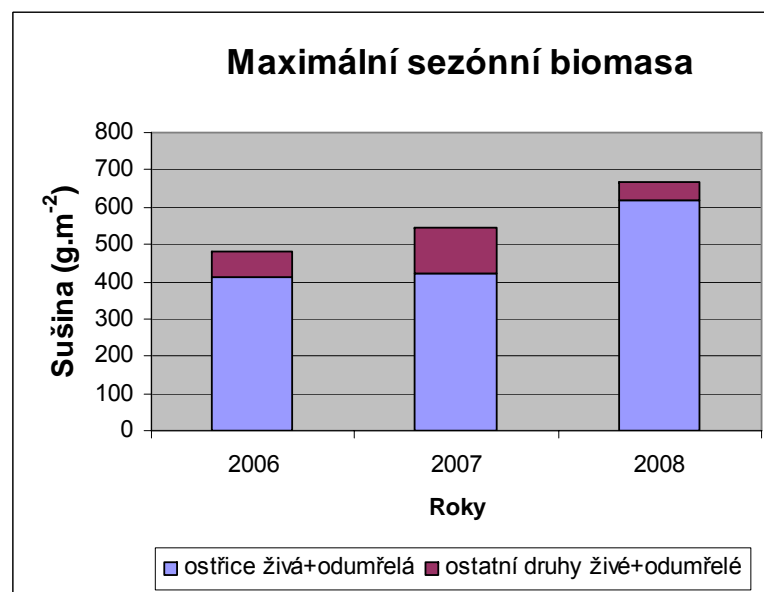
- Časová návaznost
- Varianty pokusu
- Opakování pokusu

4. Textové komentáře k tabulkám a grafům

- Ke každé souhrnné tabulce či grafu je třeba napsat krátký textový komentář. Cílem komentáře je upozornit čtenáře na nejdůležitější zjištění. Komentář musí obsahovat odkaz na danou tabulku či graf. Může obsahovat nejdůležitější číselné údaje. Komentář k jedné tabulce či grafu je obvykle dlouhý 1-2 věty až jeden odstavec. Komentář se řadí v textu nad tabulku či graf.

Příklad: srovnání hodnot

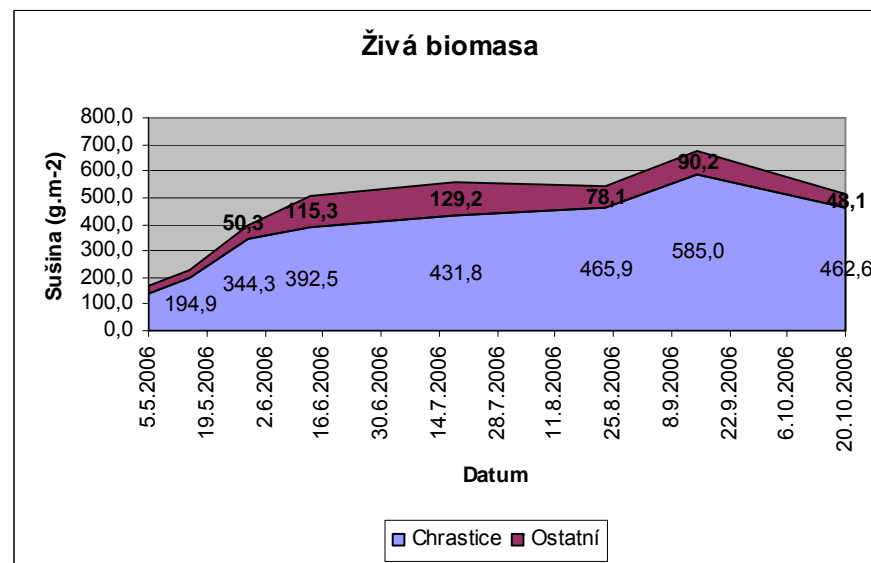
... Průměrná nadzemní biomasa ostřice štíhlé v době sezónního maxima v roce 2007 dosáhla 423 g. Hodnota nadzemní biomasy ostatních (doprovodných) druhů byla 123 g. V roce 2008 byly zjištěny výrazně vyšší hodnoty biomasy ostřice a celkové biomasy. Biomasa ostřice štíhlé dosáhla 618 g, biomasa doprovodných druhů však tvořila jen 52 g (Graf č.1).



Graf č. 1: Maximální sezónní nadzemní biomasa porostu s dominantní ostřicí štíhlou v letech 2006-2008

Příklad: sezónní chod

Sezónní chod **živé nadzemní biomasy** porostu s dominantní chrasticí ukazuje graf č.1. Biomasa rychle přirůstala zejména na počátku vegetační sezóny. Největší nárůst celkové živé biomasy byl zaznamenán v období od 15.5 do 29.5., a to ze 224,7 g.m⁻² na 394,6 g.m⁻² (tj. o 169,9 g.m⁻²). (...)



Graf č. 1. Vývoj živé nadzemní biomasy chrastice rákosovité a ostatních rostlinných druhů na Mokřých Loukách u Třeboně během roku 2006

Vhodné slovní obraty: popis časových chodů

- Hodnoty rostly (klesaly) v období od ... do ...

Příklad: Hodnoty živé nadzemní biomasy rostly v období od 5.5 do 29.5.

- Nejvyšší (nejnižší) hodnota byla zjištěna ...

Příklad: Nejvyšší hodnota nadzemní biomasy byla zjištěna na počátku srpna.

Vhodné slovní obraty: srovnání různých variant

Výnos sušiny byl větší u rostlin hnojených ve srovnání s nehnojenými.

Všechna níže uvedená tvrzení lze použít, ale každé znamená něco trochu jiného. Jaké jsou významové rozdíly mezi nimi?

- Průměrný výnos sušiny byl větší ...
- Výnos sušiny byl mnohem větší ...
- Výnos sušiny byl (statisticky) průkazně větší ...

Významové rozdíly

Jaké jsou významové rozdíly mezi tvrzeními níže?

- Průměrný výnos sušiny byl větší ...
Srovnávají se jen průměrné hodnoty, nevíme nic o rozptylech (směrodatných odchylkách). Nemůžeme říci, zda rozdíly jsou průkazné (a tedy můžeme je brát vážně)
- Výnos sušiny byl mnohem větší ...
Obecný popis, neříká nic o tom, zda rozdíl byl testován statisticky
- Výnos sušiny byl (statisticky) průkazně větší ...
Text odkazuje na výsledek statistického testu.

Vzory vhodných slovních obrátů

Najděte si vlastní vzory vhodných slovních obrátů!

- Publikované odborné články v češtině
- (Studentské práce jsou méně vhodné, častěji obsahují neobratnosti)

Popis k tabulkám a obrázkům

- Nad každou tabulku a pod každý obrázek patří popis (popisek).
- Popis má být samovysvětlující – má obsahovat všechny podstatné informace, aby je čtenář nemusel hledat v různých částech práce.
- Popis je heslovitý – neplést si s komentářem v textu!

DÚ

1. Utvořte tabulku primárních dat s popiskem
2. Na podkladě těchto primárních dat vytvořte souhrnnou tabulku nebo graf (opět s popiskem)
3. K souhrnné tabulce či grafu napište textový komentář o délce jednoho odstavce.