

Statistika

RNDr. Lucie Burešová
E-mail: 52031@mail.muni.cz

Výuka 20.2. – 28.5.2016

Přednáška: St. 9:55 – 10:45

Cvičení: St. 9:00 – 9:50 (na ½)

Zápočet:

V průběhu semestru 1 písemka (20 bodů)

Na konci semestru písemka 40 bodů

Pro získání zápočtu ≥ 30 bodů

Zkouška (strašně těžká):

Písemná část (doma):

Samostatné vyhodnocení datového souboru,
přehledný dokument s výsledky (poslat včas)

Ústní část:

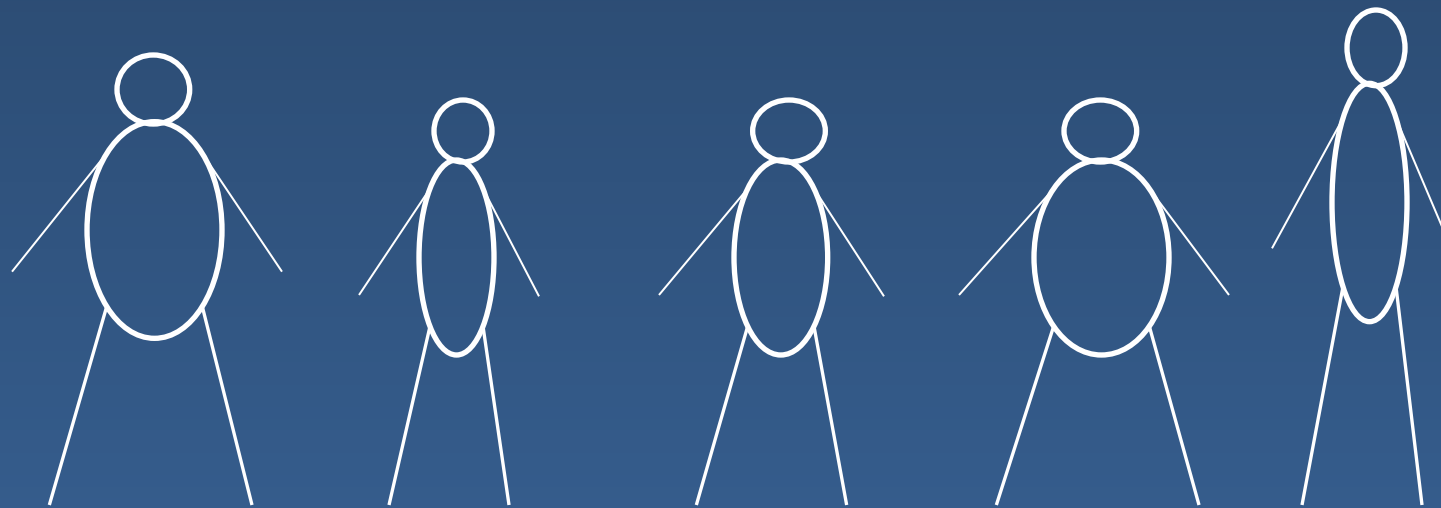
Konzultace dosažených výsledků, použitých
metod, základní teoretické otázky ...

1. Význam statistiky v biomedicínském výzkumu

K čemu potřebujeme statistiku?

- Porozumění odborné literatuře
- Diagnostika, prognóza onemocnění
- Účast na sběru nebo vlastní sběr dat
- Interpretace shromážděných dat
- Testování hypotéz
- Tvorba odborných publikací

Variabilita (všichni nejsme stejní)



135 kg

60 kg

82 kg

117 kg

94 kg

Přirozená variabilita znaku v populaci

- V případě deterministického světa by statistická analýza nebyla potřebná

Co je to statistika?

**vědecké studium dat popisujících existující
variabilitu**

(tedy zajímají nás jak rozdíly, tak podobnosti
hromadných jevů)

**a hodnotící různé domněnky (hypotézy)
vysvětlující data**

Hromadný jev – jev vyskytující se u souboru
statistických jednotek (věková struktura obyvatel
ČR)

Principy

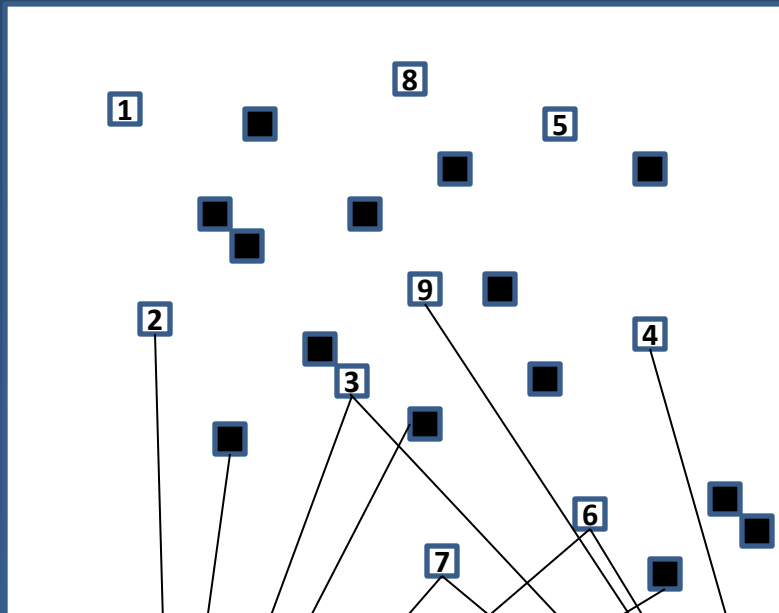
- **Popisná (deskriptivní) statistika** – uspořádání souborů, jejich popis a účelná sumarizace
- **Indukce** – proces zobecňování poznatků z výběru na celou populaci
- **Dedukce** - z obecných zákonitostí činíme závěry pro jednotlivé případy

2. Základní statistické pojmy

Data

Proměnné (znaky, veličiny)

Populace (základní soubor)
Konečná x nekonečná



Datový soubor

ID	pohlaví	věk	váha	počet onemocnění
1	Muž	29	119	1
2	Žena	35	58	4
3	Muž	70	79	0
4	Muž	58	85	6
5	Žena	33	83	4
6	Muž	40	88	0
7	Žena	82	92	0
8	Žena	64	63	2
9	Muž	47	95	3

Individuální pozorování
(statistické jednotky)

Výběr
(Úplné x výběrové šetření)

pokus x šetření

Data (naměřené hodnoty)

Kvantitativní a kvalitativní znaky

Data poměrová
(výška, hmotnost)



Data intervalová
(teplota ve °C)



Data ordinální
(závažnost onemocnění)



Data nominální
(pohlaví, typ operace)

Kolikrát?

O kolik?

Větší, menší?

Rovná se ?

**Kvantitativní
znaky**
(varianty číslem)

**Kvalitativní
znaky**
(varianty slovně)

Spojité

Diskrétní



Druhy náhodného výběru

Reprezentativní výběr (odráží strukturu základního souboru)

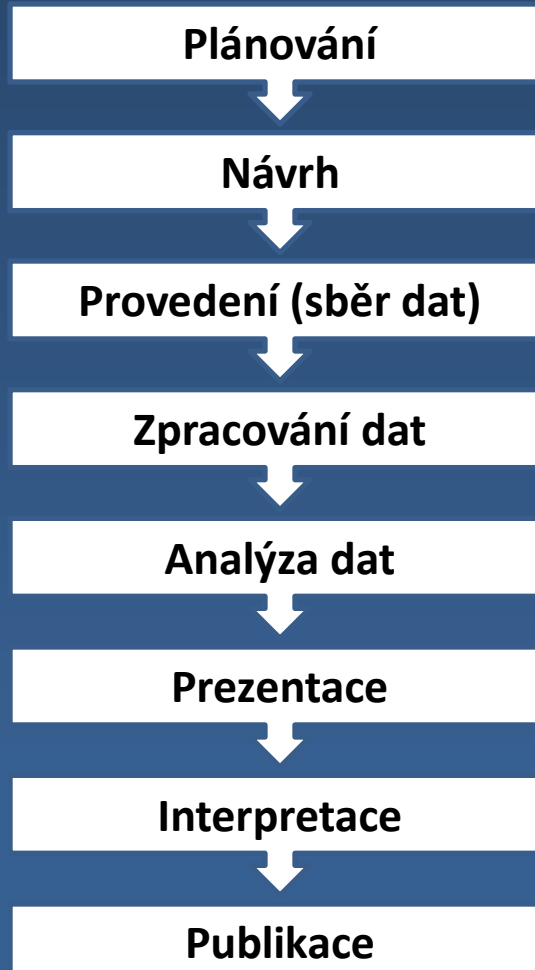
- zajistí **náhodný výběr** (každý prvek základního souboru musí mít stejnou pravděpodobnost, že bude vybrán, nezávislou na tom, zda bude vybrán prvek jiný)

- **Prostý náhodný výběr** - losování
- **Mechanický výběr** – pořadová čísla, krok
- **Oblastní výběr (stratifikovaný)** – oblasti (strata), z nich prostý nebo mechanický výběr
- **Skupinový výběr** – výběr skupin (rodina, škola, obec, okres)
- **Vícestupňový výběr** – přes vyšší výběrové jednotky (města - bloky - domy – domácnosti)
- **Párový výběr** – nemocný – zdravý se shodnými rizikovými faktory (case-control studie)

Náhodnost výběru je podmínkou statistické indukce!

3. Uplatnění statistiky v různých fázích klinického výzkumu

Schéma výzkumného projektu



1. Plánování a návrh

- přesná formulace cíle a účelu
 - výsledky předchozích výzkumů
 - Cíl: Vytipovat osoby, které se snadno mohou nakazit chřipkou.
 - Účel: Provést preventivní opatření.
- spolupráce lékaře a statistika při vytváření hypotéz
- vymezení základního souboru
- úplné x výběrové šetření (metoda výběru, rozsah výběru)
- přesné určení sledovaných znaků
- povaha znaků a měrné stupnice

Metody sběru dat

- Observační metody
 - přímé pozorování, klinické, biochemické vyšetření ...
- Rozhovor a dotazník
 - záměrně cílené otázky
- Použití dokumentace
 - původní zdravotnická dokumentace

- Výběr statistických metod
- Technika zpracování (software)
- Způsob záznamu údajů

2. Sběr a zpracování dat

- Uspořádání pilotní studie
- Sjednotit kvalitu dat mezi různými pracovníky
- Zdůvodnění vyřazení některých subjektů
- Oprava formálních chyb

Počítačové zpracování dat

Výhody	Nevýhody
Přesnost a rychlost	Chyby v softwaru
Univerzálnost	Univerzálnost
Grafika	Černá skříňka
Flexibilita	Špatná data plodí špatné závěry
Nové veličiny	
Velikost datových souborů	
Snadný přenos dat	

2. Analýza, interpretace a prezentace výsledků

- Popisná (deskriptivní statistika)
- Statistické třídění – podle třídících znaků (jednostupňové x vícešupňové)
- Kvalitativní – definované kategorie
- Kvantitativní – třídní intervaly (musí pokrýt všechny hodnoty, nepřekrývají se, hranice intervalů okrouhlá čísla)
- Intervaly stejně dlouhé x různě dlouhé
- 5 až 20 intervalů

Tabulky

Název
↓

Sloupc

Řádky

Tabulka 1.: Rozdělení dětí ve věku 9,5 - 10 let podle tělesné výšky (délka třídního intervalu 5 cm)

Hlavička →	Střed třídícího intervalu	Počet dětí	%	Kumulativní počet	Kumulativní %
	125	1	3,8	1	3,8
	130	3	11,1	4	14,9
	135	4	14,8	8	29,7
	140	12	44,4	20	74,1
	145	4	14,8	24	88,9
	150	3	11,1	27	100,0
	Celkem	27	100,0	-	-

Střed nebo hranice intervalu

Absolutní četnosti

Relativní četnosti

Absolutní kumulativní četnosti

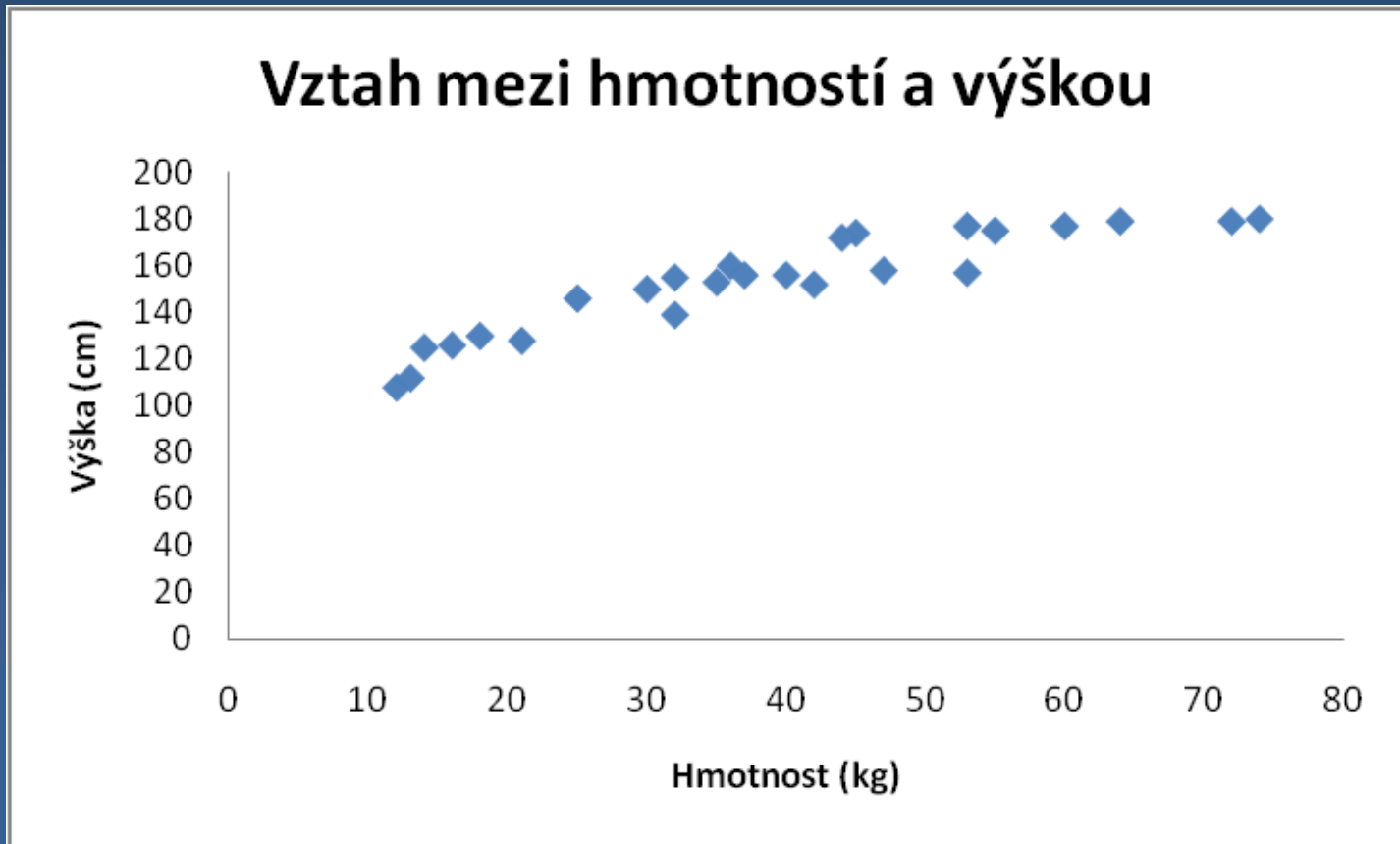
Relativní kumulativní četnosti

Grafy

- Přehlednější znázornění informací
- Pravoúhlá soustava souřadnic (rovnoměrná x nerovnoměrná stupnice (logaritmická))
- Název
- Bodový graf
- Spojnicový graf
- Sloupcový graf
- Histogram
- Koláčový (výsečový) graf
- Krabicový graf

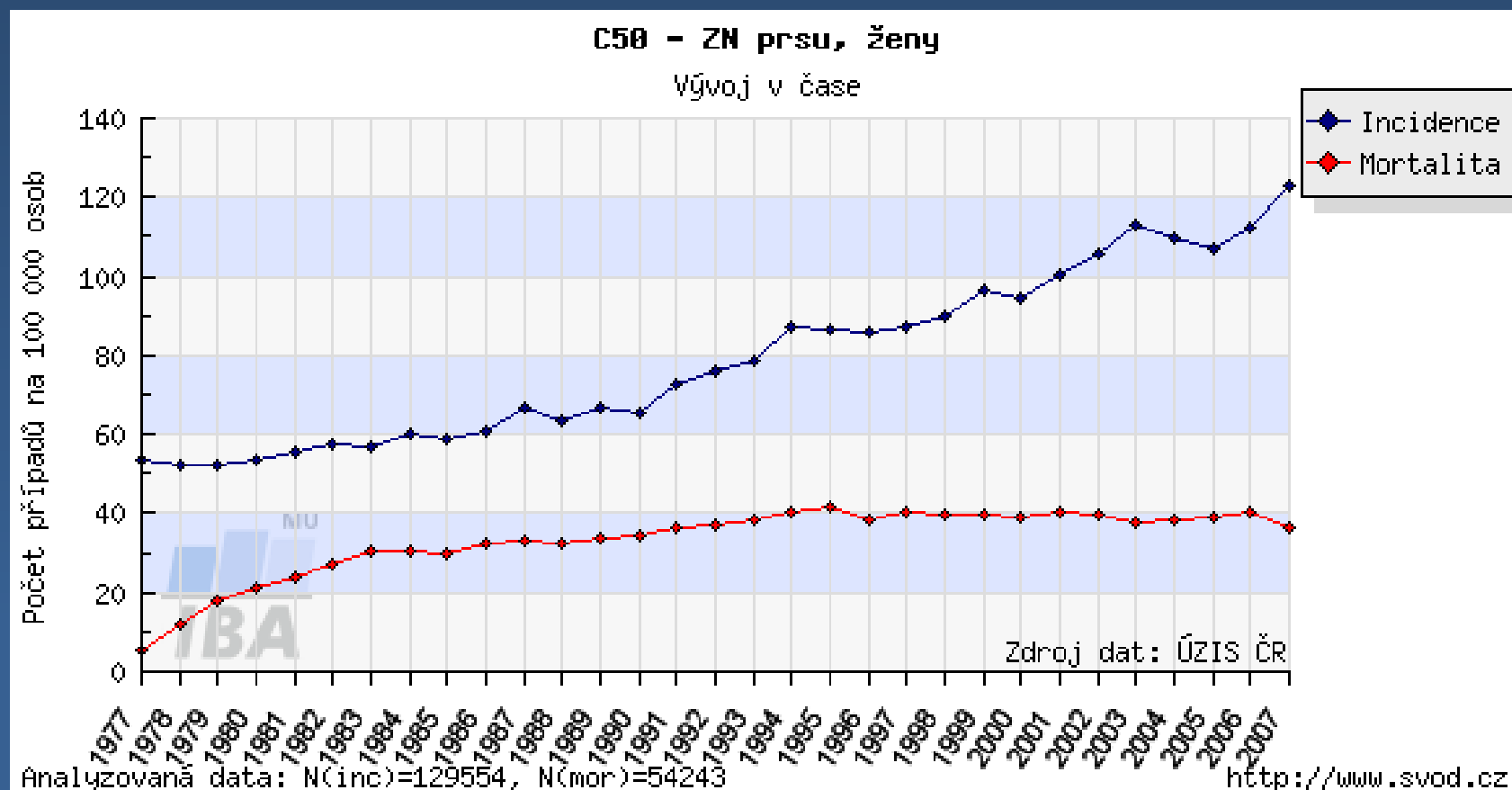
Bodový graf

- Zázornění závislosti dvou znaků

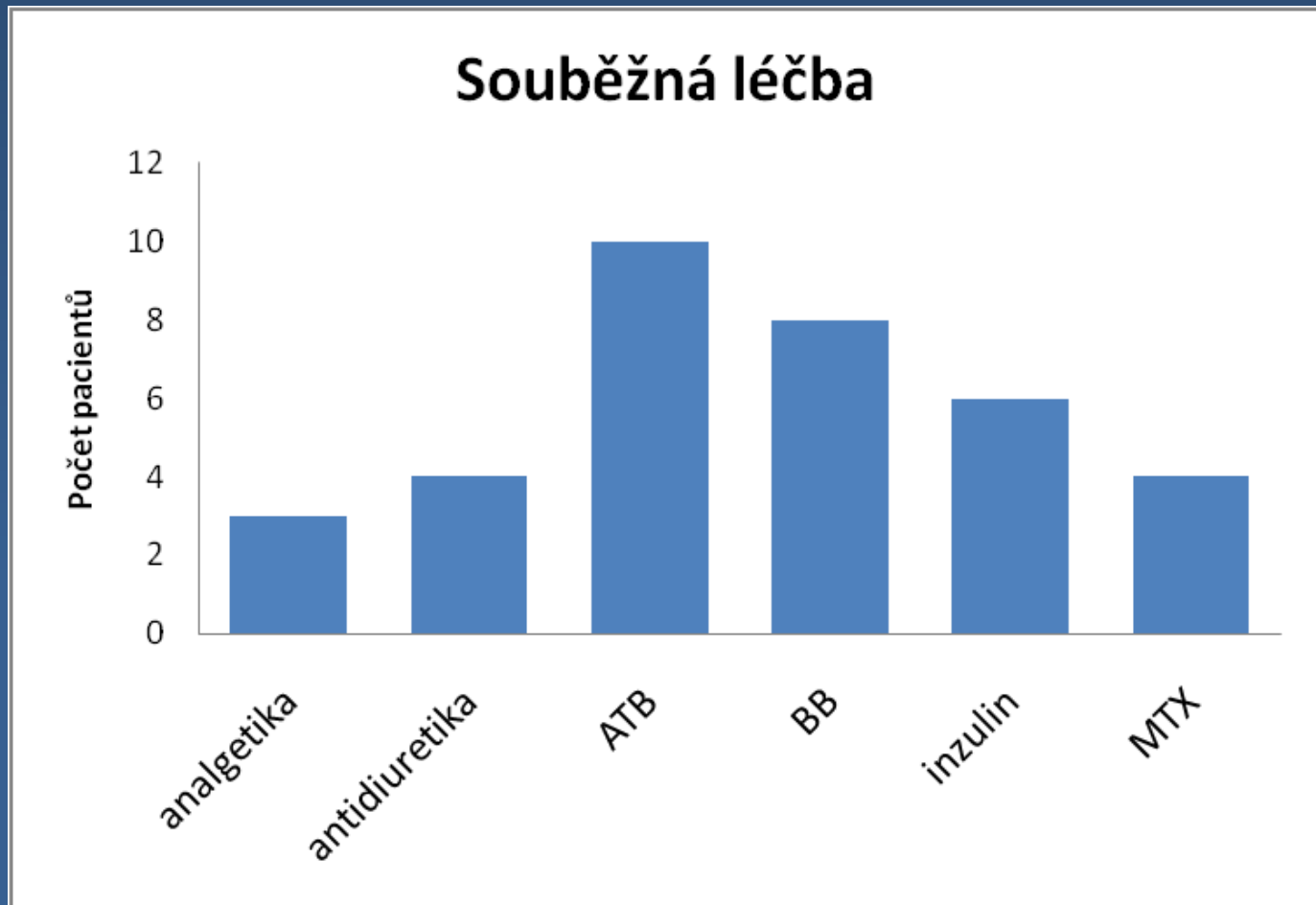


Spojnicový graf

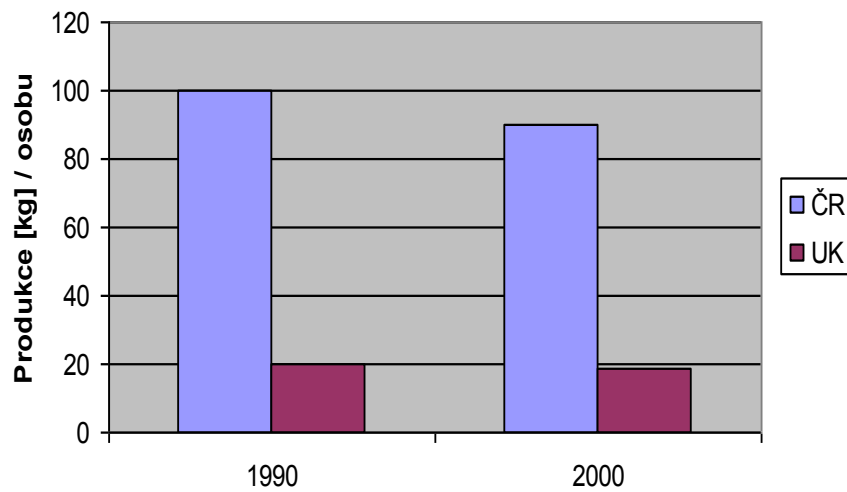
- Průběh časové řady, polygon četností



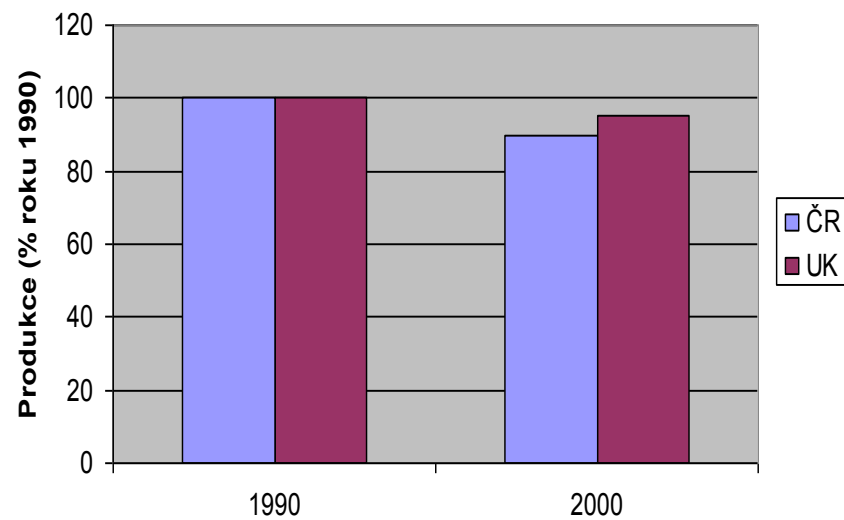
Sloupcový graf



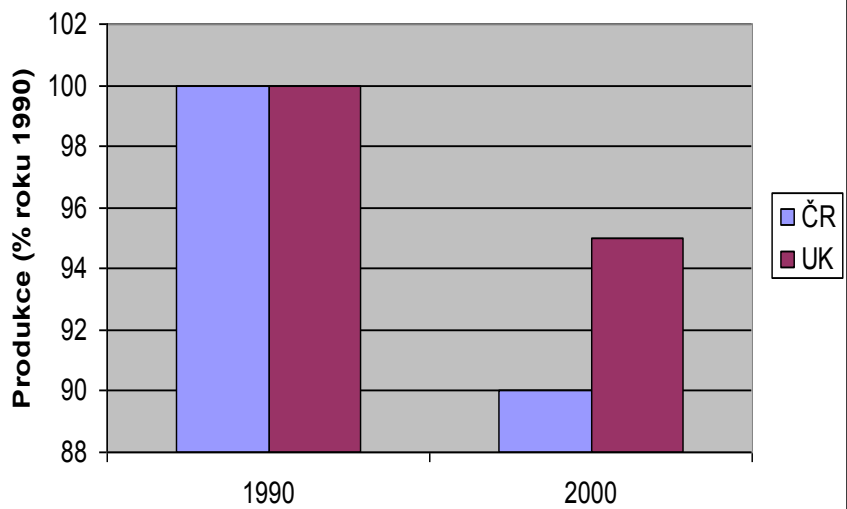
Produkce škodlivin



Produkce škodlivin

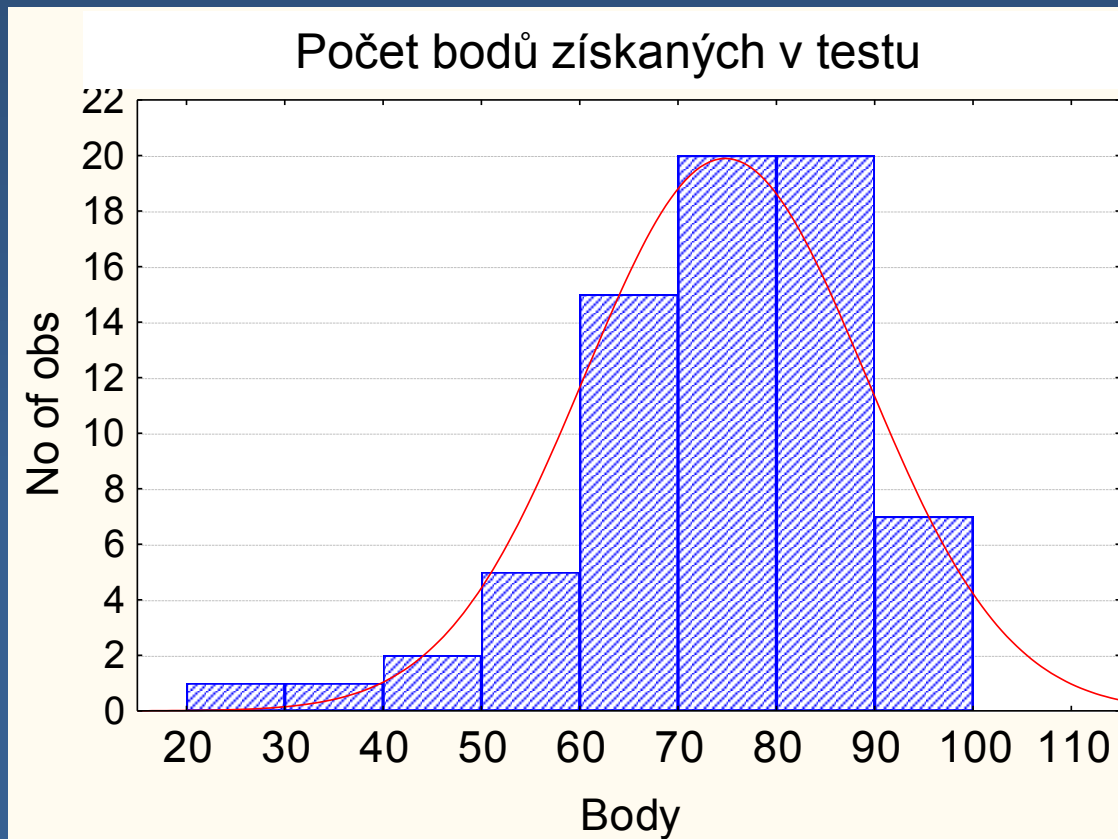


Produkce škodlivin



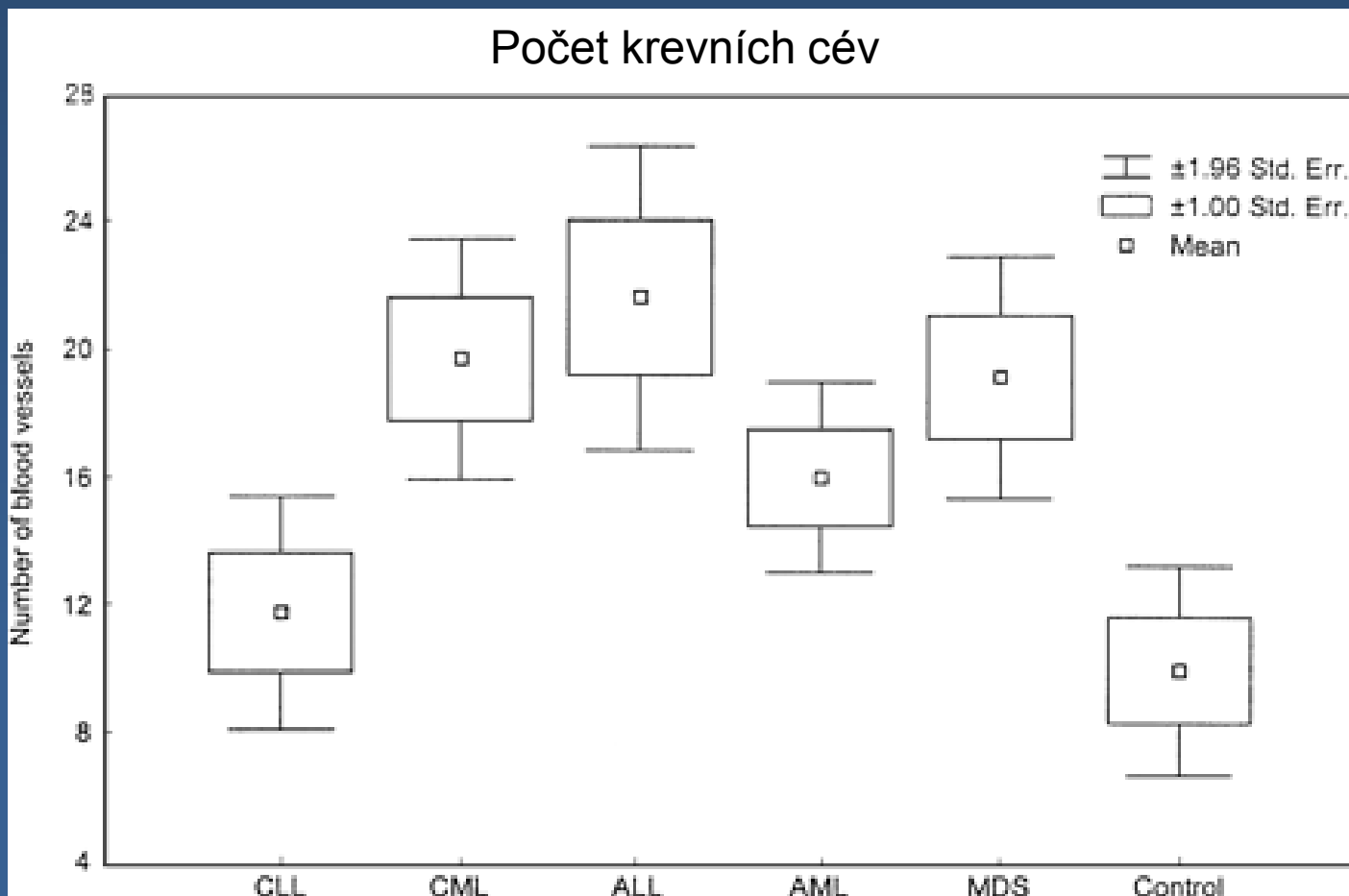
Histogram

- Rozdělení četností spojitého znaku (absolutních nebo relativních)
- Sloupce vždy vertikální
- Četnost odpovídá ploše sloupce (šířka x výška)



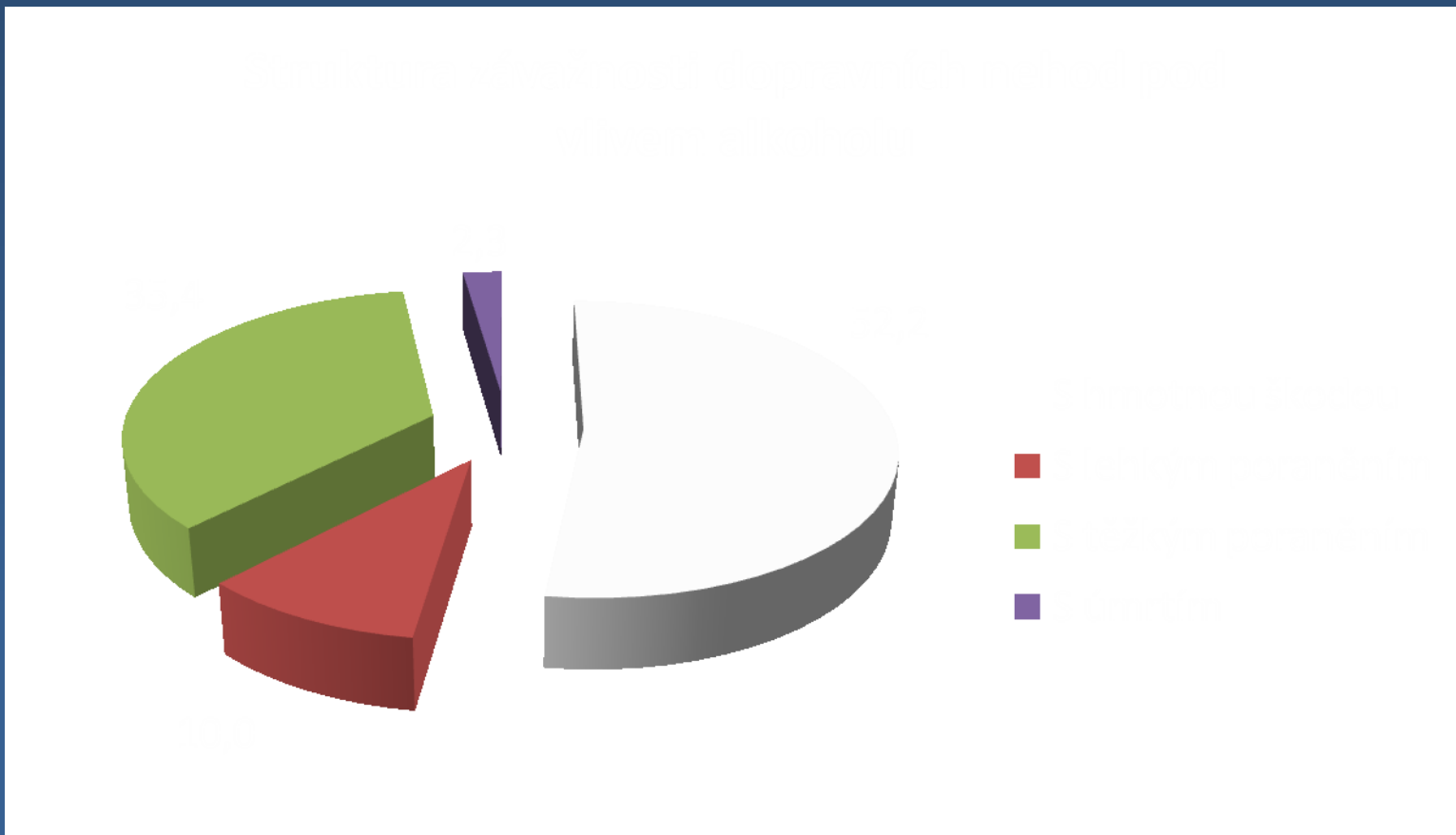
Krabicový graf (box-whisker plot)

- Obvykle medián, dolní a horní kvartil, min. – max.



Výsečový (koláčový) graf

- Plocha kruhu představuje celý soubor



3. Publikace výsledků výzkumu

- Pozornost při plánování, provádění, analýze a interpretaci výzkumu
- Pečlivě číst a interpretovat publikované výsledky
- Standard statistických metod se může lišit mezi jednotlivými časopisy

Struktura publikace

- Souhrn – několik vět x strukturovaný
- Úvod – krátké shrnutí předchozího výzkumu, vymezení cíle práce
- Metody – způsob získávání dat, návrh studie, počet subjektů, použité statistické metody, terminologie, nakládání s chybějícími údaji
- Výsledky – odpovídají cílům, tabulky a grafy sumarizující data a popisující výsledky
- Diskuse – konfrontace s jinými publikacemi, upozornění na nedostatky studie