

Statistické metody v biomedicínském výzkumu

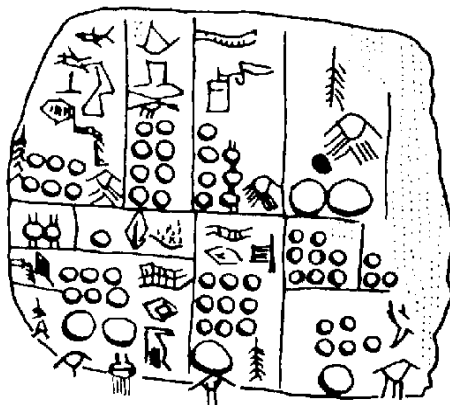
1. Úvod

Statistika je vědní obor, jehož výsledky můžeme při své práci používat prakticky denně. Neexistuje snad jediný odborný medicínský časopis, ve kterém bychom nenašli články obsahující odbornou statistickou terminologii. Často se můžeme setkat s prezentací výsledků klinického testování nového léčivého přípravku, s porovnáním různých metod léčby pacientů nebo se zprávami o incidenci konkrétní nemoci v jednotlivých regionech. (Incidence je podíl počtu nově hlášených nemocných jedinců za dané časové období a počtu všech jedinců ve sledované populaci). Takové studie používají k vyjádření nových poznatků vždy metody statistické analýzy.

Vysoká frekvence vědeckých sdělení v odborných časopisech však u čtenářů automaticky nezajišťuje znalost jazyka statistické analýzy. Dlouholeté zkušenosti v oblasti vědeckého výzkumu a využití metod statistiky ukazují spíše na to, že mnohé statistické pojmy, respektive pojmy z výzkumu jsou i odbornou veřejností chápány nepřesně, nebo dokonce nejsou pochopeny vůbec. V takovém případě však odborná sdělení nemohou přinést předpokládaný efekt.

11.1. Historie statistiky

Statistika je velmi starý vědní obor, který provází lidskou společnost už od počátku. Slovo statistika pochází z latinského „status“ – stav. Statistika byla původně nástrojem k zaznamenávání dat důležitých a dokonce nezbytných pro stát a jeho řízení. Těmito důležitými informacemi byly údaje o počtu obyvatel, výši daní, velikosti území, počtu vojáků, stavu zásob atd. Tyto informace se vždy získávaly prostřednictvím pozorování a měření. Nejstarší písemné památky mající statistickou povahu pochází z oblasti Sumeru, byly zaznamenány informace o čase, počtu osob, počtu kusů domácího zvířectva a úrodě – viz. obrázek.



Obrázek: Záznam psaný protosumerským písmem

Starověké říše byly zcela finančně závislé na objemu vybraných daní ve formě naturální, peněžní nebo ve formě pracovních povinností obyvatelstva. K zaznamenání těchto údajů se používaly opracované „statistické metodiky“.

- Egypt od r. 2850 př. n. l. – pravidelně za 2 roky se provádělo sčítání obyvatel, soupis dobytka, výše daní a stavu zásob. Zvlášť propracovaný systém evidence (census) se používal ve starověkém Římě.
- Středověkými „statistickými ročenkami“ byly vrchnostenské urbáře, které zaznamenávaly příjem poddaných, rozsah pozemkového vlastnictví šlechty a církve, a městské berní knihy a berní rejstříky, které obsahovaly výčet poplatníků spolu s údaji o jejich nemovitém majetku a výši odvedených daní.
- Ve 14. století se v Evropě objevují první církevní matriky, jako cenný zdroj informací o přirozené změně obyvatelstva.
- 16. století s sebou přineslo nový pohled na svět a člověka, rozvoj filozofie a počátky moderních vědních disciplín - **pojem STATISTIKA** poprvé použil Ital **Girolamo Ghilini** (v r. 1589).
- V 17. století dochází k vývoji moderní statistiky, především v oblasti teorie her a teorie pravděpodobnosti. V Anglii vzniká okruh statistiky, tzv. politická aritmetika (dnešní demografie) - vycházela z údajů o narozeních a úmrtích a pokoušela se na jejich základě zkoumat vývoj stavu obyvatelstva v delších časových údobích. **John Graunt** (1620-1674)



John Graunt, zakladatel demografie



Edmund Halley, teorie pravděpodobnosti

- **Edmund Halley** (1656-1742) - anglický astronom, matematik, meteorolog a fyzik - na konci 17. století při konstrukci úmrtnostních tabulek propojil statistiku s počtem pravděpodobnosti.
- Mimořádnou zásluhu o rozvoj statistiky měl belgický matematik, statistik a astronom **Adolphe Lambert Quételet** (1796-1874), který vypracoval zásady moderního sčítání lidu. Začal aplikovat numerické postupy na různé stránky společenského života (např. na zločinnost). Přišel s myšlenkou „průměrného člověka“ - na základě velkého souboru dat vypočítal základní průměrné tělesné rozměry člověka a odchylky jednotlivců od těchto průměrů. Je duchovním otcem pojmů průměr, střední hodnota, rozptyl a rozdělení.
- Statistika 19. a počátku 20. století se vyznačovala sběrem informací od co nejširšího okruhu respondentů, s cílem obsáhnout ve svém šetření celou populaci a tím získat maximálně přesný obraz stavu společnosti. Vzniká biometrická statistická škola. **K. Pearson** – aplikace statistiky v biologii.
- Počátkem 20. století se zrodila matematická statistika - disciplína, jejímž charakteristickým rysem je hledání metod, jež by umožnily vytvoření závěrů o celku na základě výběru. **R. A. Fischer** vytváří statistiku malých výběrových souborů.

I dnes se významná část statistiky zabývá zpracováním podobných dat a jejím cílem je měřit a poskytovat informace pro řízení. Zájem státu na shromažďování a vyhodnocování určitých

dat se projevuje existencí a podporou různých institucí - v našem případě je to především Český statistický úřad, který sbírá údaje o všech nemocnicích, obcích, určitých službách a ekonomických výsledcích. Dnešní statistika se však zdaleka neomezuje jen na prosté shromažďování a popis dat, ale umožňuje **zobecnovat poznatky z dat a podporovat kvalifikované rozhodování.**

Co je statistika?

Slovo statistika můžeme používat nejméně ve třech významech:

1. Číselné údaje o hromadných jevech - statistiku nezajímá to, co lze pozorovat a změřit pouze jednou, ale zabývá se pozorováním tzv. hromadných jevů (hromadný jev je jev, který se vyskytuje mnohokrát a neustále se může opakovat – např. zkoumání určité vlastnosti velkého množství objektů).
2. Praktická činnost spočívající ve sběru, zpracování a vyhodnocování dat.
3. Teoretická disciplína, která se zabývá metodami vyhodnocení hromadných jevů - složitá matematika, kterou se zabývají profesionální statistici.

Stručně bychom mohli říci, že *statistika je způsob shromažďování dat, práce s daty a jejich kvantitativní vyhodnocení, tedy interpretace.* Používá k tomu především metodu pozorování a popisu určité vybrané části reality (např. vybrané skupiny zdravotnických zařízení) s následným zobecněním těchto pozorování na celou realitu (všechna zdravotnická zařízení na sledovaném území).

1.2. Statistika a medicína

Používání statistických metod je dnes v medicíně velmi rozšířeno a základní znalosti ze statistiky jsou nezbytné nejen pro psaní odborných článků, ale i pro jejich čtení a studium.

V medicíně se často setkáváme s otázkami, které mají statistický charakter:

- „Je nový lék lepší než dosud užívaný?“
- „Jaké okolnosti ovlivňují výsledek léčby pacienta?“
- „Jaké významné ukazatele (charakteristiky) jsou spojeny s relapsem (návratem) choroby, úmrtím pacienta nebo vznikem komplikací při léčbě?“
- „Jaká je přesnost a citlivost dané diagnostické metody?“

Na tyto a podobné otázky umí odpovědět právě statistika.

1.3. Odvětví statistiky

Deskriptivní statistika

Deskriptivní (popisná) statistika se zabývá uspořádáním datových souborů, jejich popisem a sumarizací.

Induktivní statistika

Induktivní statistika se zabývá zobecňováním poznatků získaných z dat. Induktivní metoda (**indukce**) je princip usuzování, kdy z konkrétních poznatků získaných pozorováním a měřeními na vybraných částech reality vyvozujeme obecné zákonitosti platné pro celou realitu.

Při každém zobecňování poznatků dochází k ovlivnění výsledků subjektivními vjemy pozorovatele a takové poznatky mají proto jen omezenou platnost. Tento problém řeší induktivní statistika, která pracuje s metodami, umožňujícími stanovit u obecných závěrů **stupeň jejich spolehlivosti.** Výpočet stupně spolehlivosti je **objektivní**, protože je založen na poznacích **teorie pravděpodobnosti** a nikoliv na subjektivním názoru pozorovatele.

Kontrolní otázky:

1. Vyjmenujte některé významné statistiky, matematiky a astronomy, kteří se zasloužili o rozvoj vědního oboru statistika.
2. Čím se zabývá deskriptivní statistika?
3. Co je předmětem induktivní statistiky?

2. Základní pojmy

K problematice statistického výběru a vlastního výzkumu je třeba si nadefinovat několik pojmů. Hlavní roli v metodách induktivní statistiky hrají pojmy populace a výběrový soubor.

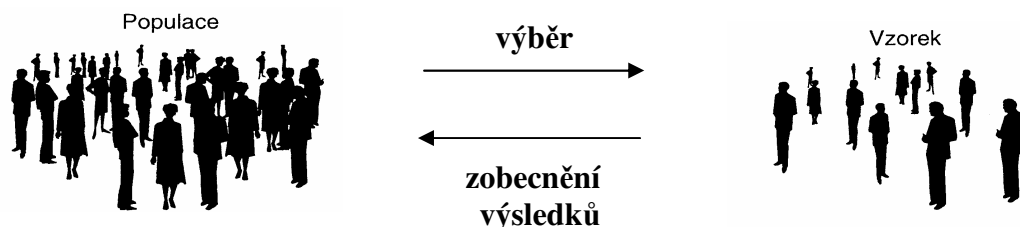
Populace (základní soubor) je definována přesným stanovením statistických jednotek buď jejich *výčtem* (v případě tzv. demografické populace), nebo *vymezením společných vlastností* statistických jednotek (v případě hypotetické populace). V medicínském výzkumu je obvyklé definovat populaci vymezením společných vlastností pacientů, kterých se daný výzkum týká.

Příklady definice hypotetické populace:

- Populaci tvoří ženy s hormonální antikoncepcí, u nichž došlo k hluboké žilní tromboze (tj. vzniku krevní sraženiny).
- Populaci tvoří pacienti s opravnou operací (tzv. revizí) totální endoprotézy kyčle (tj. celkové náhrady kyčelního kloubu).
- Populaci tvoří dospělí pacienti s diabetes mellitus (cukrovkou), kteří používají k aplikaci inzulínu inzulinovou pumpu.
- Populaci tvoří pacienti hospitalizovaní pro akutní infarkt myokardu (srdeční svaloviny).

Parametry populace (populační charakteristiky) jsou číselné charakteristiky populace. Obecně jde o neznámá čísla (např. průměrná výška dospělých mužů), která jsou předmětem odhadování.

Výběrový soubor je část (vzorek) základního souboru. V praxi jsou sledované vlastnosti zjišťovány jen u určitého počtu, neboli vzorku statistických jednotek, vybraného z populace podle určitých statistických pravidel. Počet jednotek ve výběrovém souboru značíme písmenem n . Pro účely zobecnování by měl být výběrový soubor dostatečně reprezentativní, tj. měl by dobře odrážet základní strukturu celé populace (např. stejný poměr muži/ženy, stejné věkové složení, stejné rozdělení pacientů podle stádia nemoci atd.).



Statistická jednotka – základní prvek, který je předmětem zkoumání (pacient, pokusné zvíře, transplantovaná ledvina, ...)

Znak – vlastnost sledovaná na jednotkách statistického šetření (u pacientů např. diagnóza, pohlaví, věk, délka hospitalizace, hladina celkového cholesterolu, glykémie, krevní tlak, ...).

3. Pravidla pořizování náhodných výběrů

Úplné šetření dnes provádíme v praxi jen zřídka, důvodem je většinou nedostupnost celé populace, ekonomická a časová náročnost a v některých případech i destruktivní charakter prováděného výzkumu (pokusné zvíře se musí usmrtit, testovaný výrobek se musí zničit, atd.).

Induktivní metoda, používaná ve výzkumné praxi, vyžaduje, aby byl **výběrový soubor dostatečně reprezentativní**. Při získávání reprezentativního výběrového souboru není přípustné, aby byl výběr statistických jednotek ovlivňován subjektivními postupy výzkumníků, proto používáme objektivní procedury, které jsou založeny na principech náhodnosti a nezávislosti.

Náhodností se rozumí to, že metoda, pomocí které je výběr pořizován, splňuje podmínku: každá statistická jednotka (v medicíně většinou pacient) má stejnou šanci dostat se do výběru. **Nezávislost** znamená, že výběr statistické jednotky (pacienta) není ovlivněn předchozími výběry.

3.1. Metody pořizování náhodného výběru

V medicíně se nejčastěji setkáváme se čtyřmi metodami:

- **prostý náhodný výběr**

získáme mechanismem výběru, ve kterém má každý prvek výchozí populace stejnou šanci dostat se do výběru. V praxi se používají počítačové programy tzv. generátory náhodných čísel, nebo tabulky náhodných čísel (prostý náhodný výběr lze pořídit také losováním). Předpokladem je, že všechny statistické jednotky základního souboru jsou očíslované.



Příklad:

Očíslujeme chorobopisy čísly 1 až 900. Pomocí generátoru náhodných čísel vygenerujeme řadu čísel od 0 do 9, např.: 7, 5, 6, 2, 3, 1, 4, 6, 9, 0, 8, 1, ... Do výběru zařadíme chorobopisy s čísly: 756, 231, 469, 081, ...

- **systematický náhodný výběr**

z výchozí populace vybereme každý k -tý prvek, kde $k = N/n$, N je počet prvků výchozí populace, n je požadovaný rozsah náhodného výběru. V praxi nejprve náhodně zvolíme číslo mezi 1 a 9, od kterého se výběr zahájí.

Příklad:

Výchozí populace má 100 prvků, do výběrového souboru chceme zařadit 20 prvků.

$N = 100$, $n = 20 \Rightarrow k = 5$... náhodně vybrané číslo = 6

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 ...

Do výběru budou vzaty náhodně vybrané prvky č. 6, 11, 16, 21, 26, 31, 36, 41, atd.

- **stratifikovaný náhodný výběr**

výchozí populace se nejdříve podle hodnot určitého znaku rozdělí do několika homogenních skupin (tzv. strat) - např. podle věku, po 10-ti letech. V druhém kroku se z takto vytvořených skupin provede prostý náhodný výběr.

Cílem stratifikace je zajistit, aby jednotlivé podskupiny výchozí populace byly ve výběru adekvátně zastoupeny. V medicínských studiích se kromě věku často stratifikuje podle pohlaví, věku, stádia nebo délky nemoci.

- **vícestupňový náhodný výběr**

výchozí populace je nejdříve rozdělena do několika hierarchicky uspořádaných úrovní.

Příklad: Chceme získat reprezentativní výběr žáků 7. tříd základních škol v ČR. Populaci tvoří žáci ZŠ českých a moravských měst nad 10 000 obyvatel. Populaci rozčleníme na úrovně: města - školy - třídy - žáci. V prvním kroku nejdříve ze všech českých a moravských měst (nad 10 000 obyvatel) náhodně vybereme několik měst. Ve druhém kroku potom v každém vybraném městě náhodně vybereme několik základních škol. V třetím kroku v každé vybrané škole náhodně vybereme několik 7. tříd a nakonec provedeme v těchto třídách náhodný výběr žáků (tj. jednotek na nejnižší úrovni).

Vícestupňový náhodný výběr se používá v situacích, kdy není k dispozici tzv. *výběrový rámec* (opora výběru), tj. seznam nebo jiný prostředek k identifikaci jedinců v celé populaci (v našem případě nemáme k dispozici seznam všech žáků ZŠ).

V praxi obvykle není snadné pořídit skutečně náhodný výběr. Ve většině případů je to tak, že lékař žádný náhodný výběr neuskutečňuje vzhledem k tomu, že pacientů splňujících daná kritéria není dost na to, aby se z nich dělal výběr. Daný vzorek pacientů je potom implicitně považován za náhodný výběr. V takové situaci je nutné zvážit, zda jsou pacienti, kteří se do vzorku dostali, skutečně reprezentativní pro populaci, na kterou se má zobecňovat.

3.2. Metody sběru dat

- **Observační metody** – metody založené na pouhém vizuálním pozorování nebo klinickém, mikrobiologickém nebo biochemickém vyšetření – jde o registrování a zaznamenání sledovaných událostí.
- **Rozhovor a dotazník** - o sledovaném jevu se dozvídáme zprostředkovaně dotazováním. V tomto případě existuje možnost zkreslení v důsledku nepochopení otázky nebo špatného záznamu odpovědi.
- **Studium dokumentace** - tzv. sekundární analýza, umožňuje získat informace z minulosti (zdravotní záznamy, záznamy o hospitalizaci, záznamy o pracovní neschopnosti, hlášení infekčních nemocí, hlášení úrazů, odborné posudky, MR nebo RTG snímky, ...).

3.3. Pilotní studie

Abychom se vyhnuli chybám při rozsáhlém sběru a zpracování dat, provádíme před vlastním výzkumem předběžné šetření - tzv. **pilotní studii**. Posláním pilotní studie je na malém vzorku výběrového souboru ověřit metody sběru, zpracování a vyhodnocení údajů a případně na základě tohoto poznání metody příslušným způsobem upravit. Teprve potom přikročíme k vlastnímu sběru dat v projektovaném rozsahu.

Kontrolní otázky:

1. Definujte pojem populace, uveďte příklady populací.
2. Definujte pojem výběrový soubor.
3. Co si představujete pod pojmem „parametry populace“?

4. Definujte pojmy „statistická jednotka“ a „znak“.
5. Vyjmenujte základní metody pořizování náhodných výběrů.
6. Vyjmenujte základní metody sběru dat.

4. Typy dat

Na sledovaných statistických jednotkách (pacienti, události, výrobky) zjišťujeme (měříme) určité **znaky**, tj. vlastnosti (charakteristiky) těchto jednotek.

Podle povahy rozlišujeme dva základní typy znaků - kvalitativní (kategoriální) a kvantitativní (metrický).

4.1. Kvalitativní (kategoriální) data

U kvalitativních znaků jsou varianty zkoumané vlastnosti dány většinou slovním vyjádřením - vyjadřují kvalitu měřené vlastnosti pomocí tzv. kategorií. K měření kvalitativních znaků používáme nominální nebo ordinální (tj. pořadovou) škálu. Podle použité měřicí škály potom mluvíme o datech nominálního nebo ordinálního typu.

Data nominální

Nominální měřicí škála je nejjednodušší měrnou škálou, představuje třídění statistických jednotek do dvou nebo více vzájemně se vylučujících kategorií. V případě dvou kategorií mluvíme o tzv. **Dichotomické stupnici** (např. kategorie: pozitivní/negativní, ano/ne, přítomen/nepřítomen, 0/1 atd.). Pro označení kategorií se také často používají čísla (např.: 1=muž, 2=žena), pro která však nelze použít obvyklé aritmetické operace! Kategorie nominálního znaku jsou rovnocenné a nelze je uspořádat.

Příklady nominálních znaků:

- pohlaví (kategorie: muž, žena),
- těhotenství (kategorie: normální, rizikové),
- Rh faktor (kategorie: pozitivní, negativní),
- Hypertenze (kategorie: ano, ne),
- krevní skupina (kategorie: A, B, AB, 0),
- rodinný stav (kategorie: svobodný(á), ženatý(vdaná), rozvedený(á), vdovec(vdova)).

Data ordinální (pořadová)

Ordinální měřicí škála je tvořena vzájemně se vylučujícími kategoriemi, které lze vzájemně uspořádat a lze tedy určit jejich pořadí, nemá však smysl určovat vzdálenost (rozdíl) mezi jednotlivými kategoriemi. Škále se někdy říká semi-kvantitativní.

Příklady ordinálních znaků:

- stádium nemoci (kategorie podle symptomů např.: I, II, III, IV),
- intenzita nežádoucího účinku (kategorie: mírná, střední, silná),
- invalidita (kategorie: žádná, částečná, úplná),
- GOS (kategorie: 1,2,3,4,5) ...Glasgow Outcome Score - škála používaná v traumatologii:
 - 1 = smrt,
 - 2 = vegetativní stádium,
 - 3 = závažné postižení,
 - 4 = střední postižení,
 - 5 = dobrý stav,
- mRS (kategorie: 0,2, ...,6) ... škála používaná v neurologii:

0 = bez deficitu,
1 = lehké postižení bez omezení aktivit,

...

5 = závislý, imobilní pacient,
6 = úmrtí

- reakce po očkování (kategorie: žádná, zarudnutí, otok+horečka),
- vztahy na pracovišti (kategorie: přátelské, čistě pracovní, nepřátelské),
- závažnost průběhu onemocnění, stupeň ortodontické anomálie (kategorie: definovány podle symptomů).

4.2. Kvantitativní data

Kvantitativní znaky vyjadřují měřenou vlastnost číslem. Jsou buď diskrétní (varianty znaku jsou vyjádřeny celými čísly – např. délka hospitalizace: 10 dní), nebo spojité (jednotlivé varianty znaku mohou nabývat jakékoliv hodnoty z určitého intervalu nebo rozmezí – např. tělesná hmotnost: 56,5 kg).

Kvantitativní znaky měříme na poměrové nebo intervalové stupnici. Tyto stupnice umožňují uspořádat jednotlivá pozorování podle velikosti a určit také jejich přesnou vzdálenost.

Na **intervalové stupnici** je pozice nuly věcí volby, nula může být na kterémkoli místě, aniž se tím změní povaha měření (např. libovolná kalendářní stupnice nebo teplotní stupnice ... 0°C není počátek stupnice, na stupnici lze měřit také záporné hodnoty). Lze vyjádřit rozdíl dvou hodnot, ale nelze vyjádřit jejich podíl. Nelze se ptát, kolikrát je jedna hodnota větší než druhá (Kolikrát je dnes tepleji než včera? – nemá smysl)

Na **poměrové stupnici** je pevně definována pozice nuly, 0 vyjadřuje buď naprostou absenci měřené vlastnosti, nebo pevný počátek, od něhož měříme velikost znaku. Na rozdíl od intervalové stupnice, při porovnání dvou hodnot na poměrové stupnici, má smysl se ptát, kolikrát je první hodnota větší než druhá. Měření znaků na poměrové škále dovoluje vytěžit maximální objem informací o studovaném jevu. Kvantitativní znaky dovolují použití různých matematicko-statistických operací - např. výpočet průměrů, směrodatné odchylky, korelačních koeficientů, výpočet různých indexů, atd.

Příklady kvantitativních znaků:

věk, tělesná hmotnost, tělesná výška, BMI (Body Mass Index),
průměrná velikost červených krvinek,
hladina celkového cholesterolu,
glykémie (hladina krevního cukru),
hemoglobin, hematokrit, kreatinin,
počet leukocytů, počet erytrocytů, počet trombocytů,
krevní tlak (systolický nebo diastolický),
tepová frekvence,
délka hospitalizace, délka pracovní neschopnosti,
počet zubů stálého chrupu,
počet těhotenství, počet potratů, atd.

Kontrolní otázky:

1. Uveďte charakteristiku kvalitativních dat.
2. Uveďte příklady nominálních dat.
3. Uveďte příklady ordinálních dat.

4. Uved'te příklady metrických (kvantitativních dat).
5. Zařaďte do příslušného typu dat následující znaky:
pohlaví pacienta, tělesná hmotnost (v kg), délka pracovní neschopnosti (ve dnech),
6. Zařaďte do příslušného typu dat následující znaky:
hladina celkového cholesterolu (v mmol/l), BMI (Body Mas Index), reakce po očkování (měřeno na škále: žádná, zarudnutí, otok+horečka).

5. Metody deskriptivní statistiky

Nejčastěji používanou metodou popisné statistiky je **třídění**. Třídění slouží k rozdělení prvků celého zkoumaného souboru do skupin. Skupiny jsou tvořeny podle kategorií daného třídícího znaku (obvykle podle kategorií znaku kvalitativního). Podle počtu třídících znaků rozlišujeme třídění jednostupňové (jeden třídící znak) a třídění víceúrovňové (dva a více třídících znaků). Například třídění novorozenců podle pohlaví (tj. vytvoření dvou skupin: chlapci, dívky) je jednostupňové. Třídění novorozenců podle pohlaví (chlapci / dívky) a termínu porodu (před termínem/ v termínu/ po termínu) je dvouúrovňové (vytvoření 6 skupin). Čím vyšší stupeň třídění provedeme, tím se sice dostáváme ke konkrétnější části celého souboru, ale současně spolu s tím je takto vybraný soubor menší. Vytřídíme-li například ze souboru 1500 respondentů, který je reprezentativní pro celé území České republiky, ženy z Prahy, zůstane z původních 1500 respondentů jen 90 dotázaných.

Chceme-li ke třídění použít znak kvantitativní, musíme nejdříve definovat intervaly hodnot tohoto znaku, podle kterých vytvoříme jednotlivé skupiny. Intervaly musí pokrýt všechny hodnoty znaku a nesmí se vzájemně překrývat.

Výsledkem třídění je **četnostní (frekvenční) tabulka**, která obsahuje:

Absolutní četnosti = počty prvků zařazených do jednotlivých tříd (kategorií nebo intervalů).

Relativní četnosti = absolutní četnosti dělené rozsahem výběru (tj. celkovým počtem prvků).

Kumulativní četnosti = vzniknou postupným načítáním četností z předchozích kategorií k dané kategorii. Kumulativní četnosti mohou být absolutní nebo relativní, počítají se pouze u znaků ordinálních nebo v případě definovaných intervalů pro znak kvantitativní.

Uspořádané absolutní nebo relativní četnosti vytváří **rozdělení výběrové četnosti**.

Příklad:

Četnostní tabulka - rozdělení sledovaných osob do tříd podle věkových intervalů

Třída	Absolutní četnost	Relativní četnost (v %)	Kumulativní absolutní četnost	Kumulativní relativní četnost
15-19 let	140	8,4 %	140	8,4 %
20-24 let	174	10,5 %	314	18,9 %
25-34 let	288	17,4 %	602	36,3 %
35-44 let	269	16,3 %	871	52,6 %
45-54 let	311	18,8 %	1182	71,4 %
55-64 let	209	12,7 %	1391	84,1 %
65 let a více	262	15,9 %	1653	100,0 %
Celkem	1653	100,0 %		

Kumulativní absolutní četnost 602 znamená, že v celém souboru je celkem 602 osob ve věku do 34 let (tj. 34 let a mladších). $602 = 288 + 174 + 140$.

Kumulativní relativní četnost 52,6 % znamená, že v celém souboru je celkem 52,6 % osob ve věku do 44 let (tj. 44 let a mladších). $52,6\% = 16,3\% + 17,4\% + 10,5\% + 8,4\%$.

Prezentace výsledků statistického šetření

je dána slovním výkladem, statistickými tabulkami a grafickým znázorněním. Podle stupně třídění rozeznáváme tři základní typy statistických tabulek:

- tabulku prostou (obsahuje naměřená nezpracovaná data),
- tabulku skupinovou,
- tabulku kombinační.

Tabulka prostá

Tělesná výška v cm								
130	140	136	141	139	133	149	151	139
136	138	142	127	147	139	135	141	143
132	146	151	146	141	141	131	142	141

Tabulka skupinová

Střed třídního intervalu	Počet dětí
125	1
130	3
135	4
140	12
145	4
150	3
Celkem	27

Tabulka kombinační

Střed třídního intervalu	Počet chlapců	Počet dívek	Počet dětí
125	1	-	1
130	1	2	3
135	3	1	4
140	7	5	12
145	1	3	4
150	2	1	3
Celkem	15	12	27

5.1. Grafická prezentace výsledků

Grafické znázornění je velmi účinný způsob, jak prezentovat statistické údaje. Grafy nejsou tak přesné jako tabulky, ale jsou přehlednější a názornější.

Kromě pravoúhlé soustavy souřadnic se používá i polárních souřadnic a v některých případech také nerovnoměrná – logaritmická stupnice.

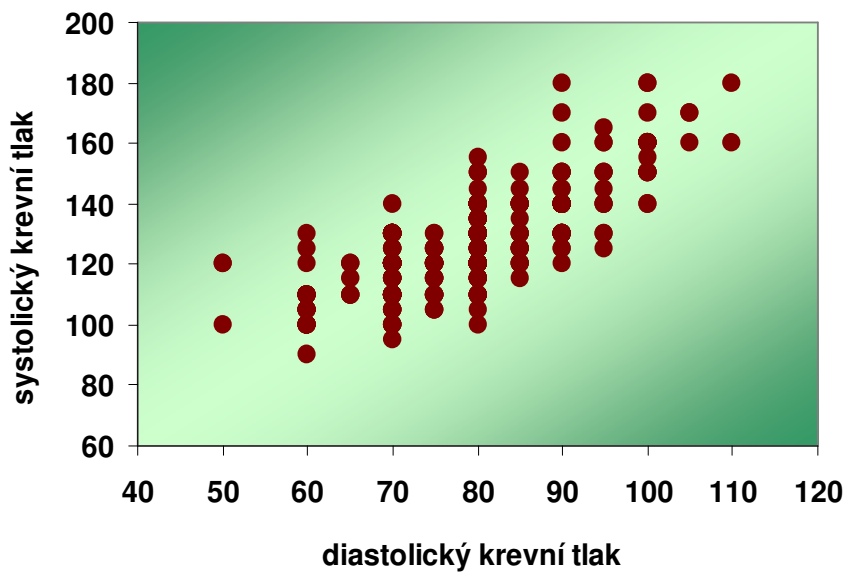
Bodový graf

používá se ke znázornění závislosti dvou kvantitativních znaků. Pro rozlišení hodnot různých skupin (např. muži, ženy) se používají jiné symboly nebo barvy.

Data pro bodový graf:

Pacient č.	Diastolický TK	Systolický TK
1	110	180
2	70	90
3	70	130
4	100	130
...
<i>n</i>	100	160

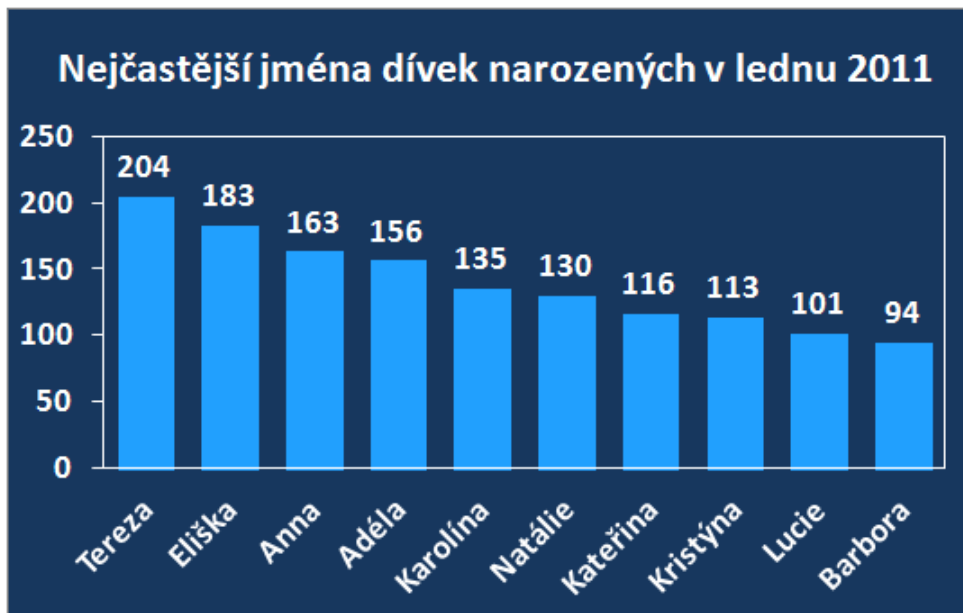
Bodový graf – závislost systolického a diastolického krevního tlaku



Sloupcový graf

používá se ke znázornění absolutních nebo relativních četností kvalitativního znaku, výška sloupců odpovídá četnosti.

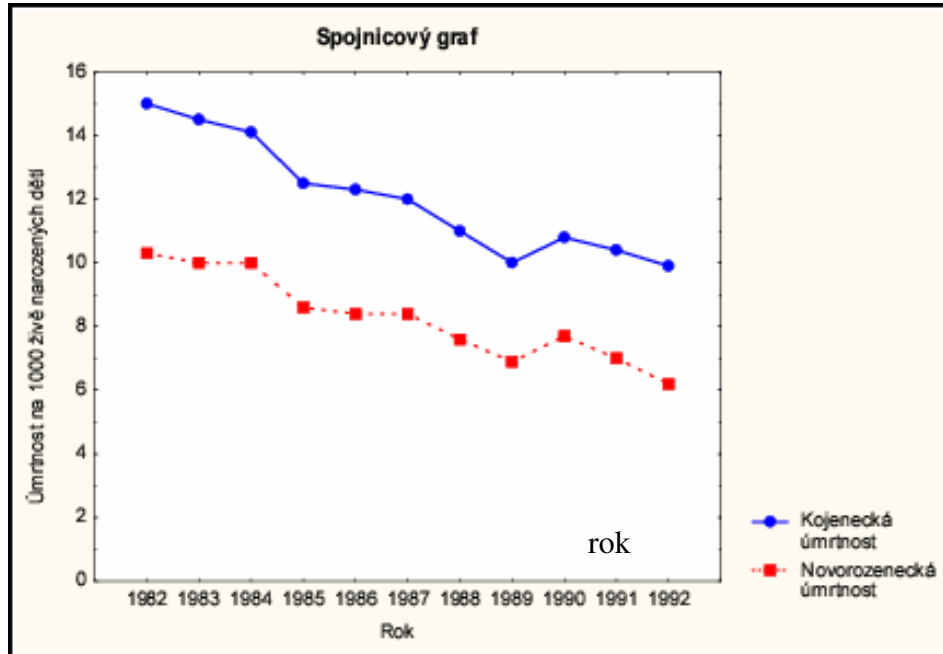
Sloupcový graf – rozdělení počtu narozených děvčat podle křestního jména



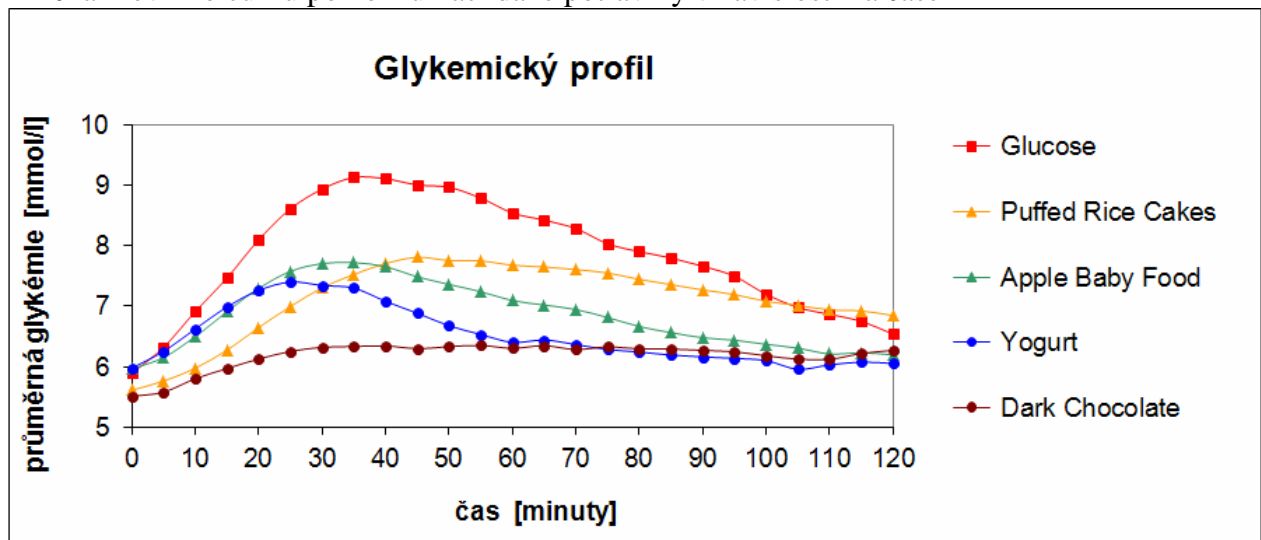
Spojnicový graf

používá se ke znázornění změny hodnot kvantitativního údaje v čase, nebo ke znázornění rozdělení absolutních nebo relativních četností spojitého znaku (v tomto případě se graf nazývá polygon četností).

Spojnicový graf – Kojenecká a novorozenecká úmrtnost na 1 000 živě narozených dětí



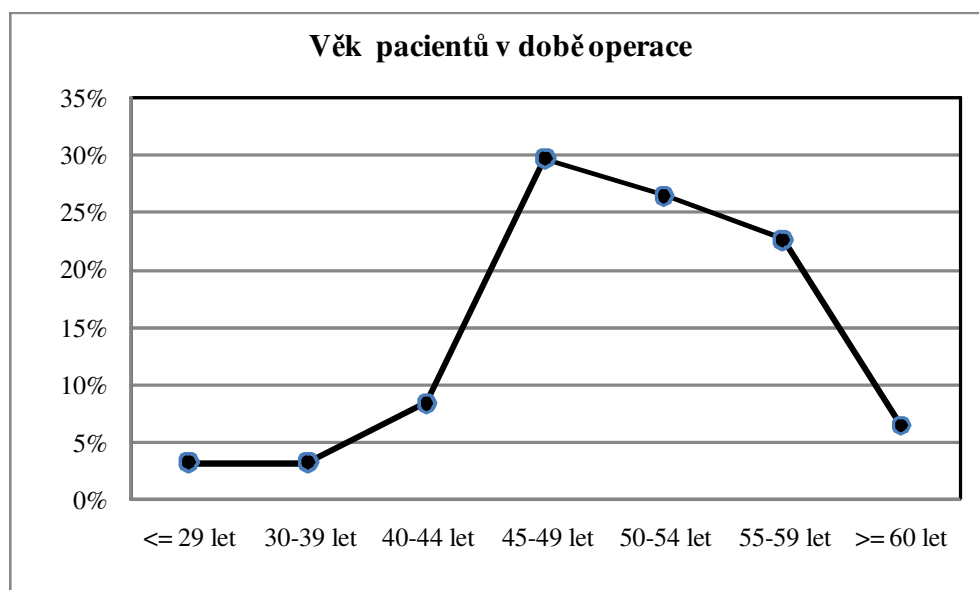
Změna krevního cukru po konzumaci dané potraviny v závislosti na čase



Polygon četností

Tabulka s daty pro polygon četností – distribuce pacientů podle věku v definovaných intervalech.

Věk	Procento pacientů
<= 29 let	3,2%
30-39 let	3,2%
40-44 let	8,4%
45-49 let	29,7%
50-54 let	26,5%
55-59 let	22,6%
>= 60 let	6,5%
Celkem	100%

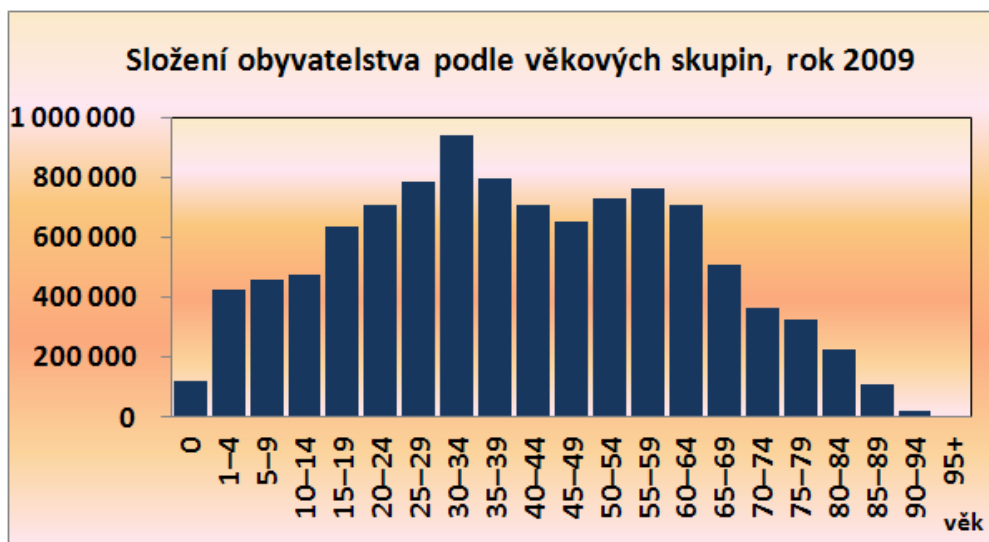


Histogram četností

používá se ke znázornění rozdělení absolutních nebo relativních četností spojitého znaku. Výška sloupců odpovídá četnosti, základna každého sloupce zahrnuje třídu hodnot znaku. Pro účely porovnání se v grafu zobrazují relativní četnosti (procenta).

Složení obyvatelstva podle věkových skupin v roce 2009

věk	počet
0	119 914
1-4	423 278
5-9	460 464
10-14	476 351
15-19	637 248
20-24	707 501
25-29	784 613
30-34	942 504
35-39	794 216
40-44	707 996
45-49	654 498
50-54	730 952
55-59	764 448
60-64	707 407

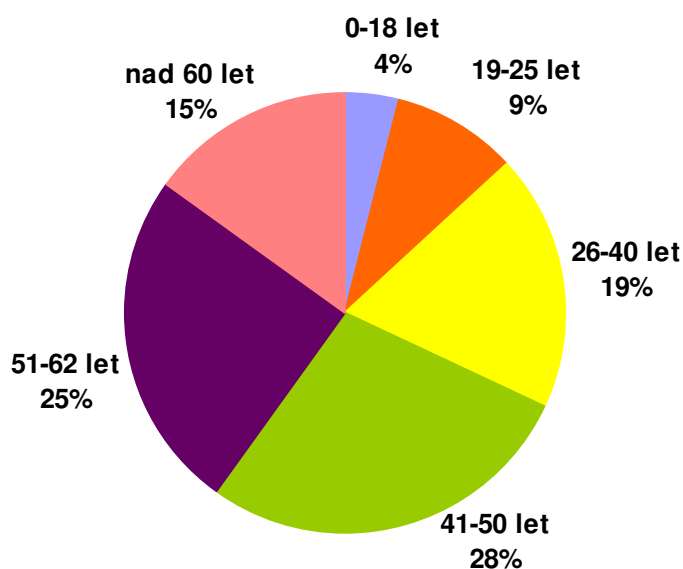


Výšečový (kruhový, koláčový) graf

Zachycuje strukturu souboru, tj. rozdělení celého souboru do jednotlivých skupin podle kategorií daného znaku. Plocha kruhu představuje celý soubor a jednotlivé části jsou znázorněny kruhovými výsečemi.

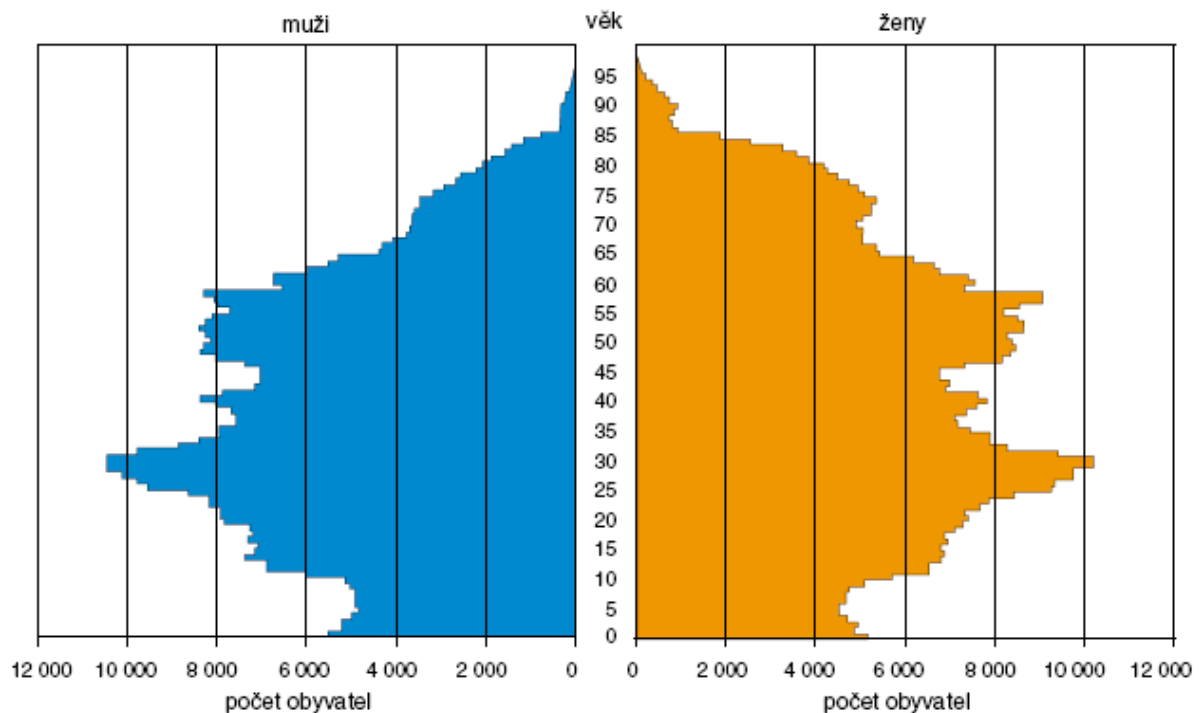
Věk	Procento osob bez domova
0-18 let	4 %
19-25 let	9 %
26-40 let	19 %
41-50 let	28 %
51-62 let	25 %
nad 62 let	15 %

Graf věkové struktury bezdomovců v ČR



Věková pyramida (strom života)

Znázorňuje věkové složení obyvatelstva, je to zvláštní typ sloupcového grafu.



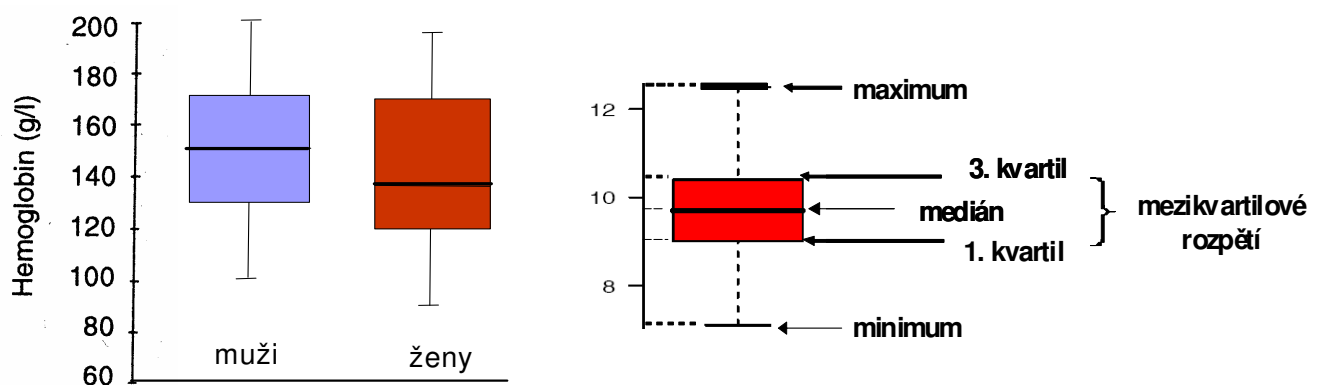
Demografické složení obyvatel v Jihomoravském kraji

Krabicový graf (box graf)

Znázorňuje rozdělení hodnot měřeného kvantitativního znaku.

Krabicový graf nejčastěji zobrazuje tzv. kvartily: medián (tj. 2. kvartil = silná čára uprostřed krabice), 1. kvartil (dno krabice), 3. kvartil (víko krabice) a nejmenší a největší hodnotu (viz. anténky), případně odlehlé nebo extrémní hodnoty (obvykle bývají znázorněny nějakým symbolem, např. kroužkem, hvězdičkou atd.). Výška krabice definuje tzv. mezikvartilové rozpětí. Kvartily budou definovány dále v textu – viz kapitola 12. Popisné statistiky – kvantitativní data.

Porovnání rozložení hemoglobinu u mužů a žen



Kontrolní otázky:

1. Definujte pojmy „frekvenční tabulka“, „absolutní četnost“, „relativní četnost“.
2. Definujte pojmy „kumulativní absolutní četnost“, „kumulativní relativní četnost“.
3. Uveďte, jaký typ grafu by byl vhodný pro grafickou prezentaci závislosti dvou kvantitativních znaků.
4. Jaký typ grafu se používá pro znázornění změny hodnot kvantitativního údaje v závislosti na čase?
5. Definujte pojem „polygon četností“, k čemu se používá?.
6. Definujte pojem „histogram četností“.
7. K čemu se používá box graf (krabicový graf)?

6. Základy teorie pravděpodobnosti

Pravděpodobnost a relativní četnost

Pokus (experiment) se liší od prostého pozorování tím, že při pokusu vědec aktivně ovlivňuje podmínky, za kterých probíhá měření. V praxi se často setkáváme s pokusy, jejichž výsledky nejsou jednoznačně předurčeny podmínkami, za kterých probíhají – takové pokusy nazýváme náhodné pokusy.

(př. házení mincí nebo hrací kostkou, losování z osudí, ...)

Náhodný jev = jakékoli tvrzení o výsledku náhodného pokusu nebo pozorování, o kterém lze po uskutečnění pokusu rozhodnout, zda je či není pravdivé.

Příklad: Náhodný jev A = „narození chlapce“

Budeme-li postupně zaznamenávat pohlaví narozených dětí, dostaneme následující posloupnost:

$A, \neg A, A, \neg A, \neg A, A, A, \neg A,$

kde $\neg A$ je náhodný jev „narození dívky“

Četnost, s jakou nastává náhodný jev A (tj. narození chlapce) pro libovolně dlouhou posloupnost pozorování, můžeme charakterizovat podílem r/n , kde n je délka posloupnosti a r je počet narozených chlapců.

Číslo r nazýváme **absolutní četnost** a podíl r/n **relativní četnost** výskytu náhodného jevu A ve výběru o rozsahu n . S rostoucím rozsahem výběru se relativní četnosti ustalují v blízkosti hodnoty 0,5 (platí pro tento konkrétní příklad).

Každý náhodný jev A je charakterizován číslem $P(A)$, které je mírou častosti výskytu jevu a nazývá se **pravděpodobnost náhodného jevu A** .

6.1. Základní vlastnosti pravděpodobnosti

$$0 \leq P(A) \leq 1$$

$P(A) = 1$, je-li A jev jistý (jev A nastává vždy)

$P(A) = 0$, je-li A jev nemožný (jev A nikdy nenastane).

Při mnohonásobném nezávislém opakování náhodného pokusu se relativní četnost výskytu jevu A jen nepatrně liší od pravděpodobnosti $P(A)$.

$$P(A) = \lim_{n \rightarrow \infty} \frac{r}{n}$$

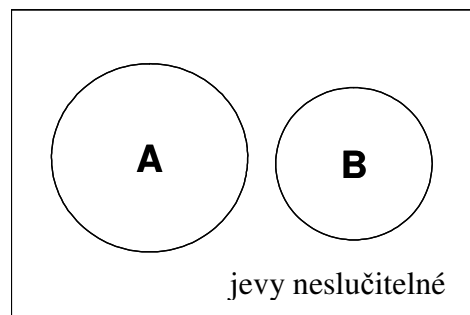
V praxi odhadujeme pravděpodobnost náhodných jevů pomocí relativních četností, kvalita odhadů závisí na počtu provedených pokusů.

6.2. Pravidla pro počítání s pravděpodobnostmi

1. Pravidlo o sčítání pravděpodobností

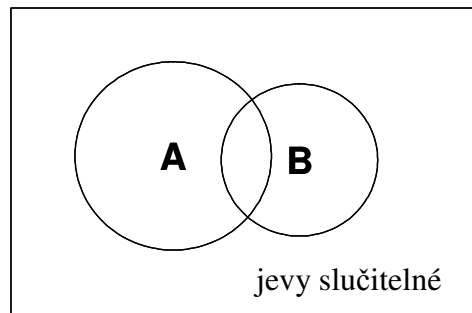
- a) Jsou-li náhodné jevy A, B vzájemně *neslučitelné* (nemohou nastat oba současně), potom pravděpodobnost, že nastane jev A nebo nastane jev B je

$$P(A \cup B) = P(A) + P(B)$$



- b) Jsou-li náhodné jevy A, B *slučitelné* (mohou nastat oba současně), potom

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



- c) Obecně: Jsou-li jevy A_1, A_2, \dots, A_k vzájemně neslučitelné, potom

$$P(A_1 \cup A_2 \cup \dots \cup A_k) = \sum_{i=1}^k P(A_i)$$

6.3. Příklad

V náhodně vybrané skupině $n=140$ mužů ve věku 40-50 let ohrožených arteriální hypertenzí se vyskytl:

- rizikový faktor "**zvýšený cholesterol**" (jev A) ve 37 případech a
- rizikový faktor "**kouření**" (jev B) v 98 případech,
- současný výskyt obou rizikových faktorů ($A \cap B$) byl zjištěn ve 31 případech.

Odhadněte pomocí relativních četností pravděpodobnosti výskytu jevů $A, B, A \cap B$ a $A \cup B$.

Řešení:

$$P(A) = 37/140 = 0,2643$$

Pravděpodobnost, že muž ve věku 40-50 let zjistíme zvýšenou hladinu celkového cholesterolu, je 0,2643.

$$P(B) = 98/140 = 0,7000$$

Pravděpodobnost, že muž ve věku 40-50 let bude kuřák, je 0,7000.

$$P(A \cap B) = 31/140 = 0,2214$$

Pravděpodobnost, že muž ve věku 40-50 let bude kuřák se zvýšenou hladinou celkového cholesterolu, je 0,2214.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,2643 + 0,7000 - 0,2214 = 0,7429$$

(jevy A a B jsou slučitelné)

Pravděpodobnost, že muž ve věku 40-50 let bude buď kuřák, nebo bude mít zvýšenou hladinu celkového cholesterolu, je 0,7429.

2. Pravděpodobnost opačného jevu ($\neg A$) k jevu A:

$$P(\neg A) = 1 - P(A)$$

3. Podmíněná pravděpodobnost

Je-li výskyt jevu A závislý na výskytu jevu B, potom pravděpodobnost, že nastane jev A, nastal-li jev B, je

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

(Jevy A, B jsou *nezávislé*, jestliže výskyt jednoho jevu neovlivňuje výskyt druhého.)

4. Pravidlo o násobení pravděpodobností

Jsou-li jevy A a B *nezávislé*, potom pravděpodobnost, že nastanou oba jevy současně je

$$P(A \cap B) = P(A) \cdot P(B)$$

Obecně $P(A_1 \cap A_2 \cap \dots \cap A_k) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_k)$

5. Pravidlo o úplné pravděpodobnosti

Jsou-li náhodné jevy B_i , $i = 1, \dots, k$ vzájemně neslučitelné a v každém pokusu nastává právě jeden, tj.

$$P(B_1 \cup B_2 \cup \dots \cup B_k) = \sum_{i=1}^k P(B_i) = 1, \text{ a známe-li podmíněné pravděpodobnosti } P(A|B_i),$$

potom

$$P(A) = \sum_{i=1}^k P(A \cap B_i) = \sum_{i=1}^k P(A|B_i) P(B_i)$$

Bayesův vzorec

Udává jakým způsobem vypočítat pravděpodobnosti $P(B_j|A)$, známe-li pravděpodobnosti $P(B_i)$ a $P(A|B_i)$ pro všechny jevy $B_i, i=1, \dots, k$.

$$P(B_j | A) = \frac{P(A | B_j) P(B_j)}{\sum_{i=1}^k P(A | B_i) P(B_i)}$$

6.4. Příklad

použití Bayesova vzorce

Jev	
A	osoba je kuřák
B_1	osoba má chronickou bronchitidu
B_2	osoba nemá chronickou bronchitidu

Pravděpodobnost jevu	
$P(B_1) = 0,40$	pravděpodobnost „výskytu osoby s bronchitidou“ v populaci
$P(B_2) = 0,60$	pravděpodobnost „výskytu osoby bez bronchitidy“ v populaci
$P(A B_1) = 0,75$	pravděpodobnost, že osoba je kuřák ve skupině s chronickou bronchitidou
$P(A B_2) = 0,50$	pravděpodobnost, že osoba je kuřák ve skupině bez chronické bronchitidy

Pravděpodobnost výskytu chronické bronchitidy u kuřáka se vypočítá podle Bayesova vzorce:

$$P(B_1 | A) = \frac{P(A | B_1) P(B_1)}{P(A | B_1) P(B_1) + P(A | B_2) P(B_2)} = \frac{0,75 \cdot 0,40}{0,75 \cdot 0,40 + 0,50 \cdot 0,60} = 0,50$$

Pravděpodobnost výskytu chronické bronchitidy u kuřáka je 0,5.

Kontrolní otázky:

1. Definujte pojmy „náhodný jev“, „pravděpodobnost náhodného jevu“.
2. Definujte základní vlastnosti pravděpodobnosti.
3. Uveďte pravidlo o sčítání pravděpodobnosti.
4. Uveďte pravidlo o násobení pravděpodobnosti.
5. Uveďte pravidlo o úplné pravděpodobnosti.

7. Skrining (screening), diagnostické screeningové testy

Screeningové vyšetření má pomoci při odhalení výskytu nemoci v raném stádiu, tedy nemoci u osob, u kterých zatím nebyla nemoc diagnostikována.

Při screening se používá vhodně zvolený a jednoduchý *screeningový test*, obvykle klinický nebo laboratorní test nebo jiná procedura.

Účelem použití screeningového testu je rozdělit populaci vyšetřovaných osob na dvě skupiny:

- osoby s pozitivním výsledkem testu, které jsou s velkou pravděpodobností nemocné a
- osoby s negativním výsledkem testu, které s velkou pravděpodobností nejsou nemocné.

Diagnostické screeningové testy jsou tedy založeny na předpokladu, že nemocní a zdraví jedinci mohou být pomocí diagnostického testu rozlišeni. Při použití testu k určení diagnózy mohou nastat 4 možné situace:

Při hodnocení přesnosti diagnostického testu porovnáváme jeho výsledky s tzv. „**zlatým standardem**“.

Zlatý standard je referenční metoda pro danou diagnózu. Je to „ideální“ vyšetření, které jednoznačně rozdělí populaci na nemocné a zdravé. Často se jedná o složité a nákladné nebo rizikové vyšetření (biopsii, použití kontrastní zobrazovací metody, pitvu, ...).

Časté chybění zlatého standardu pro určení některých diagnóz může vést k situaci, kdy je jako standard používán test, který není úplně přesný, ale je v současné době považován za nejlepší. Může dojít také k situaci, kdy je srovnáván jeden špatný test s druhým, nebo paradoxně – je nový test prohlášen za špatný i když může být ve skutečnosti lepší než používaný standard.

Postup pro hodnocení screeningového testu:

- každou osobu sledovaného souboru vyšetříme 2 krát:
 - 1. screeningovým testem a
 - 2. náročnější metodou, která dává správné výsledky (např. komplexní klinické vyšetření) – viz. „metoda zlatého standardu“.

výsledky shrneme v tabulce:

		Výsledek podle metody zlatého standardu		
		osoba je nemocná	osoba je zdravá	
Výsledek screeningového testu	Pozitivní	a	b	a+b
	negativní	c	d	c+d
		a+c	b+d	N

$N = a+b+c+d$... celkový počet osob, u nichž byl použit screeningový test.

a ... je počet nemocných osob, u nichž byl test pozitivní (správná pozitivita),

b ... je počet osob bez nemoci, u nichž byl test pozitivní (falešná pozitivita),

c ... je počet nemocných osob, u nichž byl test negativní (falešná negativita),

d ... je počet osob bez nemoci, u nichž byl test negativní (správná negativita).

a+c ... je celkový počet nemocných osob,

b+d ... je celkový počet osob bez nemoci,

a+b ... je celkový počet osob s pozitivním výsledkem testu,
c+d ... je celkový počet osob s negativním výsledkem testu.

7.1. Charakteristiky diagnostického (screeningového) testu

Mezi základní charakteristiky screeningových testů patří:

- Senzitivita = pravděpodobnost, že test bude pozitivní u nemocných.
Senzitivita = $a / (a + c)$
(počet správně pozitivních výsledků screeningového testu / počet nemocných osob)
- Specificita = pravděpodobnost, že test bude negativní u osob bez nemoci.
Specificita = $d / (b + d)$
(počet správně negativních výsledků screeningového testu / počet zdravých osob)
- Přesnost testu = pravděpodobnost, že test dává správné výsledky
Přesnost = $(a + d) / N$
(počet správně pozitivních a správně negativních výsledků screeningového testu / celkový počet osob)
- Pozitivní prediktivní hodnota = pravděpodobnost, že osoba je nemocná, když test byl pozitivní.
PPV = $a / (a + b)$
(počet správně pozitivních výsledků screeningového testu / celkový počet pozitivních testů screeningového testu)
- Negativní prediktivní hodnota = pravděpodobnost, že osoba nemá sledovanou nemoc, když test byl negativní.
NPV = $d / (c + d)$
(počet správně negativních výsledků screeningového testu / celkový počet negativních testů screeningového testu)

Charakteristiky screeningových testů se často vyjadřují v procentech, tzn., že danou pravděpodobnost násobíme číslem 100 %.

Je žádoucí, aby screeningový test byl vysoce senzitivní a vysoce specifický. Toho však v praxi nelze většinou dosáhnout.

V situaci, kdy se pro stanovení diagnózy používá klinický údaj, který má charakter spojitého znaku (např. údaj o hladině daného biochemického parametru), je třeba zvolit kompromis.

V těchto situacích je stanovena tzv. cut-off hodnota (hraniční hodnota), která rozděluje kontinuum hodnot tohoto spojitého znaku na normální a abnormální hodnoty.

7.2. Příklad

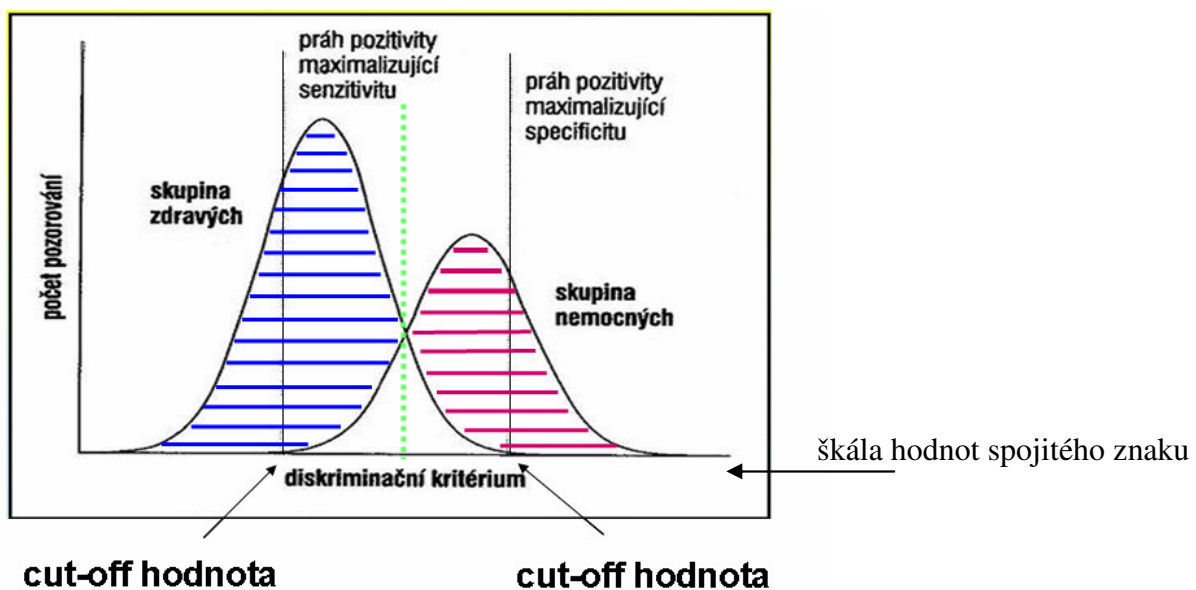
Předpokládejme, že pro danou diagnózu jsou typické vysoké hodnoty určitého biochemického parametru, který má být použit pro screening. Jako optimální cut-off hodnota byla stanovena hodnota 140. Tzn., že:

- je-li u osoby naměřena hodnota vyšší než cut-off hodnota = 140, řekneme, že daná osoba má s velkou pravděpodobností sledovanou nemoc,
- je-li u osoby naměřena hodnota menší než cut-off hodnota = 140, řekneme, že je osoba s velkou pravděpodobností zdravá.

Problém je v tom, že i u části osob, které jsou ve skutečnosti zdravé, naměříme hodnoty vyšší než cut-off 140 a naopak u části osob, které jsou ve skutečnosti nemocné, naměříme hodnoty menší než 140.

Je třeba si uvědomit, že změnou diskriminační hranice (cut-off hodnoty) se mění také vzájemný poměr senzitivity a specificity testu:

Na obrázku je znázorněno rozdělení hodnot biochemického parametru u osob zdravých a osob nemocných.



Přímky kolmé na osu x znázorňují možnou pozici cut-off hodnoty pro tento parametr. Ideální situace by nastala v případě, kdybychom u všech zdravých osob naměřily hodnotu menší než cut-off hodnota a u všech osob nemocných hodnotu větší než daná cut-off hodnota (tj. distribuce na obrázku by byly odděleny). V tomto případě by byla senzitivita testu, specifita testu i přesnost testu rovna 100 %.

7.3. Faktory ovlivňující prediktivní hodnoty

Lékař v praxi musí rozhodnout, zda pacient má nebo nemá určitou nemoc v závislosti na tom, vyjde-li test pozitivně nebo negativně. Důležitou informací pro toto rozhodnutí poskytují prediktivní hodnoty testu.

Hlavním faktorem, který ovlivňuje prediktivní hodnoty je prevalence (tj. výskyt) nemoci v populaci, která je testem vyšetřována.

Je-li známa prevalence nemoci, počítají se prediktivní hodnoty screeningového testu pomocí vzorců

$$PPV = \frac{SE \cdot prevalence}{SE \cdot prevalence + (1 - SP) \cdot (1 - prevalence)}$$

$$NPV = \frac{SP \cdot (1 - prevalence)}{SP \cdot (1 - prevalence) + (1 - SE) \cdot prevalence}$$

SE ... senzitivita testu, *SP* ... specificita testu

Je zřejmé, že screeningové testy nejsou vhodné pro určení nemoci, která má v populaci velmi malou prevalenci.

7.4. Příklad:

Použijeme-li test s 95% senzitivitou a 95% specificitou k předpovědi nemoci, která má prevalenci 1%, potom pozitivní prediktivní hodnota PPV=0,16. To znamená, že pouze 16% osob, u kterých byl test pozitivní, je skutečně nemocných a 84% nemoc nemá, i když test vyšel pozitivně.

Výpočet:

Senzitivita *SE* = 0,95,

specificita *SP* = 0,95,

prevalence = 0,01

Pozitivní prediktivní hodnota PPV = $0,95 \cdot 0,01 / (0,95 \cdot 0,01 + 0,05 \cdot 0,99) = 0,161$.

7.5. Příklady screeningových programů

- Prenatální screening vrozených vývojových vad u plodu (např. diagnostika Downova syndromu). Screening je založen na ultrazvukovém vyšetření plodu, nebo krevních testech. Jako invazivní metoda je používáno vyšetření plodové vody (amniocentéza).
- Mamografický screening pro záchyt karcinomu prsu. Screening je založen na klinickém vyšetření v prsní poradně a následně na mamografickém vyšetření a ultrazvukovém vyšetření.
- Screening karcinomu děložního hrdla / děložního čípku založen na mikroskopickém vyšetření cytologického stěru.
- Screening kolorektálního karcinomu je založen na testu na okultní krvácení ve stolici, případně kolonoskopickém vyšetření tlustého střeva.
- Screening karcinomu prostaty založen na klinickém vyšetření a zjištění hladiny PSA (prostatického specifického antigenu, hladina > 4,0 indikuje biopsii prostaty).
- screening hypertenze (realizuje se měřením krevního tlaku – běžně prováděno v ordinacích obvodních lékařů).

7.6. Příklad:

Výpočet základních charakteristik screeningového testu

Screeningový test byl použit k diagnostice určité nemoci. Výsledek screeningu byl pozitivní u 50 osob skutečně nemocných a u 8 osob zdravých (podle metody zlatého standardu). Podle zlatého standardu byl celkový počet skutečně nemocných osob 60 a celkový počet osob zdravých 90.

Určete senzitivitu, specificitu, přesnost, pozitivní a negativní prediktivní hodnotu screeningového testu.

Řešení:

1. Nejdříve vytvoříme tabulku, do které doplníme známé údaje, tj. počet osob s pozitivním výsledkem screeningového testu ve skupině nemocných a ve skupině zdravých a celkový počet nemocných a zdravých osob.
2. Dopočítáme zbývající údaje v tabulce.

	metoda „zlatého standardu“		Celkem
	Osoba je nemocná	Osoba je zdravá	
Pozitivní výsledek testu	50	8	58
Negativní výsledek testu	10	82	92
Celkem	60	90	150

3. Podle vzorců vypočítáme všechny charakteristiky screeningového testu

Výsledky

	pravděpodobnost	procento
Senzitivita	$50/60 = 0,8333$	83,3 %
Specificita	$82/90 = 0,9111$	91,1 %
Přesnost	$(50+82)/150 = 0,8800$	88,0 %
Pozitivní prediktivní hodnota	$50/58 = 0,8621$	86,2 %
Negativní prediktivní hodnota	$82/92 = 0,8913$	89,1 %

Všechny parametry screeningového testu jsou vyšší než 80%.

Senzitivita testu je 0,833 (resp. 83,3%), to znamená, že test vyjde pozitivně u 83,3% skutečně nemocných osob a u 16,7% osob dává test falešně negativní výsledek.

Specificita testu je 0,911 (resp. 91,1%), to znamená, že test vyjde negativně u 91,1% osob skutečně zdravých a u 8,9% zdravých osob dává test falešně pozitivní výsledek.

Přesnost testu je 0,880 (resp. 88,0%) tzn., že test dává správné výsledky (je správně pozitivní nebo správně negativní) u 88% osob a u 12% osob dává test chybný výsledek.

Pozitivní prediktivní hodnota testu je 0,862 (86,2%). To znamená, že 86,2% osob, u kterých byl test pozitivní, je skutečně nemocných a 13,8% nemoc nemá, i když test vyšel pozitivně.

Negativní prediktivní hodnota testu je 0,891 (89,1%). To znamená, že 89,1% osob, u kterých byl test negativní, je skutečně zdravých a 10,9% je nemocných, i když test vyšel negativně.

Kontrolní otázky:

1. Definujte pojem „screeningový test“.
2. Uveďte základní charakteristiky screeningových testů.
3. Uveďte konkrétní příklady screeningových programů.
4. Senzitivitu a specificitu screeningového testu podle zadaných údajů:
5. Výsledek screeningu byl negativní u 20 osob skutečně nemocných a u 160 osob zdravých (podle metody zlatého standardu). Celkový počet skutečně nemocných osob (podle zlatého standardu) byl 140 a celkový počet osob zdravých byl 190.

8. Náhodná veličina a rozdělení pravděpodobnosti

Za náhodnou veličinu považujeme proměnnou, jejíž hodnota je jednoznačně určena výsledkem náhodného pokusu. Hodnotu náhodné veličiny není možné jednoznačně určit před provedením pokusu.

Jak často určité hodnoty náhodné veličiny nastávají lze popsat pomocí rozdělení pravděpodobnosti.

Rozdělení pravděpodobnosti je předpis, který určuje vztah mezi možnými hodnotami náhodné veličiny x_i a jim příslušejícími pravděpodobnostmi $p_i = P(X = x_i)$.

Rozdělení pravděpodobnosti bývá zadáno tabulkou:

Hodnota náhodné veličiny x_i	x_1	x_2	...	x_n
Pravděpodobnost p_i	p_1	p_2	...	p_n

Platí $\sum p_i = 1$. Počet možných hodnot (n) může být konečný nebo nekonečný.

8.1. Typy náhodných veličin

Podle toho, jakých hodnot náhodné veličiny nabývají, je dělíme na:

- 1) **Diskrétní náhodné veličiny** – jsou definovány pouze v jednotlivých (diskrétních) hodnotách na omezeném nebo neomezeném intervalu.

Popisují se pomocí:

- rozdělení pravděpodobnosti,
- distribuční funkce $F(x)$.

- 2) **Spojitě náhodné veličiny** – jsou definovány na všech možných hodnotách v intervalu $-\infty$ až $+\infty$, nebo na omezeném intervalu a, b .

Popisují se pomocí:

- rozdělení pravděpodobnosti – tzv. **hustoty pravděpodobnosti** $f(x)$,
- distribuční funkce $F(x)$.

Distribuční funkce je kumulativní pravděpodobnost. Pro každé x udává pravděpodobnost s jakou náhodná veličina X nabude hodnot menších nebo rovno x

$$F(x) = P(X \leq x)$$

Pro diskrétní veličinu lze distribuční funkci zapsat jako součet pravděpodobností:

$$F(x) = \sum_{x_i \leq x} P(x_i)$$

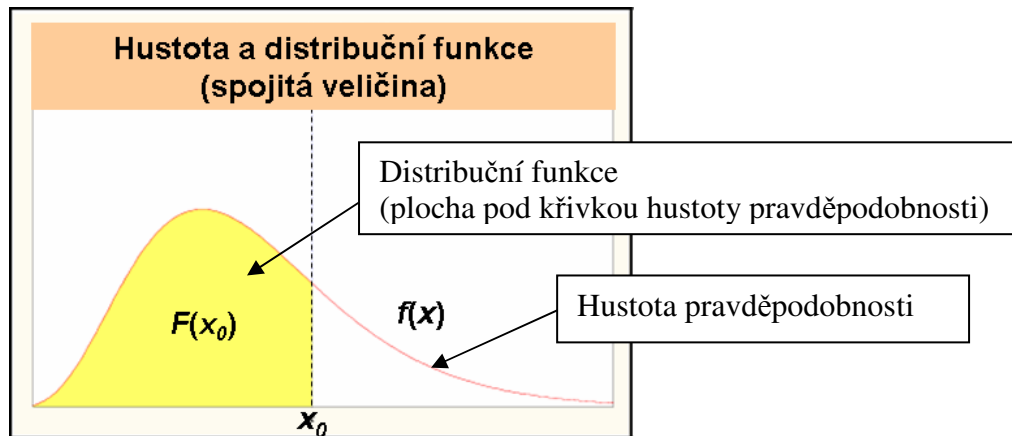
$$P(x_1 \leq X < x_2) = F(x_2) - F(x_1)$$

Pro spojitou veličinu platí:

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

$$P(a \leq X < b) = F(b) - F(a).$$

Graficky lze vztah mezi hustotou pravděpodobnosti a distribuční funkcí zobrazit následovně:



8.2. Základní typy rozdělení pravděpodobnosti

Rozdělení pro diskrétní náhodnou veličinu:

- Rovnoměrné,
- Binomické,
- Poissonovo,
- Alternativní,
- Hypergeometrické, ...

Rozdělení pro spojitou náhodnou veličinu:

- Normální,
- Exponenciální,
- Rovnoměrné,
- Studentovo,
- Weibullovo,
- Fischer-Snedecorovo,
- χ^2 rozdělení (chí-kvadrát rozdělení), ...

8.3. Modely rozdělení pravděpodobnosti pro diskrétní veličiny

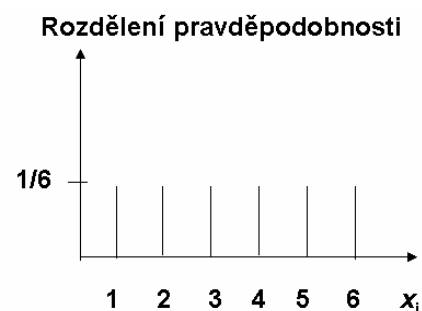
Rovnoměrné rozdělení

nejjednodušší případ diskrétního rozdělení, přiřazuje všem hodnotám náhodné veličiny stejnou pravděpodobnost

$$P(X = x_i) = 1/k, \quad i = 1, 2, \dots, k$$

Příklad

Hod hrací kostkou – pravděpodobnost padnutí každého z čísel je stejná, tj. $1/6$).



Binomické rozdělení

Uvažujme n nezávislých pokusů, v každém pokusu může sledovaný jev buď nastat s pravděpodobností p nebo nenastat s pravděpodobností $1 - p$.

Pravděpodobnost, že jev nastane právě v k pokusech z n , se vypočítá podle vzorce

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \text{ pro } k = 0, 1, 2, \dots, n$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad n! = n(n-1)(n-2)\dots 1$$

Příklad

Předpokládejme, že pravděpodobnost narození dívky je 0,49. Jaká je pravděpodobnost, že mezi třemi dětmi v rodině je právě jedna dívka?

$$n = 3, \quad k = 1, \quad p = 0,49 \quad P(X = 1) = \binom{3}{1} 0,49^1 \cdot 0,51^2 = 3 \cdot 0,127 = 0,38$$

Pravděpodobnost, že mezi třemi dětmi v rodině je právě jedna dívka, je $p = 0,38$.

Jaká je pravděpodobnost, že mezi třemi dětmi v rodině jsou právě dvě dívky?

$$P(X = 2) = \binom{3}{2} 0,49^2 \cdot 0,51^1 = 3 \cdot 0,2401 \cdot 0,51 = 0,367$$

Jaká je pravděpodobnost, že mezi třemi dětmi v rodině není žádná dívka?

$$P(X = 0) = \binom{3}{0} 0,49^0 \cdot 0,51^3 = 1 \cdot 1 \cdot 0,51^3 = 0,133$$

Příklad - Pravděpodobnostní rozložení počtu chlapců v rodině se čtyřmi dětmi.

Řešení:

Je-li pravděpodobnost narození dívky $p = 0,49$, potom pravděpodobnost narození chlapce $p = 1 - 0,49 = 0,51$.

V rodině se 4 dětmi může být celkový počet chlapců: 0, 1, 2, 3 nebo 4.

Pravděpodobnost, že v rodině se 4 dětmi není žádný chlapec:

$$n = 4, \quad k = 0, \quad \text{pravděpodobnost narození chlapce } p = 0,51$$

$$P(X = 0) = \binom{4}{0} 0,51^0 (1 - 0,51)^4 = 1 \cdot 1 \cdot 0,49^4 = 0,0576$$

$$\binom{4}{0} = \frac{4!}{0!4!} = 1$$

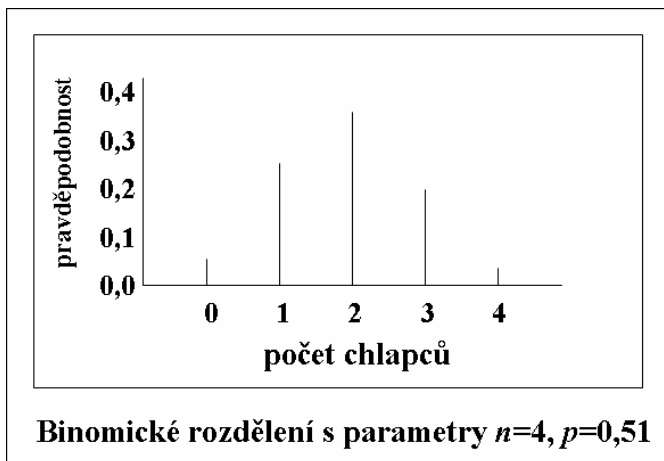
Pravděpodobnost, že v rodině se 4 dětmi není žádný chlapec je: $P(X = 0 \text{ chlapců}) = 0,058$

Pravděpodobnost, že v rodině se 4 dětmi je 1 chlapec je: $P(X = 1 \text{ chlapec}) = 0,24$.

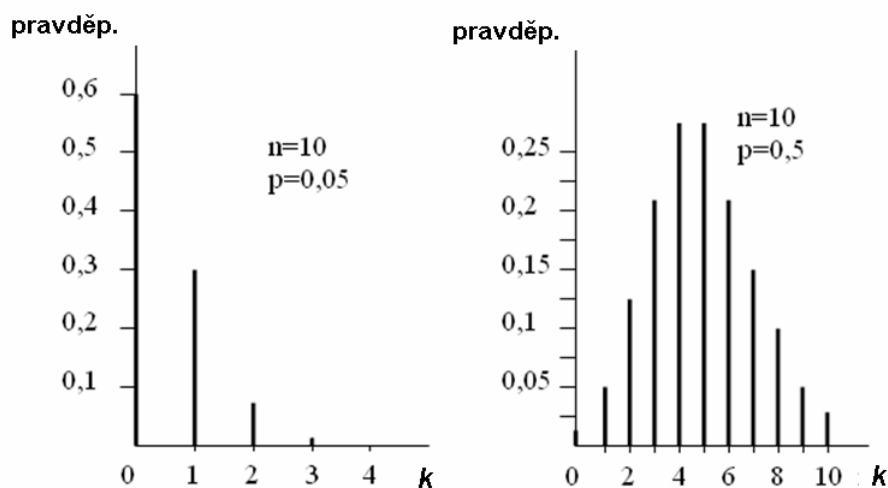
Pravděpodobnost, že v rodině se 4 dětmi jsou 2 chlapci je: $P(X = 2 \text{ chlapci}) = 0,37$.

Pravděpodobnost, že v rodině se 4 dětmi jsou 3 chlapci je: $P(X = 3 \text{ chlapci}) = 0,26$

Pravděpodobnost, že v rodině se 4 dětmi jsou 4 chlapci je: $P(X = 4 \text{ chlapci}) = 0,07$



Pro $p = 0,5$ je binomické rozdělení symetrické, pro jiné hodnoty p je asymetrické



Průběhy diskretní hustoty pravděpodobnosti pro binomické rozdělení

Poissonovo rozdělení

Používá se pro jevy, které mají malou pravděpodobnost výskytu (“vzácné jevy” – např. mutace genu, počet nehod za danou jednotku času, počet pacientů přijatých při noční službě na chirurgickém oddělení, atd.).

Uvažujme náhodnou veličinu X , která představuje počet výskytů vzácné události v daném časovém intervalu. X může nabývat celočíselných hodnot od 0 do $+\infty$. Pravděpodobnost, že vzácných událostí nastane v daném čase právě k , je

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

kde λ je konstanta, která označuje průměrný počet událostí (např. v čase), $e = 2,718$ je matematická konstanta.

Příklad – použití Poissonova rozdělení

Uvažujme jako „vzácnou událost“ hospitalizaci pacienta s ischemickou chorobou srdeční (ICHS) po chirurgickém zákroku.

Sledujeme celkem 200 pacientů s ICHS a zajímá nás pravděpodobnostní rozdělení počtu hospitalizací u těchto pacientů v průběhu 3 let.

Víme, že průměrný počet hospitalizací na jednoho pacienta: $320 / 200 = 1,6 = \lambda$.

Uvažujme $k = 0, 1, 2, 3, 4, 5$ hospitalizací.

Dosažením do vzorce za $\lambda = 1,6$ a $k = 0, 1, 2, \dots, 5$, postupně dostaneme následující pravděpodobnosti:

Pravděpodobnost, že během 3 let nedojde k hospitalizaci vůbec je:

$P(X = 0 \text{ hospitalizací}) = 0,20$.

$$P(X = 0) = \frac{e^{-1,6} \cdot 1,6^0}{0!} = e^{-1,6} = 0,20.$$

Pravděpodobnost, že během 3 let dojde k hospitalizaci jednou je:

$P(X = 1 \text{ hospitalizace}) = 0,32$.

$$P(X = 1) = \frac{e^{-1,6} \cdot 1,6^1}{1!} = e^{-1,6} \cdot 1,6 = 0,32.$$

Pravděpodobnost, že během 3 let dojde k hospitalizaci 2krát je:

$P(X = 2 \text{ hospitalizace}) = 0,26$.

Pravděpodobnost, že během 3 let dojde k hospitalizaci 3krát je:

$P(X = 3 \text{ hospitalizace}) = 0,14$.

Pravděpodobnost, že během 3 let dojde k hospitalizaci 4krát je:

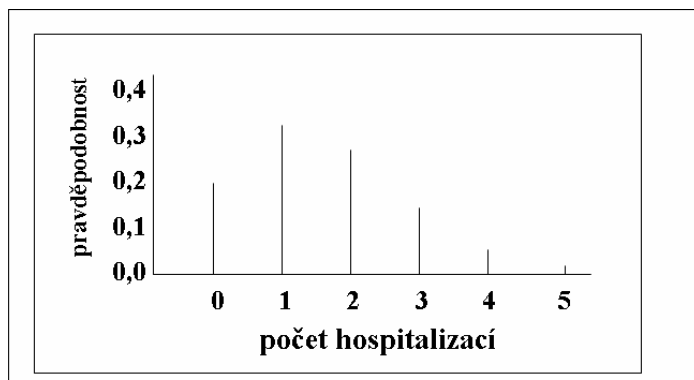
$P(X = 4 \text{ hospitalizace}) = 0,06$.

Pravděpodobnost, že během 3 let dojde k hospitalizaci 5krát je:

$P(X = 5 \text{ hospitalizací}) = 0,02$.

$$P(X = 5) = \frac{e^{-1,6} \cdot 1,6^5}{5!} = \frac{e^{-1,6} \cdot 1,6^5}{120} = 0,02.$$

Rozdělení pravděpodobnosti počtu hospitalizací



8.1. Model rozdělení pro spojité veličiny

Normální (Gaussovo) rozdělení

S normálním rozdělením se setkáváme u veličin popisujících např. výsledky biochemického vyšetření.

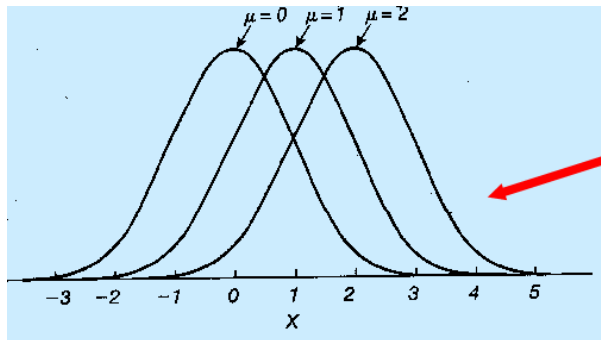
Hustota pravděpodobnosti náhodné veličiny X má tvar

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

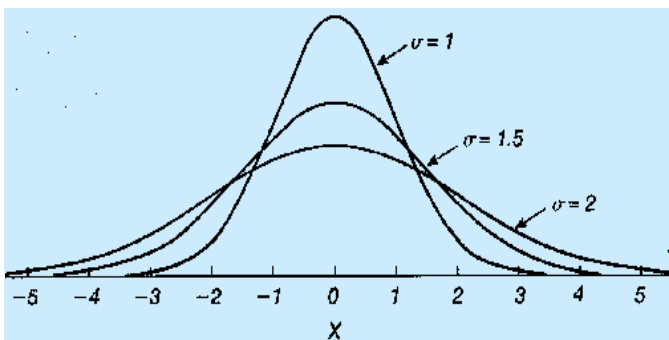
$\pi = 3,141$ a $e = 2,718$ jsou matematické konstanty, $x \in (-\infty, +\infty)$.

μ a σ jsou *parametry* určující polohu křivky na ose X a její tvar (roztážení podél osy).

Graf hustoty pravděpodobnosti normálního rozdělení



při různém μ a stejném σ



při stejném μ a různém σ

Pro různé μ a stejné σ jsou křivky totožné, liší se pouze posunutím. Pro stejné μ a různé σ jsou křivky odlišné, vrchol všech křivek však leží nad stejnou hodnotou (μ).

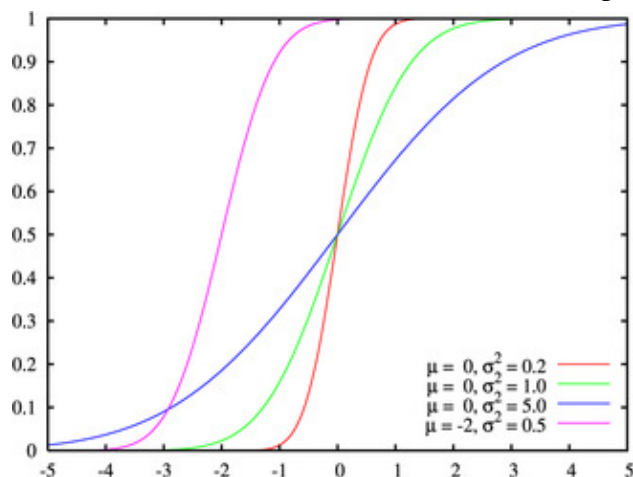
Hustota normálního rozdělení je symetrická jednovrcholová křivka zvonovitého charakteru, která nikde neprotíná osu X . Parametr μ (populační průměrná hodnota) určuje polohu vrcholu, parametr σ (populační směrodatná odchylka hodnot) určuje roztážení křivky podél osy.

Plocha pod křivkou hustoty normálního rozdělení je rovna jedné.

Distribuční funkce normálního rozdělení veličiny X

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Graf distribuční funkce normálního rozdělení pro stejný parametr $\mu (=0)$ a různý parametr σ .



Normované normální rozdělení

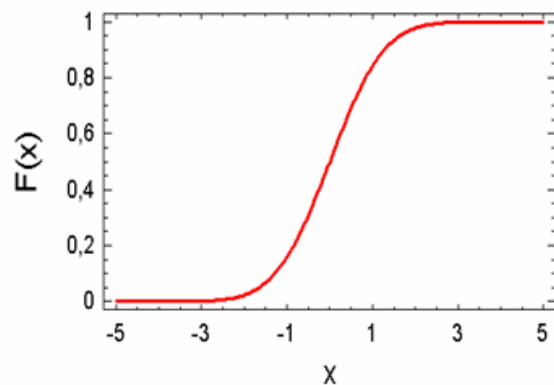
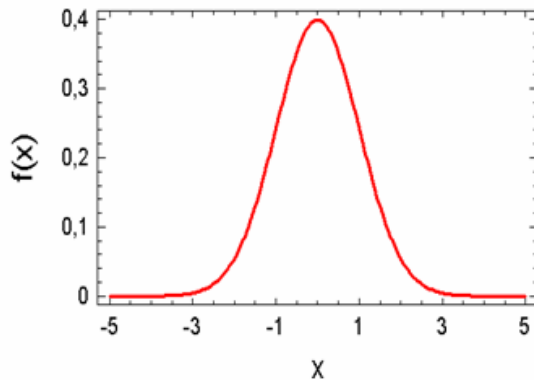
Je to rozdělení pro standardizovanou náhodnou veličinu $Z = \frac{X-\mu}{\sigma}$.

Je definováno hustotou pravděpodobnosti

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

kde $x \in (-\infty, +\infty)$.

Parametry rozdělení normovaného normálního rozdělení $\mu = 0$ a $\sigma = 1$



Vlastnosti normální distribuce

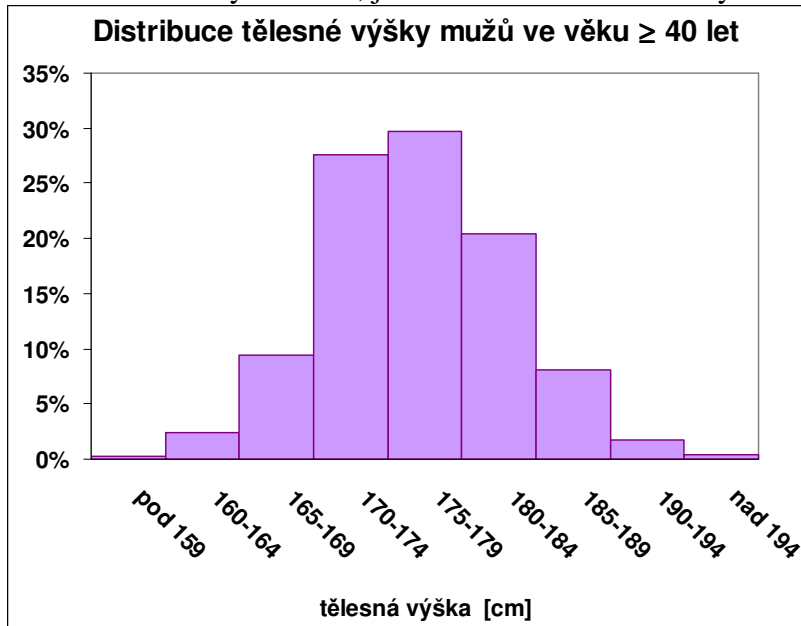
Podle tvaru distribuce rozeznáváme, zda je distribuce symetrická nebo asymetrická.

Většina distribucí v medicíně je symetrická. Je-li asymetrická, jde většinou o distribuci šikmou doprava. Většina distribucí je také jednovrcholová (unimodální), distribuce vícevrcholová je většinou způsobena nehomogenitou dat.

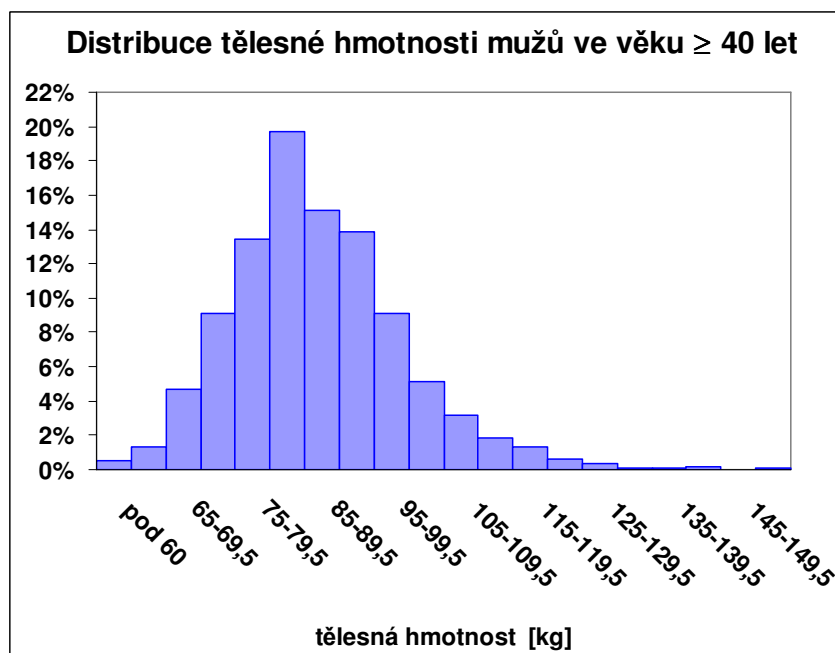
Příklad

Distribuce utonulých podle věku je dvouvrcholová – 1. vrchol: batolata a děti do 4 let věku, 2. vrchol: dospívající (výskyt v dospělosti je méně častý).

Příklad: Distribuce symetrická, jednovrcholová – tělesná výška



Příklad: Distribuce šikmá doprava – tělesná hmotnost



Pro normální rozložení hodnot měřeného kvantitativního znaku dále platí:

- a) 68,3% hodnot leží v intervalu $\mu \pm 1\sigma$
- b) 95,4% hodnot leží v intervalu $\mu \pm 2\sigma$
- c) 99,7% hodnot leží v intervalu $\mu \pm 3\sigma$.

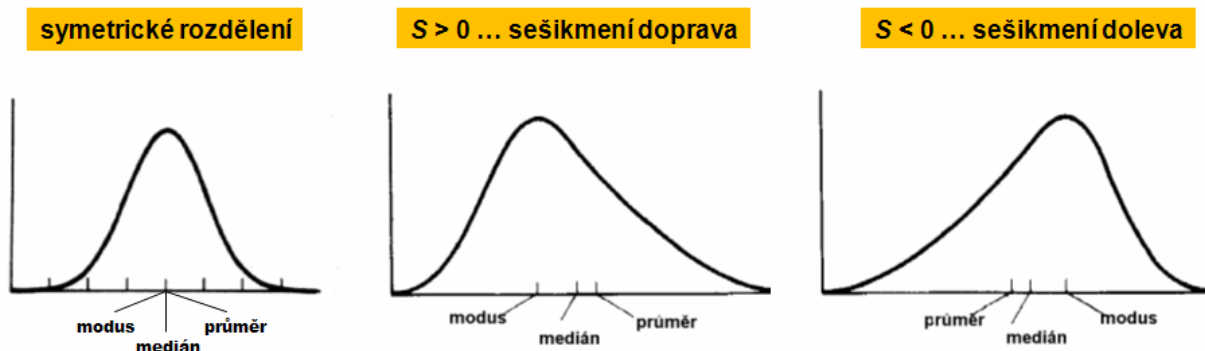
Koeficient šikmosti

je míra symetrie kolem střední hodnoty, počítá se podle vzorce

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

Hodnoty koeficientu kolem 0 znamenají, že je distribuce symetrická. Hodnoty koeficientu > 0 znamenají distribuci sešikmenou doprava, hodnoty koeficientu < 0 znamenají sešikmení doleva.

Pozice průměru a mediánu u symetrické distribuce a sešikmených distribucí

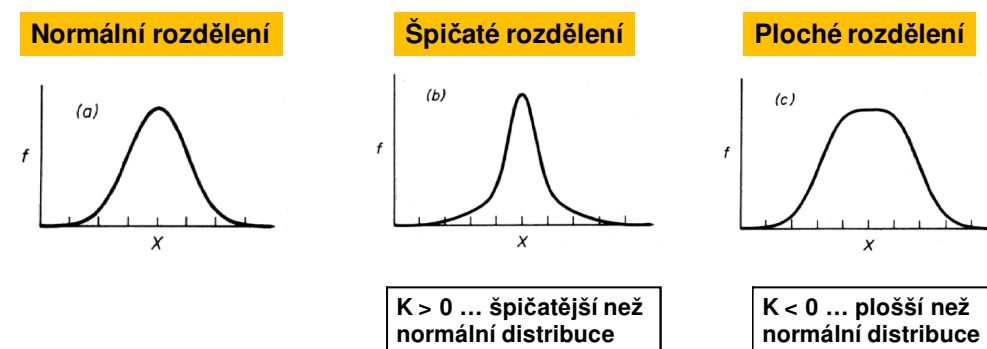


Koeficient špičatosti

Je míra koncentrace kolem střední hodnoty, počítá se podle vzorce

$$K = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})^4}{s^4} - 3$$

Hodnoty koeficientu kolem 0 znamenají, že je distribuce normální. Hodnoty koeficientu > 0 znamenají, že je distribuce špičatější než normální distribuce, hodnoty koeficientu < 0 znamenají, že je distribuce plošší než normální distribuce.



Exponenciální rozdělení

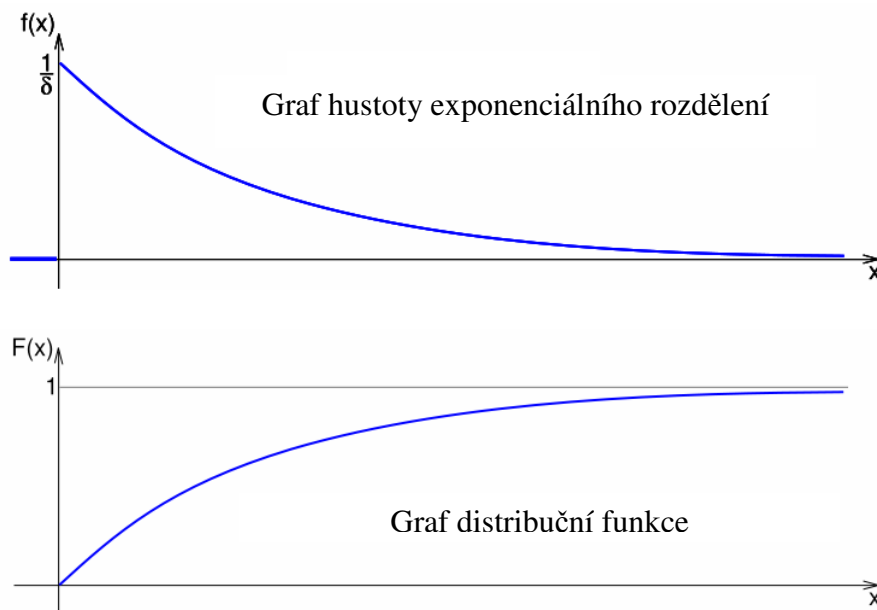
Exponenciální rozdělení pravděpodobnosti se používá pro vyjádření doby čekání na výskyt nějakého jevu. Hustota pravděpodobnosti exponenciálního rozdělení má tvar

$$f(x) = \begin{cases} 0 & \text{pro } x \leq 0 \\ \frac{1}{\delta} e^{-\frac{x}{\delta}} & \text{pro } x > 0 \end{cases}$$

Distribuční funkce exponenciálního rozdělení

$$F(x) = \begin{cases} 0 & \text{pro } x \leq 0 \\ 1 - e^{-\frac{x}{\delta}} & \text{pro } x > 0 \end{cases}$$

parametr δ je střední doba čekání na výskyt jevu.



Kontrolní otázky:

1. Definujte pojmy „náhodná veličina“, „rozdělení pravděpodobnosti náhodné veličiny“.
2. Definujte základní typy náhodných veličin a vyjmenujte známé typy rozdělení pravděpodobnosti náhodných veličin.
3. Uveďte příklad použití binomické pravděpodobnosti.
4. Jaké parametry určují tvar normální distribuce.
5. Jaký vztah je mezi průměrnou hodnotou a mediánem u doprava zešikmené distribuce.

9. Popisné charakteristiky – kvalitativní data

Četnostní (frekvenční) tabulka

Základním popisným nástrojem pro data kvalitativní je četnostní (frekvenční) tabulka, která popisuje rozdělení (distribuci) četností v jednotlivých kategoriích kvalitativního znaku.

Příklad – Kvalita spolupráce s pacientem při vyšetření

(uvažujeme 4 kategorie charakterizující kvalitu spolupráce s pacientem při vyšetření: „výborná“, „velmi dobrá“, „průměrná“, „spíše s obtížemi“).

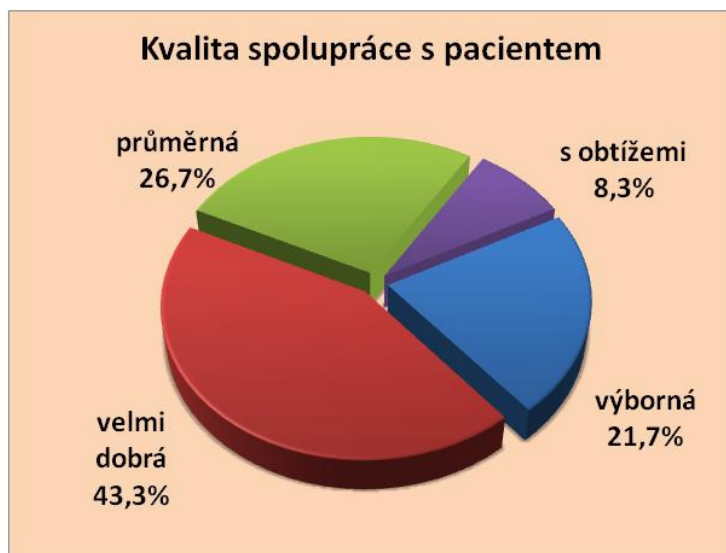
Kvalita spolupráce

	Četnost	Procenta
výborná	13	21,7
velmi dobrá	26	43,3
průměrná	16	26,7
spíše s obtížemi	5	8,3
Celkem	60	100,0

Četnost (absolutní četnost) je počet pacientů zařazených do dané kategorie.

Procenta (relativní četnost) je počet pacientů v dané kategorii vztažený k celkovému počtu pacientů. V tabulce je relativní četnost vyjádřena v procentech (je vynásobena 100). Zobrazujeme-li údaje v četnostní tabulce graficky, používáme zpravidla relativní četnosti v procentech.

Graf popisující distribuci četností v jednotlivých kategoriích měřeného znaku



Kontingenční tabulka – popis rozdělení četností podle kombinace kategorií dvou kvalitativních znaků

Příklad – Rozdělení souboru pacientů podle pohlaví a kouření.

Kontingenční tabulka popisuje počet a procento kuřáků a nekuřáků ve skupině mužů a žen

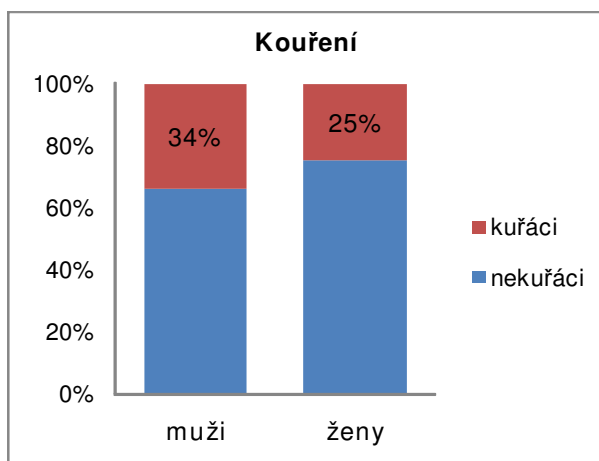
	Kuřáci počet (%)	Nekuřáci počet (%)	Celkem počet (%)
Muži	17 (34 %)	33 (66 %)	50 (100 %)
Ženy	19 (25 %)	56 (75 %)	75 (100 %)
Celkem	36 (29 %)	89 (71 %)	125 (100 %)

V souboru je 125 osob, 50 mužů a 75 žen.

Celkem je v souboru 36 kuřáků (tj. 29 %) a 89 nekuřáků (tj. 71 %).

Ve skupině mužů je 17 kuřáků, tj. 34 % a ve skupině žen je 19 kuřáček, tj. 25 %.

Graficky lze podíl kuřáků ve skupině mužů a ve skupině žen zobrazit pomocí skládaného sloupcového grafu. V grafu se opět používají procenta, aby bylo možné obě skupiny porovnat.



10. Popisné charakteristiky – kvantitativní data

Četnostní tabulky

K základnímu popisu kvantitativních dat můžeme, podobně jako u kvalitativních dat, použít četnostní tabulku. Nemá však význam zjišťovat četnosti pro každou naměřenou hodnotu zvlášť, je třeba definovat intervaly hodnot, ve kterých budeme četnosti zjišťovat. Intervaly se nesmí překrývat a musí pokrýt všechny hodnoty kvantitativního znaku.

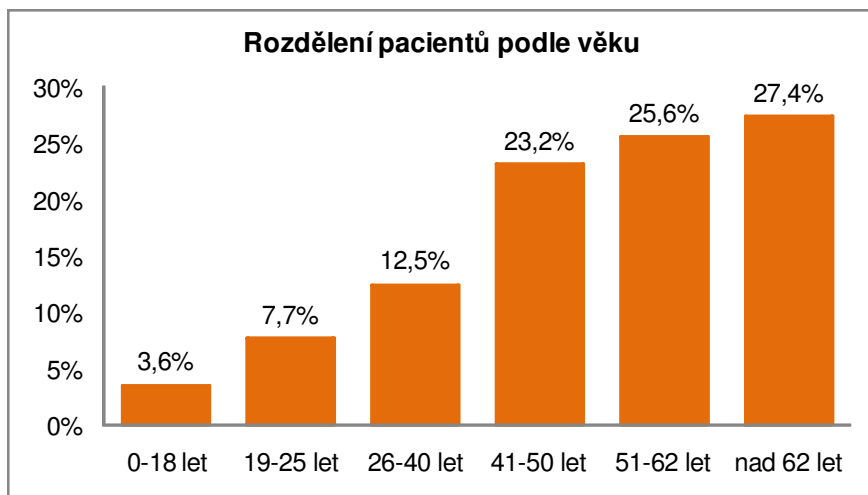
U kvantitativních znaků a znaků kvalitativních ordinálních lze počítat **kumulativní četnosti** absolutní a relativní. Kumulativní četnost se vypočítá jako součet četností z předchozích třídních intervalů. Absolutní kumulativní četnost posledního intervalu je tedy rovna velikosti výběrového souboru a jeho relativní kumulativní četnost je rovna 100 %.

Příklad

Rozdělení pacientů podle věkových intervalů popisuje následující četnostní tabulka:

Věk	Počet	Procento	Kumulativní počet	Kumulativní procento
≤ 18 let	12	3,6%	12	3,6%
19-25 let	26	7,7%	38	11,3%
26-40 let	42	12,5%	80	23,8%
41-50 let	78	23,2%	158	47,0%
51-62 let	86	25,6%	244	72,6%
nad 62 let	92	27,4%	336	100,0%
Celkem	336	100,0%		

Kumulativní relativní četnost 47 % udává, že ve sledovaném souboru je 47 % pacientů ve věku do 50 let (výpočet viz kapitola 5. Metody deskriptivní statistiky).



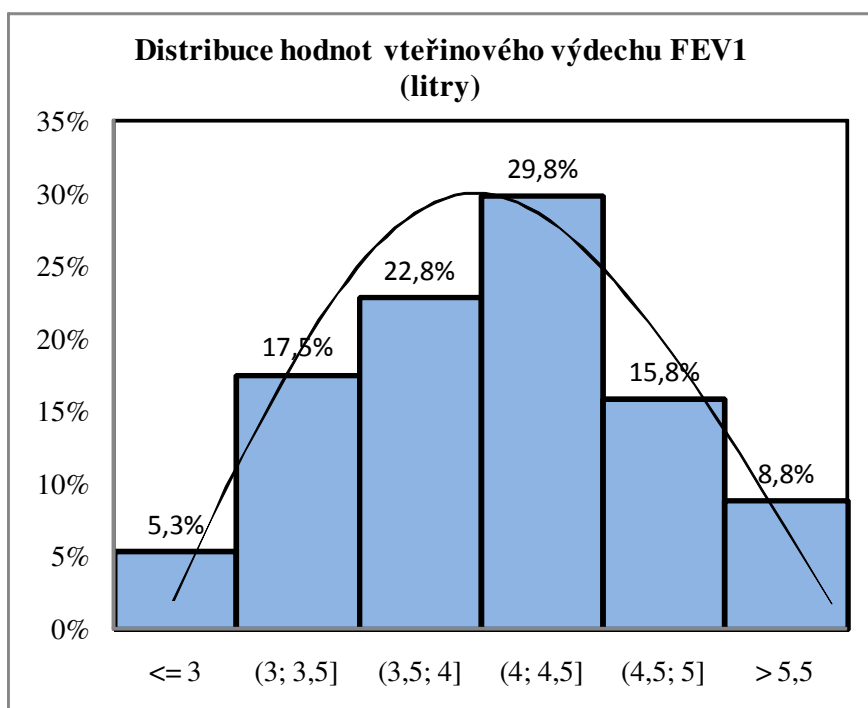
Relativní četnost výskytu dané varianty (hodnoty) znaku lze chápat jako **pravděpodobnost výskytu** této varianty a rozdělení četností hodnot tedy můžeme chápat jako **rozdělení pravděpodobností** výskytu hodnot. Názornou představu o rozložení četností hodnot daného znaku nám **poskytne grafické zobrazení distribuce**.

Pro zobrazení distribuce četností jsou vhodné sloupcové grafy (pro znaky kvantitativní) nebo koláčové grafy (pro znaky kvalitativní).

Distribuci hodnot kvantitativního znaku lze graficky zobrazit také pomocí **histogramu četností** a můžeme ji popsat pomocí nezáporné funkce $f(x)$, která se nazývá hustota rozdělení (viz dříve).

Histogram četností

Zobrazení distribuce hodnot znaku vteřinový výdech FEV1 (tj. objem vzduchu vydechnutý za 1 sekundu) s křivkou hustoty normálního rozdělení. Objem vydechnutého vzduchu je měřen v litrech (viz. jednotka na ose X).



10.1. Popisné statistiky polohy a variability dat

V souboru naměřených hodnot (výsledků pozorování) potřebujeme často určit hodnotu, kolem které se data soustřeďují. K tomu slouží řada charakteristik, které se nazývají **míry polohy**.

Míry polohy ovšem nic nevyovídají o variabilitě (různosti) dat, proto je vhodné grafické metody znázorňující variabilitu dat (histogram, polygon četností) doplnit číselnými charakteristikami popisujícími např. kolísání dat kolem průměru – **míry variability**.

CHARAKTERISTIKY POLOHY DAT

Existuje celá řada charakteristik, které lze použít k popisu polohy, nebo také střední hodnoty určitého souboru dat.

Aritmetický průměr

Nejběžněji používaná charakteristika polohy dat, určuje polohu distribuce naměřených hodnot na číselné ose (viz dříve). Není vhodnou charakteristikou pro znaky, jejichž distribuce je výrazně nesymetrická nebo vícevrcholová. Aritmetický průměr se počítá podle vzorce:

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

n je počet naměřených hodnot, x_i je hodnota naměřená u i -tého pacienta.

(Symbolem \sum (čti suma) značíme součet všech hodnot, tj, $\sum x_i = x_1 + x_2 + \dots + x_n$).

Průměr je ovlivněn extrémními hodnotami, tj. hodnotami velmi nízkými nebo vysokými vzhledem k ostatním naměřeným hodnotám. V těchto případech je vhodnější použít jako charakteristiku polohy medián hodnot.

Aritmetický průměr se používá pouze u dat kvantitativních.

Medián

Medián je prostřední hodnota v řadě hodnot uspořádaných podle velikosti. Rozděluje řadu s uspořádanými hodnotami na dvě poloviny – hodnoty v první polovině jsou menší než medián, hodnoty ve druhé polovině jsou větší než medián. Máme-li sudý počet hodnot, vypočte se medián jako průměr dvou prostředních hodnot.

Medián se počítá u dat kvantitativních a kvalitativních ordinálních.

Příklad 1: Zjistěte medián věku souboru 7 pacientů: 57, 49, 55, 54, 92, 62 a 65 let.

Řešení: Hodnoty věku seřadíme podle velikosti. Dostaneme posloupnost čísel:

49	54	55	57	62	65	92
----	----	----	----	----	----	----

Mediánem je hodnota 57 let (průměrná hodnota je 62 let). V datech se vyskytuje extrémní hodnota věku 92 let, proto je v tomto případě vhodnější použít jako míru polohy dat medián.

Příklad 2 – Zjistěte medián hodnot celkového cholesterolu naměřeného u 6 pacientů:

6,2	7,5	5,5	4,9	8,2	5,8
-----	-----	-----	-----	-----	-----

Řešení: Naměřené hodnoty seřadíme podle velikosti:

4,9	5,5	5,8	6,2	7,5	8,2
-----	-----	-----	-----	-----	-----

Medián vypočítáme jako průměr hodnot 5,8 a 6,2, tj medián = $(5,8 + 6,2) / 2 = 6,0$ mmol/l.

Příklad 3 – Zjistěte medián hodnocení testu studentů.

Studenti byli hodnoceni na ordinální škále A, B, C, D, E (A je nejlepší hodnocení, E je nejhorší hodnocení).

Student č.	1	2	3	4	5	6	7	8	9	10	11
hodnocení	B	B	C	A	D	E	A	E	C	C	D

Medián můžeme zjistit také u kvalitativních ordinálních znaků. Výsledky hodnocení studentů seřadíme, získáme řadu: A, A, B, B, C, C, C, D, D, E, E. Medián je roven hodnocení C.

Modus

Je málo používanou mírou polohy pro kvantitativní data, je definován jako hodnota, která se mezi všemi naměřenými hodnotami vyskytuje nejčastěji. Modus není vhodnou charakteristikou při malém počtu pozorování. Z definice vyplývá, že modálních hodnot může být i více. Dvojice modálních hodnot může poukazovat na tzv. bimodalitu rozdělení (na histogramu bychom viděli dva vrcholy). Modus lze zjistit také pro znaky kvalitativní, je užitečný zejména při popisu nominálních dat.

Příklad: Mějme soubor 190 pacientů, u nichž jsme zjistili krevní skupinu. Dostaneme následující výsledky:

krevní skupina	A	B	AB	0
počet pacientů	100	60	10	20

Modus je krevní skupina A, která se vyskytuje nejčastěji.

CHARAKTERISTIKY VARIABILITY DAT

Variační rozpětí

Variační rozpětí je velmi hrubá míra variability dat, vypočítá se jako rozdíl maximální a minimální naměřené hodnoty. $R = x_{max} - x_{min}$

Příklad: Zjistěte variační rozpětí věku 10 pacientů: 21, 48, 27, 52, 68, 56, 72, 61, 54, 63 let. Variační rozpětí je: $72 - 21 = 51$ let.

Směrodatná odchylka (SD)

Je nejběžněji používaná charakteristika variability dat, určuje tvar hustoty rozdělení (viz normální distribuce). Lze ji počítat jen pro data kvantitativní.

U normálního rozdělení platí - čím je odchylka větší, tím je křivka hustoty plošší, čím je odchylka menší, tím je křivka hustoty strmější. Směrodatná odchylka je obvykle označována zkratkou SD (z anglického standard deviation), počítá se podle vzorce

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

kde n je počet naměřených hodnot, \bar{x} je průměrná hodnota znaku, x_i jsou hodnoty naměřené u jednotlivých pacientů. Výraz pod odmocninou se nazývá **výběrový rozptyl** (je to průměrná odchylka naměřených hodnot od průměru).

Variační koeficient

je vhodný pro porovnání variability dvou znaků s velmi odlišnými průměry nebo pro porovnání variability veličin měřených v různých jednotkách, vyjadřuje variabilitu v procentech.

$$V = \frac{s}{\bar{x}} \cdot 100 \%, \text{ kde } s \text{ je směrodatná odchylka a } \bar{x} \text{ je průměrná hodnota znaku.}$$

Mezikvartilové rozpětí

Je vhodnou charakteristikou variability v případě nenormální distribuce dat, resp. v případě přítomnosti extrémních hodnot v datech. Počítá se jako rozdíl 3. kvartil - 1. kvartil.

Kvantily, kvartily, decily, percentily

Obecně P%-ní kvantil x_p je číslo, které odděluje z distribuce hodnot daného znaku P% nejmenších hodnot. Medián je tedy 50% kvantil (rozděluje distribuci, tj. uspořádanou řadu naměřených hodnot, na poloviny).

Kvantily, které rozdělují distribuci na 4 části po 25%, se nazývají **kvartily** (2. kvartil je medián). Kvantily, které rozdělují řadu na 10 částí (po 10%), se nazývají **decily** a po 1% **percentily**.

Percentily se používají pro stanovení normálních hodnot pro laboratorní nálezy, interval normálních hodnot se zpravidla stanovuje jako interval mezi 2,5 a 97,5 percentilem.

Box graf
zobrazuje distribuci
hodnot
kvantitativního
znaku pomocí
kvartilů.



Standardní chyba průměru (SE)

není charakteristikou variability dat, je to charakteristika přesnosti odhadu průměru. SE klesá s velikostí výběru, vypočítá se podle vzorce:

$$SE = \frac{s}{\sqrt{n}}$$

Entropie

používá se jako charakteristika variability nominálních dat, počítá se pomocí vzorce:

$$H = - \sum_{i=1}^k \frac{n_i}{n} \ln \frac{n_i}{n}$$

k je počet kategorií nominálního znaku, n_i/n je relativní četnost i -té kategorie..

Entropie nabývá minimální hodnoty 0 v případě, kdy je 100% případů zařazeno do jedné kategorie, maximální hodnota je dosažena při rovnoměrném rozdělení četností mezi kategorie.

Příklad

Zjistěte variabilitu výskytu krevních skupin zjištěných u $n = 190$ pacientů.

Krevní skupina	A	B	AB	0
počet pacientů (n_i)	100	60	10	20
Relativní četnost (n_i/n)	0,526	0,316	0,053	0,105

$$H = -\sum_{i=1}^4 \frac{n_i}{n} \ln \frac{n_i}{n} =$$

$$1 = -(0,526 \cdot \ln(0,526) + 0,316 \cdot \ln(0,316) + 0,053 \cdot \ln(0,053) + 0,105 \cdot \ln(0,105)) =$$

$$J = 1,094.$$

sobené chybami (měření, překlepy při zadávání dat do počítače). Jindy bývá příčinou skutečnost, že do sledovaného vzorku byl zařazen pacient, jehož příslušnost k ostatním lze z určitých důvodů zpochybnit. Pokud příčinou není chyba, ale jistá "odlišnost" pacienta, pak se v praxi často postupuje tak, že se statistická analýza provede dvakrát - poprvé se všemi naměřenými údaji a podruhé bez odlehle hodnoty a zváží se důsledky případné odlišnosti výsledků. Publikovat je třeba oba výsledky a nejen ten, který lépe splňuje naše očekávání!

Příklad: Popište věk pacientů pomocí základních statistických charakteristik.

Tabulka s daty

pacient č.	1	2	3	4	5	6	7	8	9	10
Věk	27	29	42	36	53	56	43	25	28	51

Aritmetický průměr

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{10} (27 + 29 + 42 + 36 + 53 + 56 + 43 + 25 + 28 + 51) = \frac{1}{10} 390 = 39 \text{ let}$$

Průměrný věk pacientů je 39 let (minimální věk je 25 let, maximální věk je 56 let).

Variační rozpětí

maximální hodnota – minimální hodnota = $56 - 25 = 31$ let

Medián věku

Hodnoty věku uspořádáme podle velikosti

Věk	25	27	28	29	36	42	43	51	53	56
------------	----	----	----	----	----	----	----	----	----	----

Medián je definován jako prostřední hodnota v řadě. Protože má řada sudý počet hodnot, vypočítá se medián jako průměr dvou hodnot, které leží uprostřed řady, tj.:

Medián = $(36+42)/2=39$ let.

Směrodatná odchylka

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad \bar{x} = 39, \quad n = 10$$

Pac. č.	Věk	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2$
1	27	$(27-39)^2$	144
2	29	$(29-39)^2$	100

3	42	$(42-39)^2$	9
4	36	$(36-39)^2$	9
5	53	$(53-39)^2$	196
6	56	$(56-39)^2$	289
7	43	$(43-39)^2$	16
8	25	$(25-39)^2$	196
9	28	$(28-39)^2$	121
10	51	$(51-39)^2$	144

$$\sum_{i=1}^n (x_i - \bar{x})^2 = 144 + 100 + 9 + 9 + 196 + 289 + 16 + 196 + 121 + 144 = 1224$$

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{10-1} 1224} = \sqrt{\frac{1224}{9}} = \sqrt{136} = 11,66 \text{ let}$$

Směrodatná odchylka věku je 11,7 let.

Variační koeficient

$$V = \frac{SD}{\bar{x}} \cdot 100\% = \frac{11,66}{39} \cdot 100 = 29,9\%$$

Variační koeficient je 30 %.

Kontrolní otázky:

1. Definujte pojem „kontingenční tabulka“, uveďte konkrétní příklad.
2. Vyjmenujte základní charakteristiky polohy kvantitativních dat.
3. Definujte základní charakteristiky variability kvantitativních dat.
4. Rozhodněte, zda je následující tvrzení pravdivé: „3. kvartil odděluje 25% menších hodnot z distribuce kvantitativního znaku“.

11. Základy statistických odhadů a testování hypotéz

11.1. Bodové a intervalové odhady charakteristik populace

Populační charakteristiky (parametry populace) jsou obecně neznámá čísla (např. průměrná výška v populaci dospělých mužů).

Parametry populace můžeme odhadnout prostřednictvím bodového nebo intervalového odhadu na základě výběru z populace.

Bodové odhady populačních parametrů

Nejjednodušší formou odhadu populačních parametrů jsou odhady bodové, jejichž výsledkem je jedno číslo. Bodovými odhady populačních charakteristik jsou výběrové charakteristiky. Bodovým odhadem populačního průměru je výběrový průměr, tj. průměr vypočítaný z dat výběrového souboru. Bodový odhad bude tím přesnější, čím větší bude náhodný výběr. Chceme-li však odhadnout populační parametr a současně vyjádřit míru nejistoty, která je s tímto odhadem spojena, musíme odhad vyjádřit intervalem.

Následující tabulka uvádí přehled nepoužívanějších parametrů populace a jejich bodových odhadů pomocí odpovídajících výběrových charakteristik:

Tabulka: Přehled populačních charakteristik a jejich bodových odhadů

Základní soubor (populace)	střední hodnota μ	rozptyl σ^2	směrodatná odchylka σ	relativní četnost π
Výběrový soubor (výběr)	výběrový průměr \bar{x}	výběrový rozptyl s^2	výběrová směrodatná odchylka s	výběrová relativní četnost p

Intervalové odhady populačních parametrů

Intervalový odhad, na rozdíl od odhadu bodového, poukazuje na variabilitu, kterou by odhad získal v jiných náhodných výběrech. Pokud bychom totiž z dané populace pořídili jiný náhodný výběr a vypočítali výběrové charakteristiky, dostali bychom jiné odhady populačních charakteristik. Postup konstrukce intervalového odhadu populačního parametru spočívá v tom, že kolem bodového odhadu stanovíme interval, který s předem zvolenou pravděpodobností obsahuje neznámý parametr. Intervalovým odhadům populačních parametrů říkáme **intervaly spolehlivosti** nebo **konfidenční intervaly** (v odborné literatuře se používá pro označení konfidenčního intervalu zkratka CI – Confidence Interval).

11.2. Interval spolehlivosti pro populační průměr

Při odhadu intervalu spolehlivosti pro populační průměr μ se využívá skutečnosti, že výběrové průměry mají normální rozdělení s průměrem μ a odchylkou $\frac{\sigma}{\sqrt{n}}$.

(Symbolem μ se označuje populační průměr kvantitativního znaku a symbolem σ se označuje populační směrodatná odchylka znaku).

Provedeme-li náhodný výběr o rozsahu n a vypočítáme výběrový průměr, potom populační průměr μ leží s pravděpodobností 0,95 ve vzdálenosti menší než $1,96 \frac{\sigma}{\sqrt{n}}$ od průměru \bar{x} .

Hodnota 1,96 je kritická hodnota standardizovaného normálního rozdělení pro koeficient spolehlivosti 0,95. Populační směrodatnou odchylku σ většinou neznáme a pro výpočty se používá pouze její odhad pomocí výběrové směrodatné odchylky s .

95% interval spolehlivosti pro průměr (95% CI) lze vypočítat podle vzorce:

$$\left(\bar{x} - 1,96 \frac{s}{\sqrt{n}}, \quad \bar{x} + 1,96 \frac{s}{\sqrt{n}} \right)$$

Tento vzorec s konstantou 1,96 je vhodnější používat pouze pro velké výběry ($n \geq 100$). Čím menší výběr máme k dispozici, tím vyšší hodnotu musíme místo čísla 1,96 do vzorce dosadit. Souvisí to s tím, že v případě malých výběrů jsou odhady populační směrodatné odchylky σ pomocí výběrové odchylky s velmi nepřesné. Přesněji lze 95% interval spolehlivosti pro průměr určit, dosadíme-li u malých výběrů ($n < 100$) místo hodnoty 1,96 hodnotu tzv. kvantilu Studentova t -rozdělení pro $n-1$ stupňů volnosti, který je označován symbolem $t_{0,975}$. Hodnota kvantilu se vyhledá ve statistických tabulkách pro požadovanou spolehlivost a daný počet stupňů volnosti $df = n-1$. K výpočtu levé a pravé meze intervalu spolehlivosti potom použijeme vzorec:

$$\left(\bar{x} - t_{0,975} \frac{s}{\sqrt{n}}, \quad \bar{x} + t_{0,975} \frac{s}{\sqrt{n}} \right)$$

Chceme-li zkonstruovat 90 % nebo 99 % interval spolehlivosti, stačí nahradit číslo 1,96 ve vzorci příslušným kvantilem standardizovaného normálního rozdělení. Pro 90 % interval spolehlivosti se používá číslo 1,645, pro 99 % interval se používá číslo 2,576. Vyžadujeme-li vyšší spolehlivost, vyjde interval širší než v případě menší spolehlivosti.

Tabulka s hodnotami kvantilů Studentova t rozdělení

Tabulka A.3: Kvantily $t_{1-\frac{\alpha}{2}}(df)$
Studentova t rozdělení o df
stupních volnosti

$t_{1-\frac{\alpha}{2}}(df)$	α		
df	0,05	0,01	0,001
1	12,706	63,656	636,578
2	4,303	9,925	31,600
3	3,182	5,841	12,924
4	2,776	4,604	8,610
5	2,571	4,032	6,869
6	2,447	3,707	5,959
7	2,365	3,499	5,408
8	2,306	3,355	5,041
9	2,262	3,250	4,781
10	2,228	3,169	4,587
11	2,201	3,106	4,437
12	2,179	3,055	4,318
13	2,160	3,012	4,221
14	2,145	2,977	4,140
15	2,131	2,947	4,073
16	2,120	2,921	4,015

$t_{1-\frac{\alpha}{2}}(df)$	α		
df	0,05	0,01	0,001
17	2,110	2,898	3,965
18	2,101	2,878	3,922
19	2,093	2,861	3,883
20	2,086	2,845	3,850
21	2,080	2,831	3,819
22	2,074	2,819	3,792
23	2,069	2,807	3,768
24	2,064	2,797	3,745
25	2,060	2,787	3,725
30	2,042	2,750	3,646
35	2,030	2,724	3,591
40	2,021	2,704	3,551
45	2,014	2,690	3,520
50	2,009	2,678	3,496
60	2,000	2,660	3,460
70	1,994	2,648	3,435
80	1,990	2,639	3,416
90	1,987	2,632	3,402
100	1,984	2,626	3,390
∞	1,960	2,576	3,290

11.3. Příklad

Odhadněte pomocí 95% CI (intervalu spolehlivosti) a 99% CI populační průměrný věk pacientek s karcinomem endometria. (Pozn.: endometrium = izolační vrstva dělohy)

V souboru 122 pacientek s karcinomem endometria byl vypočítán průměrný věk pacientek $\bar{x} = 64,49$ let (bodový odhad populačního průměrného věku) a směrodatná odchylka věku $s = 10,89$ let (bodový odhad populační směrodatné odchylky).

Řešení:

Vzhledem k tomu, že $n > 100$ můžeme pro odhad levé a pravé meze intervalu spolehlivosti použít vzorec s hodnotou kvantilu Studentova t rozdělení 1,96.

$n = 100$, $\bar{x} = 64,49$ let, $s = 10,89$ let

po dosazení do vzorců dostáváme

$$\left(\bar{x} - 1,96 \frac{s}{\sqrt{n}}, \bar{x} + 1,96 \frac{s}{\sqrt{n}} \right) = \left(64,49 - 1,96 \frac{10,89}{\sqrt{122}}, 64,49 + 1,96 \frac{10,89}{\sqrt{122}} \right) = (62,6; 66,4)$$

Populační průměrný věk pacientek s karcinomem endometria pokrývá se spolehlivostí 0,95 interval 62,6 let – 66,4 let.

99 % interval spolehlivosti pro věk pacientek:

$$\left(\bar{x} - 2,576 \frac{s}{\sqrt{n}}, \bar{x} + 2,576 \frac{s}{\sqrt{n}} \right) = \left(64,49 - 2,576 \frac{10,89}{\sqrt{122}}, 64,49 + 2,576 \frac{10,89}{\sqrt{122}} \right) = (61,9; 67,0)$$

Populační průměrný věk pacientek s karcinomem endometria pokrývá se spolehlivostí 0,99 interval 61,9 let – 67,0 let.

11.4. Interval spolehlivosti pro populační pravděpodobnost

Obdobným způsobem lze zkonstruovat interval spolehlivosti pro populační pravděpodobnost. Pomocí intervalu spolehlivosti chceme např. odhadnout, s jakou pravděpodobností můžeme v populaci naměřit celkovou hodnotu cholesterolu vyšší než norma 5 mmol/l.

Předpokládejme, že máme k dispozici náhodný výběr o rozsahu n .

Bodovým odhadem populační pravděpodobnosti je relativní četnost výskytu dané vlastnosti ve výběrovém souboru. Je-li osob s danou vlastností (tj. s hladinou celkového cholesterolu > 5 mmol/l) r , potom relativní četnost výskytu této vlastnosti ve výběru je rovna $p = r / n$.

Je-li rozsah výběru $n > 100$ a p není příliš malé, pak 95% interval spolehlivosti pro populační pravděpodobnost π má krajní body

$$p \pm 1,96 \sqrt{\frac{p(1-p)}{n}}$$

11.5. Příklad

Máme náhodný výběr 200 mladých mužů, z toho u 120 mužů jsme naměřili vyšší než doporučenou hladinu celkového cholesterolu v séru (tj. > 5 mmol/l).

Určete 95% interval spolehlivosti pro procento mladých mužů s vyšší hladinou cholesterolu v populaci.

Řešení:

$n = 200$, relativní četnost výskytu vyšší hladiny cholesterolu v uvažovaném výběru je $p = 120/200 = 0,6$. 95% interval spolehlivosti je tedy:

$$p \pm 1,96 \sqrt{\frac{p(1-p)}{n}} = 0,6 \pm 1,96 \sqrt{\frac{0,6(1-0,6)}{200}} = 0,6 \pm 0,07 \Rightarrow (0,53; 0,67)$$

Procento mladých mužů s vyšší než doporučenou hladinou celkového cholesterolu v séru pokrývá se spolehlivostí 0,95 interval (53 %, 67 %).

Poznámka: Vynásobíme-li hodnotu pravděpodobnosti výskytu jevu 100, dostaneme údaj o procentuelním výskytu tohoto jevu.

11.6. Testování hypotéz

Hlavním cílem induktivní statistiky je zobecňování výsledků výzkumu (tj. výsledků získaných z výběrových souborů) na celou populaci. Toto zobecňování se v praxi realizuje prostřednictvím ověřování tvrzení (hypotéz) týkajících se celé populace.

Statistické hypotézy (hypotézy o populaci) můžeme rozdělit do dvou skupin: na hypotézy parametrické a hypotézy neparametrické.

Parametrické hypotézy jsou tvrzení o parametrech populace, např. o střední hodnotě, mediánu, rozptylu, relativní četnosti.

Neparametrické hypotézy jsou tvrzení o jiných vlastnostech populace, např. o tvaru rozdělení nebo závislosti měřených znaků.

Parametrické hypotézy můžeme zapsat jako rovnosti (resp. nerovnosti) mezi

- testovaným parametrem a jeho předpokládanou hodnotou (např. „ $\mu = 100$ “, „ $\pi \leq 0,08$ “)
- nebo mezi testovanými parametry (např. „ $\mu_1 = \mu_2$ “, „ $\pi_1 > \pi_2$ “).

(Symbolem π se označuje relativní četnost výskytu v populaci).

Statistická hypotéza je tvrzení, jehož platnost se snažíme ověřit na základě dat, získaných z náhodného výběru.

Nulová hypotéza (H_0) představuje určitý rovnovážný stav a bývá vyjádřena rovností „ $=$ “ ($H_0: \mu = 100$ - průměrná populační hodnota daného parametru je rovna konstantě 100)

Alternativní hypotéza (H_A) je tvrzení, které s nulovou hypotézou nesouhlasí, představuje porušení rovnovážného stavu a zapisujeme ji tedy jedním ze tří možných zápisů nerovnosti („ \neq “, „ $<$ “, „ $>$ “).

- Zvolíme-li alternativní hypotézu ve tvaru „ $<$ “ nebo „ $>$ “, mluvíme o jednostranné alternativní hypotéze (např. $\mu < 100$, $\mu > 100$),
- Zvolíme-li alternativní hypotézu ve tvaru „ \neq “, mluvíme o oboustranné alternativní hypotéze (např. $\mu \neq 100$)

Platnost statistické hypotézy se prověřuje na základě naměřených dat pomocí statistického testu.

Příklad: Formulace nulové a alternativní hypotézy

H_0 : Kouření nezávisí na pohlaví.

H_A : Kouření závisí na pohlaví.

H_0 : Průměrná hodnota systolického krevního tlaku v populaci zdravých dospělých osob je 120 mmHg.

H_A : Průměrná hodnota systolického krevního tlaku v populaci zdravých dospělých osob se nerovná 120 mmHg.

H_0 : Index krvácivosti PBI u dětí s rovnátky a bez rovnátek se neliší.

H_A : Index krvácivosti PBI u dětí s rovnátky a bez rovnátek se liší.

11.7. Chyba I. a II. druhu

Vzhledem ke skutečnosti, že máme k dispozici data pouze z výběrového souboru, můžeme se při rozhodování o platnosti nulové hypotézy dopustit chyby. Při rozhodování mohou nastat situace, které popisuje následující tabulka:

Tabulka: Možné situace při rozhodování o platnosti nulové hypotézy

Naše rozhodnutí	Skutečnost	
	Ho neplatí	Ho platí
zamítáme Ho	správné rozhodnutí (síla testu $1-\beta$)	chyba I. druhu (α)
nezamítáme Ho	chyba II. druhu (β)	správné rozhodnutí

Chyba I. druhu

Zamítneme-li nulovou hypotézu, která ve skutečnosti platí, dopustíme se chyby I. druhu. Pravděpodobnost, že k takové chybě dojde, nazýváme **hladina významnosti** a označujeme ji symbolem α .

Chyba II. druhu

Nezamítneme-li nulovou hypotézu v případě, kdy platí hypotéza alternativní, dopustíme se chyby II. druhu. Pravděpodobnost této chyby je β .

Při testování hypotéz se snažíme minimalizovat obě chyby, což však není možné, neboť snížením β se zároveň zvýší hladina významnosti α a naopak. Proto je třeba najít kompromis mezi požadavky na α a β .

Platí-li nulová hypotéza a my jsme ji nezamítli, rozhodli jsme správně. Pravděpodobnost tohoto rozhodnutí označujeme $(1-\alpha)$ a nazýváme ji **spolehlivost**. Správným rozhodnutím je rovněž zamítnutí nulové hypotézy v případě, kdy platí hypotéza alternativní. Tohoto rozhodnutí se dopouštíme s pravděpodobností $(1-\beta)$, což bývá označováno jako **síla testu**.

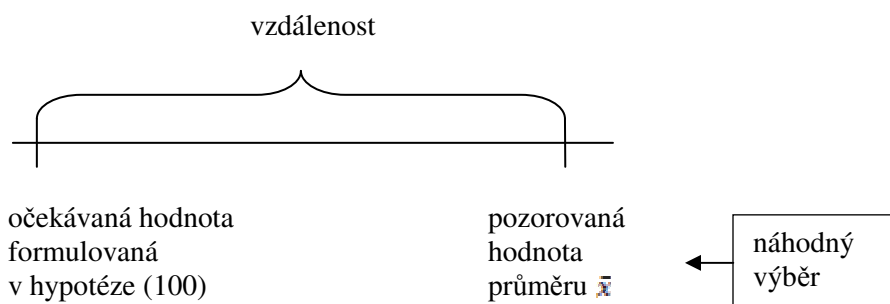
Ve statistice se volí jako rozhodující vstupní parametr testu pravděpodobnost chyby I. druhu – hladina významnosti α . Obvykle volíme 5%-ní nebo 1%-ní hladinu významnosti, tj. $\alpha=0,05$, resp. $\alpha=0,01$.

Jediný způsob jak snížit β , aniž bychom tím zvýšili α , je zvětšit rozsah výběrového souboru.

11.8. Princip testování hypotéz

Základní princip testování hypotéz spočívá v tom, že měříme vzdálenost pozorované statistiky od očekávané hodnoty stanovené hypotézou.

V případě ověřování platnosti nulové hypotézy $H_0: \mu = 100$ potřebujeme zjistit jaká je vzdálenost populačního průměru μ od konstanty 100. Populační průměrnou hodnotu μ neznáme, ale můžeme ji odhadnout pomocí výběrového průměru \bar{x} počítaného z naměřených dat. Bude-li vzdálenost mezi \bar{x} a konstantou 100 příliš velká, bude to pro nás signál zpochybňující platnost testované hypotézy.



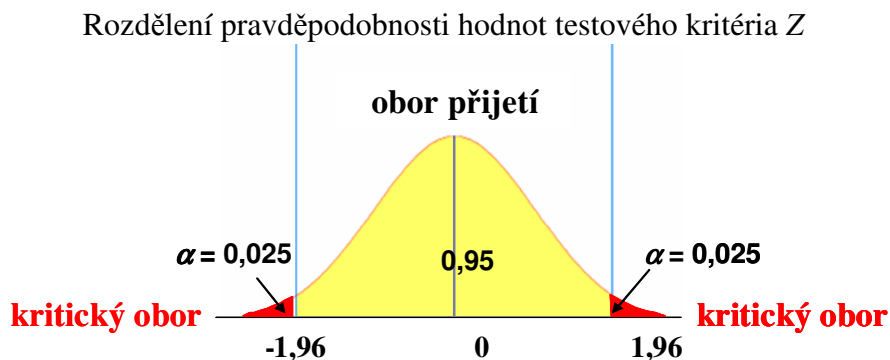
K posouzení toho, jaká vzdálenost je dost velká na to, abychom zpochybnili platnost nulové hypotézy, potřebujeme znát variabilitu, jakou by výběrový průměr měl, kdybychom experiment opakovali vícekrát. Hrubé pravidlo říká, že je-li pozorovaná hodnota od

očekávané hodnoty dále než dvě standardní chyby (tj. chyby výběrového průměru SE), pak se hypotéza zamítá. Přitom vycházíme z toho, že rozdělení výběrových průměrů je normální a hodnotám, které jsou od očekávaného průměru dále než dvě SE, u normálního rozložení, odpovídá dost malá pravděpodobnost (zhruba 4,6%).

Vzdálenost pozorované a očekávané hodnoty můžeme měřit pomocí testové statistiky:

$$z = \frac{\text{pozorovaná hodnota} - \text{očekávaná hodnota}}{\text{standardní chyba pozorované hodnoty}}$$

Takto zkonstruovaná statistika Z má přibližně standardní normální rozdělení (s parametry: průměr=0, směrodatná odchylka=1). Pro konkrétní (vypočítanou) hodnotu Z můžeme zjistit odpovídající pravděpodobnost.



Kritický obor a obor přijetí pro testovou statistiku

Pro statistiku Z můžeme stanovit tzv. kritický obor, resp. obor přijetí.

V podstatě jde o to, že prostor všech možných hodnot testové statistiky je třeba rozdělit na dvě části: na **obor přijetí** obsahující hodnoty testové statistiky svědčící pro přijetí nulové hypotézy a na tzv. **kritický obor** obsahující hodnoty svědčící pro zamítnutí nulové hypotézy. Hranice mezi kritickým oborem a oborem přijetí se nazývá **kritická hodnota testu**. Kritický obor konstruujeme tak, aby pravděpodobnost, že hodnota testové statistiky bude ležet v kritickém oboru, přestože je nulová hypotéza platná, byla rovna předem zvolené hodnotě α . Je-li alternativní hypotéza ve tvaru " \neq ", svědčí ve prospěch alternativy jak extrémně nízké, tak extrémně vysoké hodnoty testové statistiky.

Kritický obor se zpravidla definuje jako oblast hodnot testového kritéria, kterým odpovídají pravděpodobnosti $< 0,05$. Předem zvolené hranici (0,05) říkáme **hladina významnosti**.

Padne-li hodnota testového kritéria Z do kritického oboru, H_0 zamítáme.

Čím větší je vzdálenost pozorované hodnoty od očekávané, tím vyšší je hodnota testového kritéria Z a tím menší je pravděpodobnost dostat takto vzdálené hodnoty za platnosti nulové hypotézy.

Při testování hypotéz se můžeme setkat se dvěma přístupy: 1) klasickým testem nebo 2) čistým testem významnosti.

Klasický test významnosti

Klasický test významnosti se skládá z následujících kroků:

- **Formulace nulové a alternativní hypotézy.**
- **Volba vhodné testové statistiky** (na základě typu dat a konkrétní situace).
- **Sestrojení kritického oboru a oboru přijetí pro hodnoty testové statistiky.**
- **Výpočet pozorované hodnoty testové statistiky.**

- **Formulace závěru testu:**

Padne-li vypočítaná hodnota testové statistiky do kritického oboru, potom zamítáme nulovou hypotézu ve prospěch alternativní hypotézy.

Padne-li testová statistika do oboru přijetí, potom nulovou hypotézu nelze zamítnout.

Čistý test významnosti

Čistý test významnosti řeší otázku, zda hodnoty naměřené v náhodném výběru podporují nulovou hypotézu. V klasickém testu nepotřebujeme znát předem hladinu významnosti, protože výsledek klasického testu nám umožní rozhodnout, na jakých hladinách významnosti můžeme nulovou hypotézu zamítnout.

Čistý test významnosti se skládá z následujících kroků:

- **Formulace nulové a alternativní hypotézy.**
- **Volba vhodné testové statistiky.**
- **Výpočet pozorované hodnoty testové statistiky a výpočet statistiky p -value (tj. p -hodnoty, signifikance testu)**
- **Formulace závěru testu.**

Statistika p -value nám říká jaká je minimální hladina významnosti, na níž bychom při daném výběrovém souboru mohli nulovou hypotézu zamítnout. Je-li např. p -value = 0,001 znamená to, že nulovou hypotézu můžeme zamítnout na hladinách významnosti 0,001 a vyšších. Čím menší je hodnota p -value, tím silnější je výpověď náhodného výběru proti nulové hypotéze. Obecné schéma, podle kterého se řídíme při rozhodování o platnosti nulové hypotézy, pracuje s nejběžněji používanou hladinou významnosti 0,05 se kterou vypočítanou hodnotu p -value porovnáváme:

Situace	Naše rozhodnutí
p -value < 0,05	H_0 zamítáme
p -value \geq 0,05	H_0 nelze zamítnout

Kontrolní otázky:

1. Vysvětlíte pojem „bodové odhady“ populačních parametrů.
2. Jakým způsobem se konstruuje 95% interval spolehlivosti pro populační průměr v případě, že rozsah výběrového souboru $n > 100$?
3. Definujte pojem „statistická hypotéza“, uveďte konkrétní příklady hypotéz.
4. Co znamená „chyba I. druhu“ a „chyba II. druhu“ při ověřování platnosti nulových hypotéz?
5. Definujte pojmy „kritický obor“ a „obor přijetí“ pro testovou statistiku.

12. Testy hypotéz pro kvalitativní znaky

12.1. Chí-kvadrát test

Kvalitativní znaky popisujeme pomocí četnostních tabulek, které popisují rozdělení (distribuci) měřených osob do kategorií kvalitativního znaku. V případě zjišťování závislosti dvou kvalitativních znaků používáme kontingenční tabulku (kategorie jednoho znaku definují řádky kontingenční tabulky a kategorie druhého znaku definují sloupce tabulky).

	Kouření		Celkem
	Ano (kouří)	Ne (nekouří)	
Muži	17 (34,0 %)	33 (66,0 %)	50 (100 %)
Ženy	19 (25,3 %)	56 (74,7 %)	75 (100 %)
Celkem	36 (28,8 %)	89 (71,2 %)	125 (100 %)

Četnostní tabulky umožňují testování

- hypotéz o typu distribuce (např. distribuce je normální, distribuce je rovnoměrná).

Kontingenční tabulky umožňují testování

- hypotéz o stejné distribuci hodnot znaku ve dvou nebo více výběrech,
- hypotéz o nezávislosti dvou kategoriálních znaků,
- hypotézu o symetrii (v případě opakovaného měření kvalitativního znaku).

Univerzálním metodou, která se používá pro testování hypotéz o kategoriálních znacích je **chí-kvadrát test** (χ^2 -test).

Chí-kvadrát test je neparametrická metoda, která se používá ke zjištění, zda mezi dvěma kategoriálními znaky existuje prokazatelný výrazný vztah. Test chí-kvadrát je založen na porovnání pozorovaných četností s četnostmi očekávanými za platnosti nulové hypotézy. **Pozorované četnosti** jsou četnosti zjištěné na výběrovém souboru (viz četnosti v kontingenční tabulce). **Očekávané četnosti** jsou četnosti, které bychom očekávali za platnosti nulové hypotézy.

Testová statistika pro chí-kvadrát test má obecný tvar:

$$\chi^2 = \sum \frac{(\text{pozorované četnosti} - \text{očekávané četnosti})^2}{\text{očekávané četnosti}}$$

Testová statistika má χ^2 rozdělení o $df = (\text{počet řádků} - 1) \text{ krát } (\text{počet sloupců} - 1)$ stupních volnosti.

Kritickou hodnotu χ^2 rozdělení $\chi^2_{1-\alpha}(df)$ najdeme ve statistických tabulkách, zpravidla pro $\alpha=0,05$.

12.2. Příklad 1

Chceme ověřit platnost hypotézy H_0 : **Kouření nezávisí na pohlaví** proti alternativní hypotéze H_A : **Kouření závisí na pohlaví**.

Řešení:

Z naměřených dat vytvoříme kontingenční tabulku pozorovaných četností

	Kouření		Celkem
	kuřák	nekuřák	
Muži	17	33	50
Ženy	19	56	75
Celkem	36	89	125

Zvolíme vhodnou testovou statistiku, tj. statistiku chí-kvadrát test

$$\chi^2 = \sum \frac{(\text{pozorované četnosti} - \text{očekávané četnosti})^2}{\text{očekávané četnosti}}$$

K výpočtu testové statistiky potřebujeme zjistit očekávané četnosti za platnosti nulové hypotézy. Vytvoříme tedy tabulku očekávaných četností:

	Kouření		Celkem
	kuřák	nekuřák	
Muži	14,4 (28,8%)	35,6 (71,2%)	50 (100%)
Ženy	21,6 (28,8%)	53,4 (71,2%)	75 (100%)
Celkem	36 (28,8%)	89 (71,2%)	125 (100%)

Výpočet očekávaných četností

Očekávané četnosti můžeme pro každé políčko kontingenční tabulky vypočítat podle obecného vztahu

$$\text{očekávaná četnost} = \frac{\text{součet v příslušném řádku} \times \text{součet v příslušném sloupci}}{\text{celkový počet pozorování}}$$

Očekávaná četnost pro pozorovanou četnost 17 (tj. muži kuřáci) = $(50 \cdot 36) / 125 = 14,4$,
 Očekávaná četnost pro pozorovanou četnost 33 (tj. muži nekuřáci) = $(50 \cdot 89) / 125 = 35,6$,
 Očekávaná četnost pro pozorovanou četnost 19 (tj. ženy kuřáčky) = $(75 \cdot 36) / 125 = 21,6$,
 Očekávaná četnost pro pozorovanou četnost 56 (tj. ženy nekuřáčky) = $(75 \cdot 89) / 125 = 53,4$.

Pokud by mezi muži a ženami nebyl v kouření rozdíl, potom by se pozorované četnosti měly blížit očekávaným četnostem a případná odchylka by byla způsobena pouze náhodou. Způsob jak zjistit rozdíly mezi pozorovanými a očekávanými četnostmi je vypočítat χ^2 statistiku.

Výpočet hodnoty testové statistiky

$$\chi^2 = \frac{(17 - 14,4)^2}{14,4} + \frac{(33 - 35,6)^2}{35,6} + \frac{(19 - 21,6)^2}{21,6} + \frac{(56 - 53,4)^2}{53,4} = 1,098$$

Stanovení kritického oboru pro hodnoty testové statistiky

Statistika χ^2 má chí-kvadrát rozdělení o df stupních volnosti, kde

$$df = (\text{počet řádků} - 1) \cdot (\text{počet sloupců} - 1),$$

kde počet řádků a počet sloupců znamená počet řádků a počet sloupců v kontingenční tabulce.

V tomto příkladu má kontingenční tabulka 2 řádky a 2 sloupce, tedy $df = (2-1)(2-1) = 1$.

Ze statistických tabulek zjistíme tzv. kritickou hodnotu rozdělení χ^2 o df stupních volnosti, což je kvantil $\chi^2_{1-\alpha}(df)$.

Tabulka A.5: Kvantily $\chi^2_{1-\alpha}$ rozdělení χ^2 o df stupních volnosti

$\chi^2_{1-\alpha}(df)$	α		
	0,05	0,01	0,001
df			
1	3,84	6,63	10,83
2	5,99	9,21	13,82
3	7,81	11,34	16,27
4	9,49	13,28	18,47
5	11,07	15,09	20,51

$\chi^2_{1-0,05}(1) = 3,84$ ($\alpha = 0,05$ je spolehlivost testu). Kritickou oblast tedy tvoří hodnoty testové statistiky $\geq 3,84$.

Rozhodnutí o platnosti nulové hypotézy

Hodnota testové statistiky je menší než kritická hodnota $\chi^2 = 1,098 < 3,84$ (tj. nepadne do kritického oboru), nulovou hypotézu tedy nelze zamítnout.

Ověření platnosti nulové hypotézy - čistý test významnosti

V případě, že máme k dispozici statistický software, nebo použijeme-li např. funkci CHITEST v programu MS Excel, můžeme kromě hodnoty testové statistiky vypočítat také hodnotu **p-value** (tj. **signifikanci chí-kvadrát testu**) a rozhodnout o platnosti nulové hypotézy porovnáním hodnoty p-value s hraniční hodnotou 0,05.

Výsledek chí-kvadrát testu ze statistického programu SPSS:

chí-kvadrát test

	Hodnota	df	Signifikance (p-value)
Pearson Chi-Square	1,099	1	,295
N	125		

Signifikance testu chí-kvadrát pro data z tohoto příkladu je $p = 0,295$. Hodnota $p = 0,295$ je větší než hraniční hodnota 0,05.

Nulovou hypotézu tedy nelze zamítnout. Kouření nezávisí na pohlaví.

12.3. Příklad 2

V rámci průzkumu prováděného na gymnáziích a středních odborných učilištích (SOU) byl zjištěn počet kouřících studentů. Zajímá nás, zda je míra kouření na gymnáziích a SOU stejná.

Řešení:

Formulace nulové a alternativní hypotézy:

H_0 : Podíl kuřáků na gymnáziích a SOU je stejný.

H_A : Podíl kuřáků na gymnáziích a SOU se liší.

Pomocí kontingenční tabulky popíšeme distribuci studentů podle kouření pro gymnázia a pro SOU

	Kuřáci	Nekuřáci	Celkem
Gymnázium	16 (14 %)	99 (86 %)	115 (100 %)
SOU	61 (38 %)	101 (62 %)	162 (100 %)
Celkem	77 (28 %)	200 (72 %)	277 (100 %)

Výpočet očekávaných četností za platnosti nulové hypotézy

	Kuřáci	Nekuřáci
Gymnázium	31,97 (=115*77/ 277)	83,03 (=115*200/ 277)
SOU	45,03 (=162*77/ 277)	116,97 (=162*200/ 277)

Výpočet hodnoty testového kritéria chí-kvadrát testu

$$\chi^2 = \frac{(16 - 31,97)^2}{31,97} + \frac{(99 - 83,03)^2}{83,03} + \frac{(61 - 45,03)^2}{45,03} + \frac{(101 - 116,97)^2}{116,97} = 18,89$$

Stanovení kritického oboru pro hodnoty testové statistiky a rozhodnutí o platnosti nulové hypotézy

Kritická hodnota chí-kvadrát testu pro $df=(2-1)*(2-1)=1$ stupeň volnosti a spolehlivost $\alpha = 0,05$ je $\chi^2_{1-0,05}(1) = 3,84$. Kritickou oblast tvoří hodnoty testové statistiky $\geq 3,84$.

Vypočítaná hodnota testové statistiky je větší než kritická hodnota, $\chi^2 = 18,89 > 3,84$, nulovou hypotézu tedy zamítáme na hladině signifikance 0,05 ve prospěch alternativní hypotézy.

Výpočet hodnoty p-value (signifikance testu chí-kvadrát) a rozhodnutí o platnosti nulové hypotézy

Výsledek chí-kvadrát testu ze statistického programu SPSS (value = hodnota testové statistiky):

Chi-Square Tests

	Value	df	Signifikance
Pearson Chi-Square	18,888	1	,00001
N of Valid Cases	277		

Signifikance testu chí-kvadrát $p = 0,00001$ je mnohem menší než hraniční hodnota 0,05. Nulovou hypotézu zamítáme, chí-kvadrát test prokázal statisticky signifikantně vyšší podíl kuřáků na SOU (38 %) oproti gymnáziím (14 %).

12.4. Příklad 3

V náhodném výběru 50 obézních dětí ve věku 6-14 let byla u každého dítěte zjištěna obezita u matky a obezita u otce. Zajímá nás, zda obezita rodičů spolu souvisí.

Formulace nulové a alternativní hypotézy:

H_0 : Obezita rodičů spolu nesouvisí

H_A : Obezita rodičů spolu souvisí

Vzhledem k tomu, že oba sledované znaky, tj. „obezita matky“ a „obezita otce“ jsou znaky kategoriální s dvěma kategoriemi (obézní, neobézní), můžeme zjištěnou situaci popsat pomocí kontingenční tabulky a nulovou hypotézu lze testovat pomocí chí-kvadrát testu.

Crosstabulation

			otec		Total
			obézní	neobézní	
matka	obézní	Count	15	9	24
		%	62,5%	37,5%	100,0%
	neobézní	Count	7	19	26
		%	26,9%	73,1%	100,0%
Total		Count	22	28	50
		%	44,0%	56,0%	100,0%

Výpočet očekávaných četností za platnosti nulové hypotézy

	Otec obézní	Otec neobézní
Matka obézní	10,56 (=24*22/50)	13,44 (=24*28/50)
Matka neobézní	11,44 (=26*22/50)	14,56 (=26*28/50)

Výpočet hodnoty testového kritéria chí-kvadrát testu

$$Chi^2 = \frac{(15 - 10,56)^2}{10,56} + \frac{(9 - 13,44)^2}{13,44} + \frac{(7 - 11,44)^2}{11,44} + \frac{(19 - 14,56)^2}{14,56} = 6,41$$

Stanovení kritického oboru pro hodnoty testové statistiky a rozhodnutí o platnosti nulové hypotézy

Kritická hodnota chí-kvadrát testu pro $df=(2-1)*(2-1)=1$ stupeň volnosti a spolehlivost $\alpha = 0,05$ je $\chi^2_{1-0,05}(1) = 3,84$. Kritickou oblast tvoří hodnoty testové statistiky $\geq 3,84$.

Vypočítaná hodnota testové statistiky je větší než kritická hodnota, $\chi^2 = 6,41 > 3,84$, nulovou hypotézu tedy zamítáme na hladině signifikance 0,05 ve prospěch alternativní hypotézy.

Výpočet hodnoty p-value (signifikance testu chí-kvadrát) a rozhodnutí o platnosti nulové hypotézy

Výsledek chí-kvadrát testu ze statistického programu SPSS (value = hodnota testové statistiky):

Chi-Square Tests

	Value	df	Signifikance
Pearson Chi-Square	6,411	1	,011
N of Valid Cases	50		

Signifikance chí-kvadrát testu $p = 0,011$ je menší než hraniční hodnota 0,05.

Nulovou hypotézu zamítáme, chí-kvadrát test prokázal statisticky signifikantní závislost mezi obezitou obou rodičů.

Kontrolní otázky:

1. Vysvětlete princip chí-kvadrát testu.
2. Pro jaký typ dat lze použít chí-kvadrát test.

13. Testy hypotéz pro kvantitativní data

Znaky kvantitativní poskytují o měřené vlastnosti více informace než znaky kvalitativní.

U kvantitativních znaků lze popsat, kromě distribuce hodnot sledovaného znaku v definovaných intervalech, také parametry této distribuce, tj. charakteristiky polohy dat (průměr, medián) a charakteristiky variability dat (rozptyl, směrodatnou odchylku, ...).

Hypotézy pro kvantitativní znaky jsou především:

- tvrzení o typu distribuce (např. distribuce hodnot daného znaku je normální) a
- tvrzení o hodnotách parametrů této distribuce (tj. tvrzení o průměru, mediánu, rozptylu, atd.).

V této kapitole se zaměříme, vzhledem k rozsahu této problematiky, pouze na testy hypotéz o průměrné hodnotě znaku.

K testování hypotéz o populačním průměru daného kvantitativního znaku se používá tzv. **Studentův t-test**, resp. metoda **Analýza rozptylu** v situaci, kdy porovnáváme 3 a více populačních průměrů.

Jednotlivé situace shrnuje následující tabulka:

Přehled parametrických testů pro testování hypotéz o populačním průměru pro kvantitativní data

Situace	Testovaná hypotéza	Test pro ověření platnosti hypotézy
Jeden výběrový soubor	$H_0: \mu = \text{očekávaná hodnota}$	Studentův t-test jednovýběrový
Jeden výběrový soubor, opakovaná měření znaku	$H_0: \mu_{\text{před léčbou}} = \mu_{\text{po léčbě}}$	Studentův t-test párový
Dva výběrové soubory	$H_0: \mu_1 = \mu_2$	Studentův t-test dvouvýběrový
Tři a více výběrových souborů	$H_0: \mu_1 = \mu_2 = \dots = \mu_n$	Analýza rozptylu (ANOVA)

(μ je populační průměr znaku, který odhadujeme pomocí výběrového průměru)

13.1. Testování hypotéz o průměru - jeden výběrový soubor

Máme náhodný výběr n osob, u nichž měříme hodnotu nějakého spojitého znaku (např. hladinu albuminu, Body Mass Index (BMI), celkový cholesterol, atd.).

Chceme testovat nulovou hypotézu $H_0: \mu = \mu_0$

tj. průměrná hodnota znaku v populaci (μ), z níž výběr pochází, je rovna určitému číslu μ_0 (tj. očekávané hodnotě - např. referenční hodnotě, tj. normě pro zdravou populaci).

Populační průměrnou hodnotu μ daného znaku neznáme, odhadujeme ji pomocí výběrového průměru.

K testování nulové hypotézy použijeme **jednovýběrový Studentův t -test**.

K výpočtu hodnoty testové statistiky použijeme vzorec:
$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

kde \bar{x} je výběrový průměr, s je výběrová směrodatná odchylka, n je rozsah výběru, resp. počet naměřených hodnot.

Testová statistika T má tzv. **Studentovo t rozdělení** o $n-1$ stupních volnosti. **Počet stupňů volnosti** (df) se počítá jako rozdíl rozsahu výběrového souboru (tj. počet naměřených hodnot) mínus počet statistik použitých při výpočtu hodnoty testového kritéria (kromě statistiky, která je předmětem testování) – v tomto případě se tedy od rozsahu výběru n odečítá hodnota 1 za použití směrodatné odchylky s . (Statistika T nemá normální rozdělení, protože při výpočtu používáme místo populační směrodatné odchylky σ její odhad, tj. výběrovou směrodatnou odchylku s .)

Předpoklad použití Studentova t -testu

Použití t -testu předpokládá normální distribuci hodnot sledovaného znaku.

13.2. Příklad 1

Studie o dětech s vrozenou srdeční vadou zjišťovala, v jakém věku začínají děti chodit. Ve studii bylo 35 dětí, průměrný věk, kdy začínaly chodit, byl 12,8 měsíců, směrodatná odchylka byla 2,5 měsíce. Z předchozích výzkumů je známo, že v celé populaci zdravých dětí byl průměrný věk, kdy děti začínaly chodit 11,4 měsíce.

Zajímá nás, zda je možné na základě této studie prohlásit, že srdeční vada negativně ovlivňuje věk, kdy začíná dítě chodit.

Formulace nulové a alternativní hypotézy: $H_0: \mu = 11,4$ měsíce, $H_A: \mu > 11,4$ měsíce.

(v tomto případě jde o test nulové hypotézy proti jednostranné alternativě).

Řešení:

Pro odhad populačního průměrného věku μ použijeme výběrový průměr \bar{x} . Základní popisné statistiky věku shrnuje následující tabulka:

	N	Průměr \bar{x}	Směrodatná odchylka s	Chyba průměru s / \sqrt{n}
věk	35	12,83	2,48	0,826

Výpočet hodnoty testového kritéria Studentova jednovýběrového t -testu

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{12,83 - 11,4}{\frac{2,48}{\sqrt{35}}} = 1,74$$

Stanovení kritického oboru pro testovou statistiku a rozhodnutí o platnosti nulové hypotézy

Ve statistických tabulkách najdeme kvantil Studentova t rozdělení pro $df = 35-1=34$ stupňů volnosti a spolehlivost $\alpha = 0,05$. $t_{1-\alpha}(34) = 1,69$

Tabulka A.4: Kvantily $t_{1-\alpha}(df)$			
Studentova t rozdělení o df stupních volnosti			
$t_{1-\alpha}(df)$	α		
df	0,05	0,01	0,001
1	6,314	31,821	318,289
2	2,920	6,965	22,328
3	2,353	4,541	10,214
30	1,697	2,457	3,385
35	1,690	2,438	3,340
40	1,684	2,423	3,307
45	1,679	2,412	3,281
50	1,676	2,403	3,261
60	1,671	2,390	3,232
70	1,667	2,381	3,211
80	1,664	2,374	3,195
90	1,662	2,368	3,183
100	1,660	2,364	3,174
∞	1,645	2,326	3,090

Kritickou oblast tvoří hodnoty testové statistiky $\geq 1,69$.

Hodnota testové statistiky je menší než kritická hodnota, $T = 1,74 > 1,69$, nulovou hypotézu tedy nelze zamítnout. Na základě našeho výběru tedy nelze prokázat, že by vrozená srdeční vada měla negativní vliv na věk, v němž začíná dítě chodit.

Výpočet hodnoty p -value (signifikance testu chí-kvadrát) a rozhodnutí o platnosti nulové hypotézy

Výsledek chí-kvadrát testu ze statistického programu SPSS (t = hodnota testové statistiky):

One-Sample Test						
	Test Value = 11.4				95% CI pro rozdíl průměrů	
	t	df	Signifikance	rozdíl průměrů	Lower	Upper
věk	1,74	34	,121	1,433	-,471	3,338

Hodnota $t = 1,74$ je hodnota testové statistiky jednovýběrového t -testu, této hodnotě odpovídá vypočtená hladina signifikance testu $p = 0,121 > 0,05$.

Testovanou hypotézu $H_0: \mu = 11,4$ měsíce tedy nelze zamítnout.

Na základě našeho výběru nelze prokázat, že by vrozená srdeční vada měla negativní vliv na věk, v němž dítě začíná chodit.

13.3. Testování hypotéz o průměru pro dva výběry

V lékařském výzkumu se často zajímáme, zda se soubor pacientů liší v hodnotách sledovaného kvantitativního znaku od kontrolního souboru zdravých osob. Máme tedy dva výběrové soubory, výběrový soubor pacientů a kontrolní soubor, tj. výběr z populace zdravých osob. U obou výběrových souborů měříme hodnoty daného kvantitativního znaku a chceme testovat hypotézu

$$H_0: \mu_1 = \mu_2,$$

tj. populační průměrné hodnoty znaku se neliší.

μ_1 je průměr znaku v populaci pacientů, z níž pochází výběrový soubor pacientů, μ_2 je průměr znaku v populaci zdravých osob, z níž pochází výběrový kontrolní soubor. Populační průměry odhadujeme pomocí výběrových průměrů \bar{x}_1 a \bar{x}_2 .

K testování hypotézy použijeme **dvouvýběrový Studentův t -test**

Pro výpočet hodnoty testové statistiky použijeme vzorec:

1) v případě shodných rozptylů měřeného znaku

$$T = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

2) v případě různých rozptylů měřeného znaku

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}_1, s_1 \dots$ průměr a směrodatná odchylka znaku v 1. výběru (výběrový soubor pacientů)

$\bar{x}_2, s_2 \dots$ průměr a směrodatná odchylka znaku ve 2. výběru (kontrolní skupina)

$n_1, n_2 \dots$ rozsahy výběrů.

Testová statistika T má Studentovo t rozdělení pro $df = n_1 + n_2 - 2$ stupňů volnosti. Počet stupňů volnosti se počítá jako rozdíl počtu hodnot naměřených v obou výběrech mínus počet statistik použitých při výpočtu hodnoty testového kritéria (kromě statistiky, která je předmětem testování) – v tomto případě tedy mínus 2 za použití výběrových rozptylů s^2 .

Poznámka:

Použití t -testu předpokládá normální distribuci hodnot sledovaného znaku v obou výběrech.

Pro ověření hypotézy o shodnosti rozptylů se v programu Excel používá F -test shodnosti rozptylů.

13.4. Příklad

Na ortopedické klinice byla u 92 pacientů provedena operace, při níž byl pacientům vyměněn celý kyčelní kloub (tj. byla implantována totální endoprotéza kyčle). Z toho u 42 pacientů byla provedena operace navigovaná počítačem, u 50 pacientů operace klasická, nenavigovaná. U obou skupin pacientů byla zaznamenána délka operace v minutách. Zajímá nás, zda se doba operace u navigovaných a nenavigovaných operací liší.

Řešení:

Formulace nulové a alternativní hypotézy:

$H_0: \mu_1 = \mu_2$, (tj. doba operace u navigované a nenavigované TEP kyčle se neliší.)

$H_A: \mu_1 \neq \mu_2$.

K ověření platnosti nulové hypotézy použijeme **dvouvýběrový Studentův t -test**.

Populační průměrné hodnoty délky operace μ_1 a μ_2 odhadneme pomocí výběrových průměrů. Z dat výběru dále vypočítáme výběrové směrodatné odchylky s_1 a s_2 .

Základní popisné statistiky délky operace shrnuje následující tabulka:

Délka operace (minuty)	N	min – max	Průměr	Směrodatná odchylka
Navigované operace	42	80-140	106,3	14,75
Nenavigované operace	50	55-115	85,0	14,71

Výpočet hodnoty testového kritéria Studentova dvouvýběrového t -testu

Za předpokladu stejných rozptylů hodnot měřeného znaku v obou výběrech (viz. Zvárová F-test shodnosti rozptylů), můžeme k výpočtu hodnoty testového kritéria použít následující vzorce:

Nejdříve vypočítáme sdružený odhad rozptylů

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(42 - 1) \cdot 14,75^2 + (50 - 1) \cdot 14,71^2}{42 + 50 - 2}} = 14,73$$

Hodnota testového kritéria

$$T = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{106,3 - 86,0}{14,73 \sqrt{\frac{1}{42} + \frac{1}{50}}} = \frac{20,3}{3,08} = 6,58$$

Stanovení kritického oboru pro hodnoty testové statistiky a rozhodnutí o platnosti nulové hypotézy

Ve statistických tabulkách najdeme kvantil Studentova t rozdělení pro $df = n_1 + n_2 - 2 = 42 + 50 - 2 = 90$ stupňů volnosti a spolehlivost $\alpha = 0,05$.

$$t_{1-\alpha/2}(90) = 1,987.$$

$t_{1-\frac{\alpha}{2}}(df)$	α		
	0,05	0,01	0,001
90	1,987	2,632	3,402
100	1,984	2,626	3,390
∞	1,960	2,576	3,290

Kritickou oblast tvoří hodnoty testové statistiky $\geq 1,987$ (při tomto stanovení kritické oblasti předpokládáme, že při výpočtu hodnoty testové statistiky dostaneme kladnou hodnotu, tj. odečítáme od vyšší hodnoty průměru průměr menší, resp. že pro porovnání s kritickou hodnotou použijeme absolutní hodnotu testového kritéria).

Hodnota testové statistiky je větší než kritická hodnota, $T = 6,58 > 1,987$, nulovou hypotézu tedy zamítáme ve prospěch alternativní hypotézy.

Na základě našeho výběru jsme tedy prokázali, že délka operace u navigovaných operací je signifikantně delší než u nenavigovaných operací.

Výpočet hodnoty p -value (signifikance testu chí-kvadrát) a rozhodnutí o platnosti nulové hypotézy

Výsledek Studentova t -testu ze statistického programu SPSS (t = hodnota testové statistiky)

	Levenův test rovnosti rozptylů		t-test rovnosti průměrů				
	F	Sig.	t	df	Signifikance	Rozdíl průměrů	Chyba rozdílu
Doba operace	,137	,712	6,596	90	< 0,0001	20,33	3,082

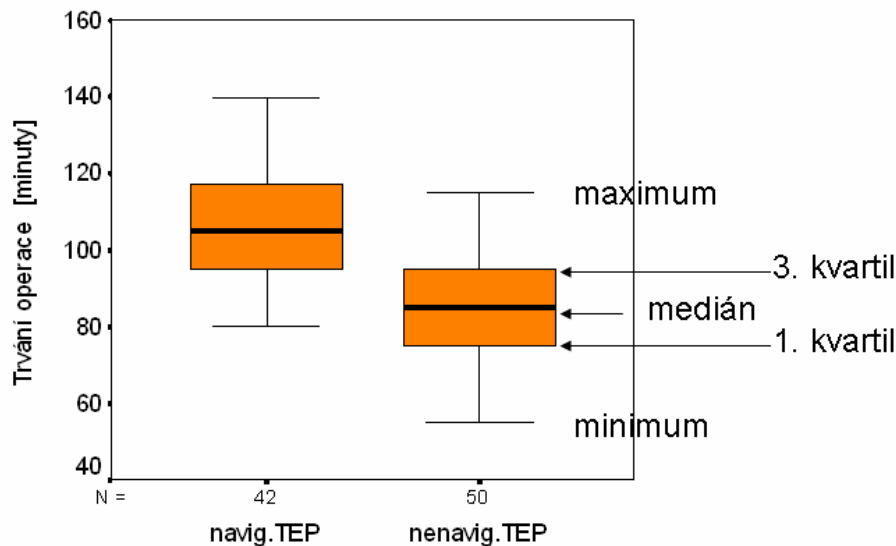
Hodnota $t = 6,596$ je hodnota testové statistiky dvouvýběrového Studentova testu t -testu, této hodnotě odpovídá vypočtená hladina signifikance testu $p < 0,0001 < 0,05$.

Testovanou hypotézu $H_0: \mu_1 = \mu_2$ (délka operace u navigovaných a nenavigovaných operací je stejná) tedy zamítáme.

Na základě našeho výběru jsme tedy prokázali, že délka operace u navigovaných operací je signifikantně delší (106 minut) než u nenavigovaných operací (86 minut).

Poznámka: V tabulce s výsledky programu SPSS, jsou také výsledky Levenova testu shodnosti rozptylů, které nám říkají, že rozptyly délky operace obou typů operace se neliší statisticky významně ($p = 0,712 > 0,05$). Při výpočtu hodnoty testového kritéria podle vzorce se proto použil vzorec pro případ shodných rozptylů.

Distribuci trvání operace lze graficky popsat pomocí kvartilového box grafu



13.5. Testování hypotéz o průměru pro jeden výběr, opakovaná měření

Na výběrovém souboru často měříme hodnoty daného kvantitativního znaku opakovaně ve dvou různých situacích (např. před léčbou a po léčbě), tzn. že máme k dispozici tzv. párová data. Zajímá nás, zda se vlivem léčby změnila průměrná hodnota sledovaného znaku (tj. zda měla léčba nějaký efekt).

Na těchto datech potom můžeme testovat hypotézu

$H_0: \mu_{\text{před léčbou}} = \mu_{\text{po léčbě}}$, tj. populační průměrná hodnota znaku před léčbou se neliší od populační průměrné hodnoty znaku po léčbě. K ověření platnosti hypotézy použijeme Studentův t -test párový s testovou statistikou

$$T = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$$

kde \bar{d} je průměrná hodnota diferencí (hodnota před léčbou – hodnota po léčbě)

s_d je směrodatná odchylka diferencí (hodnota před léčbou – hodnota po léčbě).

Pokud testovaný lék nemá žádný účinek, budeme očekávat, že populační průměrná hodnota diferencí (před léčbou – po léčbě) bude rovna nule. Populační průměr diferencí neznáme, lze

ho však odhadnout pomocí výběrového průměru diferencí \bar{d} hodnot před léčbou – po léčbě.

Testová statistika T má Studentovo t rozdělení o $df = n - 1$ stupních volnosti.

Předpokladem použití Studentova párového t -testu je normální distribuce diferencí před léčbou – po léčbě.

13.6. Příklad

V souboru $n=396$ zaměstnanců byly v rámci preventivních prohlídek v roce 2005 a opakovaně v roce 2006 zjištěny hodnoty LDL cholesterolu. Zajímá nás, zda se hodnoty zjištěné při 1. a 2. měření liší.

Řešení:

Formulace nulové a alternativní hypotézy

Testovaná nulová hypotéza: $H_0: \mu_1 = \mu_2$, tj. průměrné hodnoty LDL cholesterolu zjištěné při 1. a 2. měření se neliší.

Alternativní hypotéza: $H_A: \mu_1 \neq \mu_2$.

K ověření platnosti nulové hypotézy použijeme Studentův t -test párový.

Populační průměrnou hodnotu difference LDL cholesterolu \bar{d} odhadneme pomocí výběrového průměru diferencí (1. měření – 2. měření). Z dat dále vypočítáme směrodatnou odchylku diferencí. Pro každého pacienta v souboru tedy vypočítáme rozdíl hodnota zjištěná při 1. měření mínus hodnota zjištěná při 2. měření a dále vypočítáme průměrnou hodnotu a směrodatnou odchylku těchto rozdílů.

Základní popisné statistiky diferencí hodnot LDL cholesterolu shrnuje následující tabulka:

N	Průměrná diference \bar{d}	Směrodatná odchylka diferencí s_d
396	1,63	1,05

N ... počet opakovaných měření,

\bar{d} ... průměrná hodnota diferencí 1.měření-2.měření

Výpočet hodnoty testového kritéria Studentova párového t -testu

$$T = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{1,63}{1,05 / \sqrt{396}} = 30,89$$

Stanovení kritického oboru pro hodnoty testové statistiky a rozhodnutí o platnosti nulové hypotézy

Ve statistických tabulkách najdeme kvantil Studentova t rozdělení pro $df = n-1 = 395$ stupňů volnosti a spolehlivost $\alpha = 0,05$. Hodnota kvantilu je $t_{1-\alpha/2}(395) = 1,96$

Kritickou oblast tvoří hodnoty testové statistiky $\geq 1,96$.

Hodnota testové statistiky je větší než kritická hodnota, $T = 30,89 > 1,96$, nulovou hypotézu tedy zamítáme ve prospěch alternativní hypotézy.

Na základě našeho výběru jsme tedy prokázali, že při opakovaném měření se naměřily statisticky významně nižší hodnoty LDL cholesterolu než při 1. měření.

Výpočet hodnoty p -value (signifikance testu) a rozhodnutí o platnosti nulové hypotézy

Výsledek Studentova párového t -testu ze statistického programu SPSS

	Charakteristiky diferencí			hodnota testového kritéria t	df	Signifikance
	Průměrná diference	Odchylka diference	Chyba prům. diference			
LDL cholesterol, (1. - 2.) měření	1,63	1,05	,053	30,9	396	< 0,0001

Hodnota $t = 30,9$ je hodnota testové statistiky Studentova párového t -testu, této hodnotě odpovídá vypočtená hladina signifikance testu $p < 0,0001 < 0,05$.

Testovanou hypotézu ($H_0: \mu_1 = \mu_2$, tj. průměrné hodnoty LDL cholesterolu zjištěné při 1. a 2. měření se neliší), tedy zamítáme.

Na základě našeho výběru jsme prokázali, že při opakovaném měření byly naměřeny statisticky významně nižší hodnoty LDL cholesterolu než při 1. měření.

13.7. Testování hypotéz o průměru pro více než dva výběry

V mnoha situacích potřebujeme porovnávat průměrné hodnoty více než dvou skupin.

Mějme k nezávislých náhodných výběrů. Zajímá nás, zda všechny výběry pochází ze stejné populace (tj. neliší se průměrnou hodnotou měřeného kvantitativního znaku).

Chceme testovat nulovou hypotézu : $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$

Alternativní hypotéza předpokládá, že alespoň jedna průměrná hodnota se liší od ostatních.

Pro ověření platnosti hypotézy použijte metodu, která se nazývá **Analýza rozptylu (ANOVA)**.

Analýza rozptylu je založena na předpokladu, že každý z výběrů pochází z populace s normálním rozdělením se stejnou směrodatnou odchylkou. Zajímá nás, zda střední hodnoty (průměry) skupin jsou všechny shodné, nebo zda se navzájem liší.

Podmínky pro použití analýzy rozptylu:

- předpokládáme normální rozdělení sledovaného kvantitativního znaku ve všech výběrech,
- předpokládáme stejné rozptyly znaku ve všech výběrech.

Princip analýzy rozptylu:

Celkový rozptyl sledovaného znaku rozdělíme na dvě části

- na variabilitu uvnitř výběrů (intraindividuální) a
- variabilitu mezi výběry (interindividuální).

Pokud by neexistoval rozdíl mezi výběrovými průměry, měla by variabilita mezi výběry i variabilita uvnitř výběrů popisovat stejný jev, tj. stejný populační rozptyl. K porovnání těchto variabilit použijeme F -test

Testová statistika F -testu

$$F = \frac{\text{variabilita mezi výběry}}{\text{variabilita uvnitř výběrů}}$$

Statistika F má F rozdělení o ν_1 a ν_2 stupňů volnosti. Počet stupňů volnosti ν_1 je pro čítelek, $\nu_1 = k-1$, kde k je počet porovnávaných skupin. Počet stupňů volnosti ν_2 je pro jmenovatele, $\nu_2 = n_1+n_2+\dots+n_k - k$, kde n_1 až n_k jsou počty pozorování ve skupině 1 až ve skupině k .

V případě platnosti nulové hypotézy bude čítelek F statistiky (zhruba) stejný jako jmenovatel, statistika F tedy bude rovna přibližně jedné.

V případě, že zamítneme nulovou hypotézu, musíme data dále analyzovat, protože tento výsledek nám neříká, které průměry se od sebe navzájem liší. K další analýze použijeme testy mnohonásobného porovnávání.

13.8. Příklad

Mějme tři nezávislé náhodné výběry, tj. sledované osoby rozdělené podle vzdělání. Na těchto výběrech sledujme hodnotu celkového skóre vědomostního testu.

Naměřené údaje shrnuje tabulka

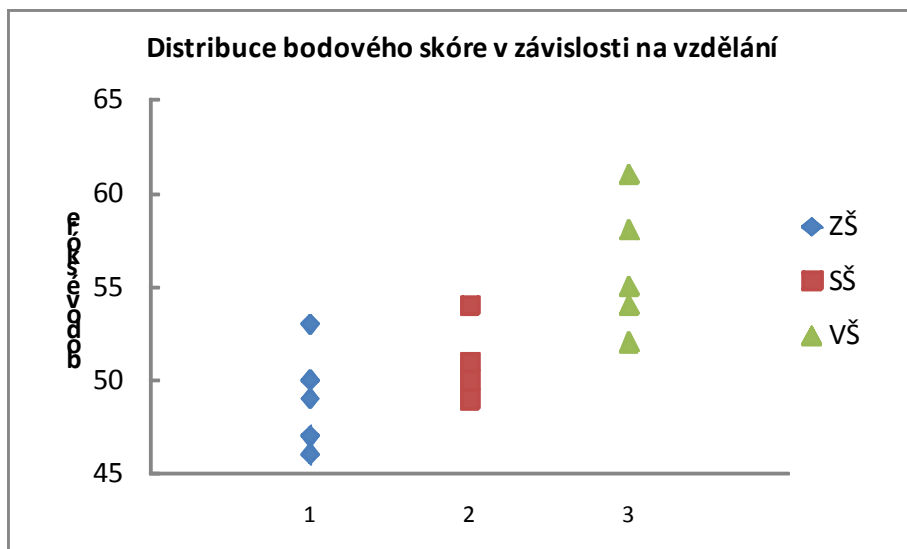
	ZŠ	SŠ	VŠ
Výběr	1	2	3
	47	54	55
	53	50	54
	49	51	58
	50	51	61
	46	49	52
	$\bar{X}_1 = 49$	$\bar{X}_2 = 51$	$\bar{X}_3 = 56$
	$s_1^2 = 7.5$	$s_2^2 = 3.5$	$s_3^2 = 12.5$

X_1, X_2, X_3 jsou výběrové průměrné hodnoty vědomostního skóre, s_1^2, s_2^2, s_3^2 je výběrová variabilita hodnot skóre (tj. druhá mocnina výběrové směrodatné odchylky).

Chceme ověřit platnost hypotézy $H_0: \mu_1 = \mu_2 = \mu_3$, (tj. populační průměrné hodnoty skóre se rovnají).

Řešení:

Prvním krokem při analýze tohoto typu je prozkoumat data vizuálně. Vhodným grafem pro malé výběry je bodový graf, pro větší rozsah výběrových souborů by byl vhodným grafem box graf.



V našem případě je zřejmé, že data neobsahují žádná odlehlá pozorování. Možná vás napadla otázka, proč nelze provést všechny možné dvouvýběrové t testy nezávislosti, tedy testy pro následující dvojice:

ZŠ vs. SŠ,
 ZŠ vs. VŠ,
 SŠ vs. VŠ?

Důvodem je skutečnost, že toto opakované porovnávání neoprávněně zvyšuje pravděpodobnost chyby prvního druhu. U každého testu je řekněme 5% možnost chybného pozitivního výsledku (tj. chyby prvního druhu), pokud neexistuje žádný rozdíl. Máme-li tři skupiny a provedeme všechny tři testy, pravděpodobnost, že dostaneme nejméně jeden chybný pozitivní výsledek (chybu prvního druhu) je větší než 5%. Abychom se tomuto problému vyhnuli, použijeme k testování hypotézy metodu analýzy rozptylu.

Postup

1. Zjistíme variabilitu uvnitř výběrů (tj. průměrný rozptyl ve výběrech)

$$S_{res}^2 = \frac{s_1^2 + s_2^2 + s_3^2}{3} = \frac{7,5 + 3,5 + 12,5}{3} = 7,83$$

2. Zjistíme variabilitu mezi výběry (mezi výběrovými průměry)

$$S_{\bar{x}}^2 = \frac{n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 + n_3(\bar{X}_3 - \bar{X})^2}{k-1} = \frac{5(49-52)^2 + 5(51-52)^2 + 5(56-52)^2}{2} = 65$$

kde $\bar{X} = 52$ je celkový průměr (tj. průměr ze všech naměřených hodnot),
 k je počet výběrových souborů.

Vypočteme hodnotu testového kritéria F

$$F = \frac{S_{\bar{x}}^2}{S_{res}^2} = \frac{65}{7,83} = 8,3$$

V případě platnosti nulové hypotézy H_0 (všechny populační průměry jsou shodné), by byla hodnota F přibližně rovna jedné.

Stanovíme kritický obor pro hodnoty testové statistiky a rozhodneme o platnosti nulové hypotézy

Testová statistika F má F rozdělení o ν_1 (pro čítelel) a ν_2 (pro jmenovatel) stupních volnosti.

$\nu_1 = k - 1$, kde k je počet porovnávaných skupin (výběrových souborů) a

$\nu_2 = n_1 + n_2 + \dots + n_k - k$.

V našem případě budeme ve statistických tabulkách hledat kvantil F rozdělení pro $\nu_1 = 2$ a $\nu_2 = 15-3=12$ stupňů volnosti a pro hladinu spolehlivosti $\alpha = 0,05$.

Tabulka A.6: 95% kvantily F rozdělení o ν_1 (čítatel) a ν_2 (jmenovatel) stupních volnosti

$F_{0,95}(\nu_1, \nu_2)$	ν_1								
	ν_2	1	2	3	4	5	6	8	10
1	161,45	199,50	215,71	224,58	230,16	233,99	238,88	241,88	
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,40	
3	10,13	9,55	9,28	9,12	9,01	8,94	8,85	8,79	
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,64	
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,35	
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,14	
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,98	
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,85	
12	4,75	3,89	3,49	3,26	3,11	3,00	2,85	2,75	
13	4,67	3,81	3,41	3,18	3,03	2,92	2,77	2,67	

$F_{1-\alpha}(\nu_1=2, \nu_2=12) = 3,89$.

Kritickou oblast tedy tvoří hodnoty testové statistiky $\geq 3,89$.

Hodnota testové statistiky je větší než kritická hodnota, $F = 8,3 > 3,89$, nulovou hypotézu tedy zamítáme ve prospěch alternativní hypotézy.

Na základě našeho výběru jsme tedy prokázali, že alespoň jeden z průměrů se liší od ostatních.

Při zpracování těchto dat pomocí statistického software bychom získali pro hodnotu testového kritéria $F = 8,3$ dosaženou hladinu významnosti (p -hodnotu) $p = 0,005 < 0,05$. Hypotézu H_0 bychom tedy zamítli.

Tabulka s výsledky Analýzy rozptylu (statistický software program SPSS)

ANOVA

skóre	Suma čtverců	df	průměrný čtverec	F	Signifikance
Variabilita mezi skupinami	130,0	2	65,00	8,30	,005
Variabilita uvnitř skupin	94,0	12	7,83		
Celková variabilita	224,0	14			

Analýza rozptylu nám pouze říká, že průměrné hodnoty nejsou stejné. Abychom zjistili, jak se liší, je třeba provést další analýzu. Jednou z možností je porovnat každou dvojici průměrů, nebo dvojice, které nás zajímají. Problémem je však skutečnost, že mnohonásobné testování

významnosti dává vysokou pravděpodobnost, že bude nalezen významný rozdíl pouze náhodou. Pro řešení tohoto problému existuje několik metod, jako například Bonferroniho, Tukeyova, Newman-Keulsova, Duncanova, Fisherovo LSD (nejmenší významný rozdíl - Least Significant Difference) a Scheffého. Úkolem každé metody je udržet danou hladinu pravděpodobnosti chyby prvního druhu 5 % a v podstatě ji rozdělit mezi všechna porovnání. Některé z těchto testů jsou velmi konzervativní. Může se stát, že F test zamítne hypotézu o rovnosti průměrů, a přitom žádná dvojice průměrů se od sebe podle výsledků metod mnohonásobného porovnávání navzájem významně neliší nebude.

Mnohonásobné porovnávání - Bonferroniho metoda

Pro porovnání, která nás zajímají, provedeme modifikované t testy s *upravenou hladinou významnosti*. Úprava hladiny α spočívá v jejím vydělení počtem testů, které chceme provést. S touto modifikovanou hladinou významnosti budeme potom porovnávat přesné hladiny signifikance jednotlivých t -testů. V případě našich dat budeme pracovat s hladinou významnosti $\alpha = 0,05/3 = 0,0167$.

Nulová a alternativní hypotéza jsou stejné jako v případě Studentova dvouvýběrového t -testu. Vzorec pro výpočet hodnoty testové statistiky se nepatrně liší od toho, o kterém jsme již hovořili.

Upravený vzorec pro testovou statistiku má tvar

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_e}{df_e} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Testová statistika T má za platnosti nulové hypotézy ($\mu_1 = \mu_2$) Studentovo t rozdělení o df_e stupních volnosti. V našem případě $df_e = 12 (= n_1 + n_2 + \dots + n_k - k)$. s_e je odhad sdruženého rozptylu všech skupin (tedy včetně těch, které nepoužíváme při porovnávání), odpovídá rozptylu uvnitř skupin, tj. reziduálnímu rozptylu ($s_e = 7,83$).

Kritická hodnota pro statistiku T je také jiná. Uvědomme si, že pro každé porovnání se hladina významnosti rovná 1,67 % a nikoli 5 %. K výpočtu kritické hodnoty potřebujeme (pokud nepoužíváme počítač) velmi dobré tabulky tohoto rozdělení.

Hodnoty testového kritéria upraveného t testu

Porovnání skupin	Hodnota testové statistiky T
ZŠ vs. SŠ	3,915
ZŠ vs. VŠ	11,744
SŠ vs. VŠ	5,872

Pokud použijeme počítač, lze spočítat dosaženou hladinu významnosti neboli pravděpodobnost, že získáme naše, nebo ještě silnější výsledky za předpokladu platnosti nulové hypotézy. Získané p -hodnoty jsou již korigovány na mnohonásobné porovnávání, tj. jsou vyděleny počtem provedených porovnání (v našem případě 3 porovnání).

Výsledné p -hodnoty lze tedy přímo porovnat s hraniční hodnotou signifikance 0,05.

Ukázka počítačového výstupu mnohonásobného porovnávání pomocí Bonferroniho metody (program SPSS)

Mnohonásobné porovnávání

Dependent Variable: skóre

Bonferroni

(I) vzdělání	(J) vzdělání	Rozdíl průměrů (I-J)	Chyba rozdílu	Signifikance (p-hodnota)
ZŠ	SS	-2,00	1,770	,842
	VŠ	-7,00*	1,770	,006
SŠ	VŠ	-5,00*	1,770	,046

* rozdíl průměrů je signifikantní (statisticky významný) na hladině signifikance 0.05

Z výsledků mnohonásobného porovnávání podle Bonferroniho metody je zřejmé, že statisticky významný rozdíl v průměrných hodnotách skóre testu byl prokázán mezi skupinou s VŠ vzděláním (průměr 56 bodů) a skupinou se ZŠ vzděláním (průměr 49 bodů) a dále mezi skupinou s VŠ vzděláním SŠ vzděláním (průměr 51 bodů).

14. Neparametrické metody

U všech předchozích testů (Studentovy *t*-testy, ANOVA) jsme předpokládali normální distribuci hodnot sledovaného znaku. V praxi však tento předpoklad není často splněn, nebo jej nemůžeme ověřit (např. u malých výběrů).

Máme-li podezření na výraznou odchylku od normality, nebo v případě výskytu extrémních hodnot v datech, použijeme testy, které předpoklad o normalitě nepotřebují – **neparametrické testy**.

Neparametrické testy neporovnávají parametry, ale spíše distribuci hodnot, používají se také pro znaky ordinální.

Přehled nejznámějších neparametrických testů

- **Mann-Whitney *U*-test** - porovnání dvou výběrových souborů
- **Wilcoxonův test** - pro jeden výběrový soubor, opakovaná měření (tj. párová data)
- **Kruskal-Wallisův test** - pro tři a více výběrových souborů (obdoba metody ANOVA)

Shrnutí

typ distribuce měřeného znaku	normální distribuce	spojitá nenormální distribuce	bi(multi)nomické (kvalitativní znaky)
populační parametr, který se testuje	průměr	medián (distribuční fce)	pravděpodobnosti
jeden výběr	Studentův t-test jednovýběrový	znaménkový test	
párová data	Studentův t-test párový	Wilcoxonův test, znaménkový test	McNemarův test
2 nezávislé výběry	Studentův t-test dvouvýběrový	Mann-Whitney U test, Kolmogorov-Smirnov test	chi-kvadrát test
k nezávislých výběrů	Analýza rozptylu	Kruskal-Wallis test	chi-kvadrát test

Kontrolní otázky:

1. Uveďte přehled základních parametrických testů pro test hypotéz o průměrné hodnotě kvantitativního znaku.
2. Jaký typ statistického testu byste použili pro porovnání souboru pacientů s kontrolním souborem v hodnotách celkového cholesterolu, v případě normální distribuce hodnot v obou souborech?
3. Jaký typ testu byste použili pro ověření závislosti mezi pohlavím pacientů a kouřením?
4. Jaký typ testu byste použili pro porovnání 3 souborů v hodnotách kvantitativního parametru (za předpokladu normální distribuce hodnot ve všech souborech)?
5. Jaký typ testu byste použili pro porovnání opakovaných měření metrického parametru (za předpokladu normálního rozdělení diferencí hodnot)?
6. V jaké situaci byste použili neparametrickou metodu pro porovnání dvou souborů v hodnotách kvantitativního znaku?

15. Měření závislosti

Na měřených objektech často sledujeme více znaků, hodnoty znaků spolu mohou souviset, navzájem se ovlivňovat.

Některé vztahy jsou zřejmé (např. souvislost tělesných rozměrů, systolického a diastolického krevního tlaku, krevního tlaku a věku atd.), jiné se snažíme popsat.

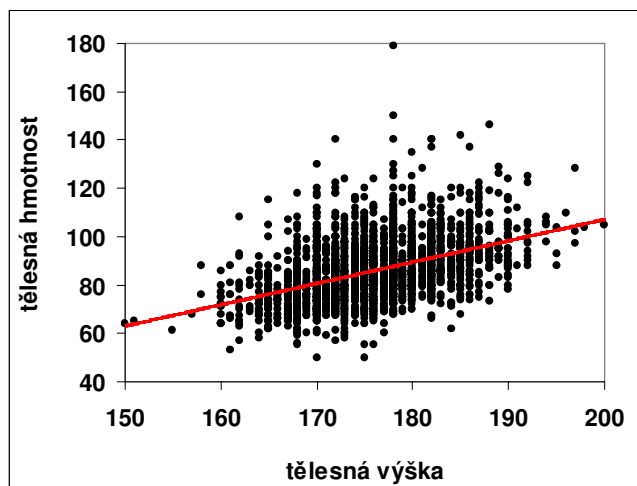
Předmětem statistické analýzy je popsat sílu závislosti a typ závislosti. **Sílu závislosti (korelaci)** vyjadřujeme pomocí různých měř závislosti, např. pomocí korelačních koeficientů. Hodnota korelačních koeficientů se obvykle pohybuje mezi nulou a jedničkou, resp. mezi hodnotou -1 a +1, nuly dosahuje při nezávislosti, jedničky při funkční závislosti.

Při měření závislosti dvou kvantitativních znaků můžeme druh a sílu závislosti posoudit graficky pomocí **bodového grafu**. **Typ závislosti** odhadujeme pomocí křivky, která se dá proložit systémem bodů v bodovém grafu. Bodový graf odhalí také extrémní hodnoty, které mohou mít na posuzování závislosti zásadní význam.

Příklad

Bodový graf znázorňuje závislost (korelaci) dvou kvantitativních znaků - tělesné hmotnosti a tělesné výšky. Každý bod odpovídá výsledkům měření u konkrétní osoby, na osu x se vynáší hodnota tělesné výšky, na osu y hodnota tělesné hmotnosti.

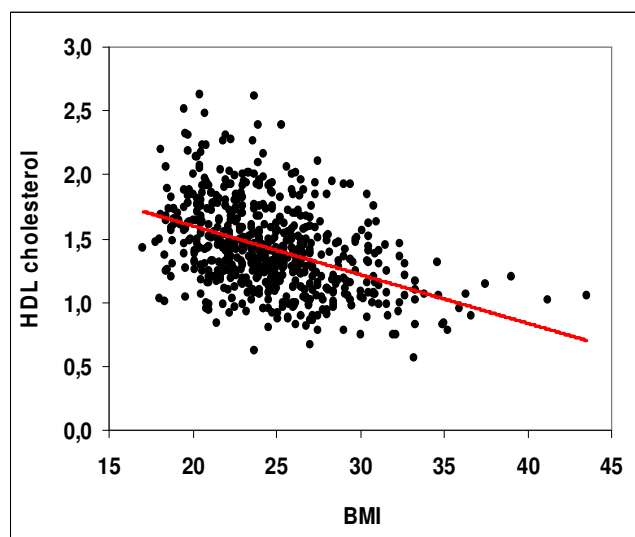
Z grafu je zřejmé, že mezi oběma znaky existuje **přímá lineární závislost**, systém bodů v grafu se dá proložit přímkou, která má rostoucí tendenci (kladnou směrnici).



Příklad

Bodový graf znázorňuje závislost dvou kvantitativních znaků – HDL cholesterolu a celkového cholesterolu.

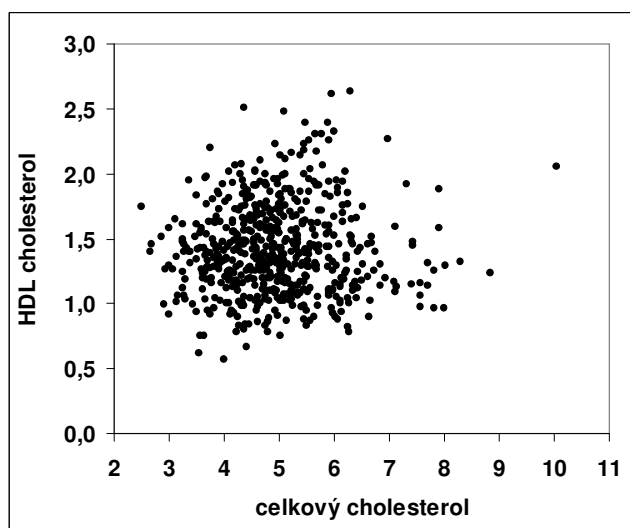
Z grafu je zřejmé, že mezi oběma znaky existuje **nepřímá lineární závislost**, systém bodů v grafu se dá proložit přímkou, která má klesající tendenci (zápornou směrnici).



Příklad

Bodový graf znázorňuje vztah mezi hodnotami HDL cholesterolu a celkového cholesterolu.

Z grafu je zřejmé, že mezi oběma znaky neexistuje **žádná závislost**, systém bodů v grafu se nedá proložit křivkou, která by situaci popsala.



15.1. Pearsonův korelační koeficient

U lineární závislosti můžeme sílu závislosti vyjádřit pomocí Pearsonova korelačního koeficientu, který se vypočítá podle vzorce

$$r_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) / s_X s_Y$$

X a Y je označení znaků, jejichž závislost posuzujeme, \bar{x} a \bar{y} jsou průměry znaků, s_X a s_Y jsou směrodatné odchylky znaků, n je počet párových hodnot.

Hodnoty Pearsonova korelačního koeficientu se pohybují od -1 do +1:

- hodnoty r blízké nule znamenají, že mezi znaky X a Y není lineární závislost,
- hodnoty záporné ($r < 0$) ukazují na nepřímou lineární závislost,
- hodnoty kladné ($r > 0$) ukazují na přímou lineární závislost.

Čím vyšší je hodnota korelačního koeficientu, tím větší je míra lineární závislosti znaků.

V případě přítomnosti extrémních hodnot v datech nebo výrazné nenormality rozložení hodnot jednoho nebo obou znaků je vhodnější mírou závislosti neparametrický **Spearmanův korelační koeficient**

Příklad

Výsledky korelační analýzy (statistický program SPSS)

V následující tabulce jsou výsledky Pearsonovy korelační analýzy. Trojici hodnot v každé buňce tabulky tvoří hodnota Pearsonova korelačního koeficientu (hvězdičkou jsou označeny korelační koeficienty, které jsou statisticky významně nenulové), dalším údajem je hladina signifikance testu významnosti korelačního koeficientu (tj. p -hodnota) a N , tj. počet párových hodnot.

V tomto příkladu jde o posouzení závislosti mezi výsledky testů z fyziky, chemie a biologie u studentů, kteří zúčastnili přijímacího řízení na Lékařskou fakultu. Korelovány byly počty bodů dosažené studenty v jednotlivých testech.

Tabulka: Korelační matice – výstup ze statistického programu SPSS

Correlations

		FYZIKA	CHEMIE	BIOLOGIE
FYZIKA	Pearson Correlation	1,000	,881**	,874**
	Signifikance	,	,000	,000
	N	1257	1257	1257
CHEMIE	Pearson Correlation	,881**	1,000	,860**
	Signifikance	,000	,	,000
	N	1257	1257	1257
BIOLOGIE	Pearson Correlation	,874**	,860**	1,000
	Signifikance	,000	,000	,
	N	1257	1257	1257

** . Correlation is significant at the 0.01 level (2-tailed).

Nejvyšší míra přímé lineární závislosti (nejvyšší hodnota korelačního koeficientu) byla zjištěna mezi hodnotami bodového skóre testu z fyziky a testu z chemie ($r = 0,881$), nejnižší hodnota korelačního koeficientu byla zjištěna mezi bodovým skóre testu z chemie a biologie ($r = 0,860$). Všechny zjištěné závislosti lze klasifikovat jako silné, hodnota korelačního koeficientu je vyšší než 0,8.

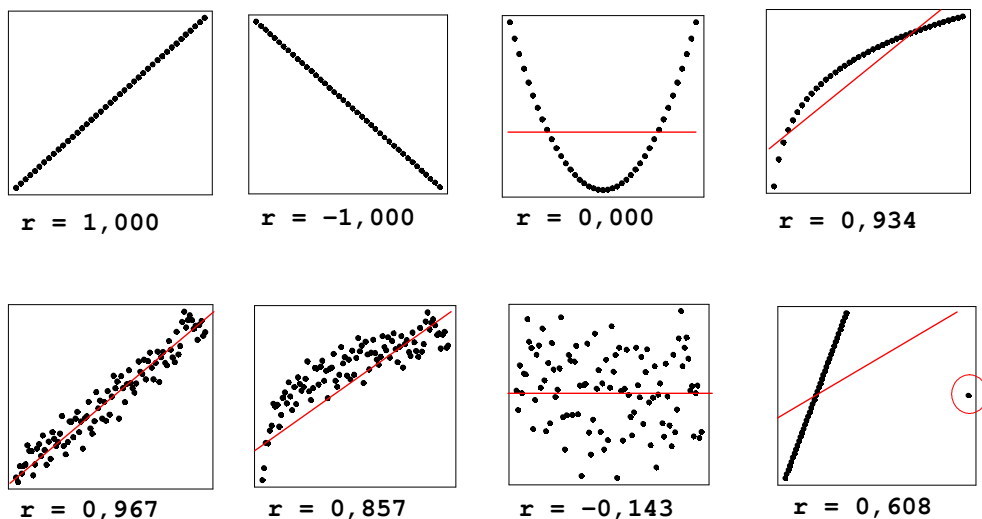
Test významnosti korelačního koeficientu

Je-li korelační koeficient r statisticky významně nenulový (signifikance pro test hypotézy $H_0: r = 0$ je $< 0,05$), používají se ke slovnímu vyjádření síly vztahu tyto hranice:

0,1 - 0,3 korelace slabá
 0,4 - 0,6 střední
 0,7 - 0,8 silná
 nad 0,9 velmi silná.

Tvary závislostí

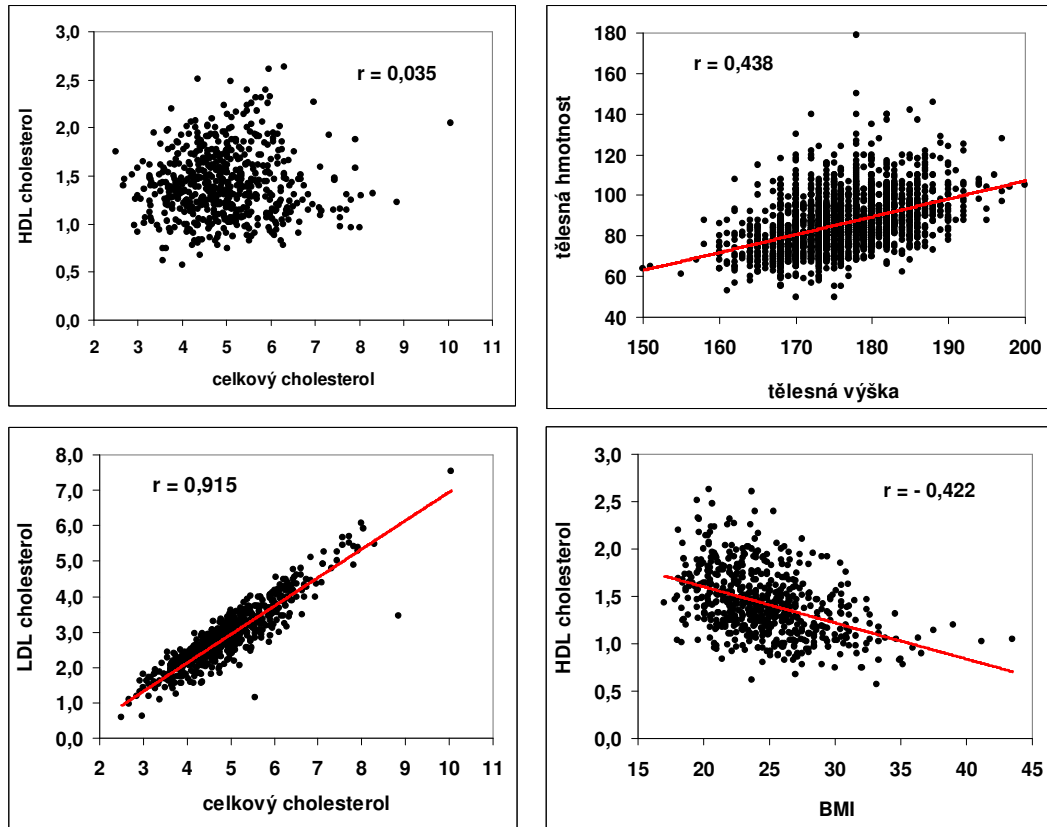
Pearsonův koeficient lineární závislosti r



První čtyři grafy zobrazují funkční závislosti (body jsou zarovnané do přímky nebo křivky). Na prvním grafu je zobrazena přímá absolutní lineární závislost ($r = 1,0$), nepřímá absolutní lineární závislost ($r = -1,0$), funkční kvadratická závislost ($r = 0$), funkční logaritmická

závislost ($r = 0,934$). Další tři grafy zobrazují postupně přímá lineární závislost ($r = 0,967$), logaritmická závislost ($r = 0,857$), žádná závislost ($r = -0,143$) a poslední graf přímou absolutní (funkční) lineární závislost s jednou odlehlou hodnotou, která výrazně snižuje hodnotu korelačního koeficientu ($r = 0,608$).

Bodové grafy a odpovídající hodnoty Pearsonova korelačního koeficientu.



Kritické hodnoty pro posouzení významnosti korelačního koeficientu

Vypočtenou hodnotu korelačního koeficientu r lze porovnat s tabulkovou kritickou hodnotou. Pokud bylo k dispozici n párových hodnot, má vypočtený korelační koeficient $n-2$ stupňů volnosti.

Přesahuje-li absolutní hodnota korelačního koeficientu r příslušnou kritickou hodnotu, nulovou hypotézu $H_0 : r = 0$ zamítáme.

Z tabulky je zřejmé, že pro malá n je k zamítnutí H_0 zapotřebí vyšších hodnot korelačního koeficientu r a pro velký počet párových hodnot stačí k zamítnutí nulové hypotézy poměrně nízká hodnota korelačního koeficientu.

Stupně volnosti n-2	Hladina významnosti α 5 %
1	0,996
2	0,950
3	0,878
4	0,811
5	0,754
10	0,576
15	0,482
20	0,423
25	0,381
30	0,349
35	0,325
40	0,304
45	0,288
50	0,273
60	0,250
70	0,232
80	0,217
90	0,205
100	0,195

Tabulka kritických hodnot Pearsonova korelačního koeficientu

15.2. Spearmanův pořadový koeficient korelace

Spearmanův korelační koeficient se používá v případě nenormálního rozložení hodnot jednoho nebo obou kvantitativních znaků, jejichž závislost chceme posuzovat. Používá se také v případech výskytu extrémních hodnot v datech. Předpokládaný vztah obou kvantitativních znaků přitom nemusí být ani lineární, může být trvale vzestupný nebo sestupný.

K výpočtu Spearmanova korelačního koeficientu se nepoužívají vlastní naměřené hodnoty znaků, ale používá se pořadí těchto hodnot. To znamená, že sledované osoby uspořádáme podle hodnot jednoho sledovaného znaku X a přiřadíme jim pořadovou hodnotu q_i . Stejným způsobem je uspořádáme podle hodnot druhého sledovaného znaku Y a přiřadíme jim pořadovou hodnotu r_i . Každá sledovaná osoba má tedy přidělena dvě pořadová čísla, jedno číslo odpovídá pořadí v uspořádání podle hodnot znaku X a druhé pořadové číslo odpovídá pořadí v uspořádání podle hodnot znaku Y .

Kdyby s rostoucími hodnotami X vzrůstaly i hodnoty Y , byla by zřejmě pořadí obou veličin shodná, tj. $q_i = r_i$ pro každou osobu i .

Jestliže s rostoucími hodnotami X klesají hodnoty Y , jsou pořadí obou veličin právě opačná. Při nezávislosti jsou pořadí zpřeházena zcela náhodně.

Pro výpočet hodnoty korelačního koeficientu se použije vzorec

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

kde d_i^2 je druhá mocnina rozdílu pořadí pro osobu i , tj. $(q_i - r_i)^2$.

Je-li pořadí X shodné s Y , jsou všechny rozdíly nulové a koeficient r_s nabývá hodnoty 1.

Je-li pořadí přesně opačné, je $r_s = -1$.

Hodnoty korelačního koeficientu blízké nule naznačují, že pořadí jsou náhodně zpřeházená a mezi sledovanými veličinami tedy není závislost.

Kritické hodnoty Spearmanova korelačního koeficientu pro oboustrannou alternativu (tj. alternativní test $H_A: r \neq 0$) jsou uvedeny ve statistických tabulkách pouze pro $n \leq 30$. Pro větší

rozsah výběru můžeme použít testové statistiky $Z = r_s \sqrt{n-1}$, která má za platnosti nulové hypotézy ($H_0: r=0$) standardizované normální rozdělení. Jestliže tedy vyjde $|Z| \geq$ než kvantil $z_{1-\alpha/2}$, zamítáme nulovou hypotézu ve prospěch alternativy na zvolené hladině významnosti α .

15.3. Příklad

(viz učebnice Zvárová)

Na interním oddělení bylo léčeno na určitou chorobu $n=10$ pacientů. Za čtyři týdny po ukončení hospitalizace byli pacienti pozváni ke kontrolní prohlídce, při níž se mimo jiných údajů sledovala též sedimentace červených krvinek. Závažnost klinického průběhu onemocnění pacientů lze vyjádřit na základě uspořádání podle závažnosti zjištěných klinických příznaků do řady od 1 (nejlehčí průběh) až po 10 (nejtěžší průběh).

Ověřte na hladině významnosti 0,05, zda hodnota sedimentace nezávisí na klinickém průběhu onemocnění. Hodnocení pacientů podle závažnosti klinického průběhu onemocnění, zjištěné hodnoty sedimentace a pořadí podle hodnot sedimentace jsou uvedeny v tabulce.

Pacient	1	2	3	4	5	6	7	8	9	10
Pořadí pacienta podle klinického průběhu (q_i)	4	6	1	5	10	2	7	3	9	8
Hodnota sedimentace	19	22	26	30	33	23	28	21	59	39
Pořadí pacienta podle hodnoty sedimentace (r_i)	1	3	5	7	8	4	6	2	10	9
Diference pořadí ($d_i = q_i - r_i$)	3	3	-4	-2	2	-2	1	1	-1	-1
Čtverec diference (d_i^2)	9	9	16	4	4	4	1	1	1	1

Výpočet hodnoty korelačního koeficientu

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 650}{10(100 - 1)} = 1 - \frac{3900}{990} = 0,697$$

Kritická hodnota Spearmanova korelačního koeficientu pro spolehlivost $\alpha = 0,05$ a $n = 10$ je 0,6364 (tabulka kritickým hodnot viz např. Zvárová).

$r_s = 0,697 > 0,6364$, proto nulovou hypotézu zamítáme na 5% hladině významnosti a můžeme říci, že jsme prokázali statisticky významnou závislost mezi hodnotou sedimentace a klinickým průběhem onemocnění.

16. Regresní analýza

Pokud z bodového grafu zjistíme, že mezi sledovanými znaky existuje nějaký typ závislosti (lineární, exponenciální, logaritmický, atd.), můžeme se pokusit popsat tuto závislost pomocí matematického modelu (rovnice).

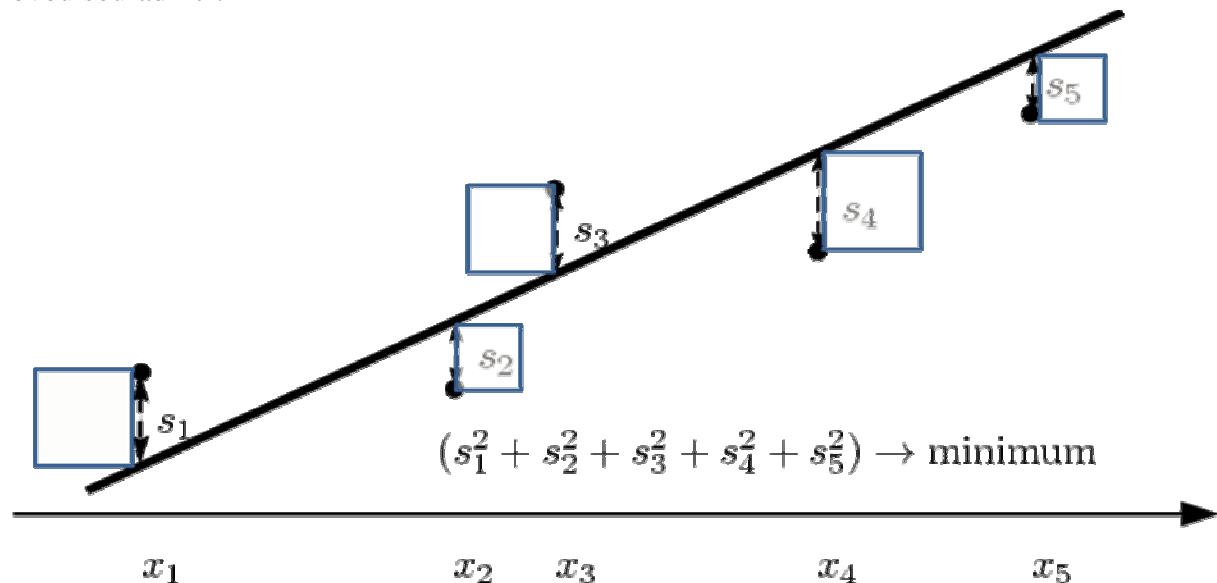
Lineární závislost znaků X a Y popisujeme pomocí rovnice přímky (tj. pomocí jednoduchého lineárního regresního modelu)

$$Y = a + bX,$$

Proměnnou na levé straně rovnice (Y) nazýváme **závisle proměnná**, proměnnou na pravé straně rovnice (X) **nezávisle proměnná**, a a b jsou koeficienty, které jsou předmětem odhadování v regresní analýze. Koeficienty regresní přímky se odhadují pomocí metody nejmenších čtverců (MNC).

Metoda nejmenších čtverců

slouží k nalezení takového řešení (tj. rovnice přímky), aby byl součet druhých mocnin chyb nalezeného řešení minimální. Čtvercem chyb přitom rozumíme druhou mocninu rozdílu y-ové souřadnice „naměřeného“ bodu a y-ové souřadnice bodu na přímce, oba body mají stejnou x-ovou souřadnici.



Parametr b (regresní koeficient) udává o kolik se změní závisle proměnná, zvětší-li se hodnota nezávisle proměnné o jednotku.

Nezávisle proměnných může být v rovnici více, vícenásobný lineární regresní model (pro odhad závisle proměnné Y pomocí více nezávisle proměnných) má tvar

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_kX_k.$$

Postup regresní analýzy:

- 1) sestrojíme bodový graf a posoudíme typ závislosti
- 2) zvolíme vhodný regresní model a odhadneme koeficienty modelu
- 3) zhodnotíme kvalitu nalezeného řešení

V bodovém grafu vynášíme hodnoty závisle proměnné na svislou osu, hodnoty nezávisle proměnné na osu vodorovnou. Typ regresního modelu volíme na základě posouzení bodového grafu, podle křivky, která se dá proložit systémem bodů a dobře charakterizuje typ závislosti.

Hodnocení kvality modelu

- z grafického vyjádření
- posouzením reziduálního (zbytkového) rozptylu

$$s^2 = \frac{\sum_{i=1}^n (y_i \text{ naměřené} - y_i \text{ odhad})^2}{(n-2)}$$

- posouzení koeficientu determinace R^2 , který je mírou vhodnosti modelu. Udává jaké procento variability závisle proměnné, lze vysvětlit pomocí modelu. Vyjde-li $R^2=0,87$ znamená to, že regresní model vysvětluje 87 % variability závisle proměnné.

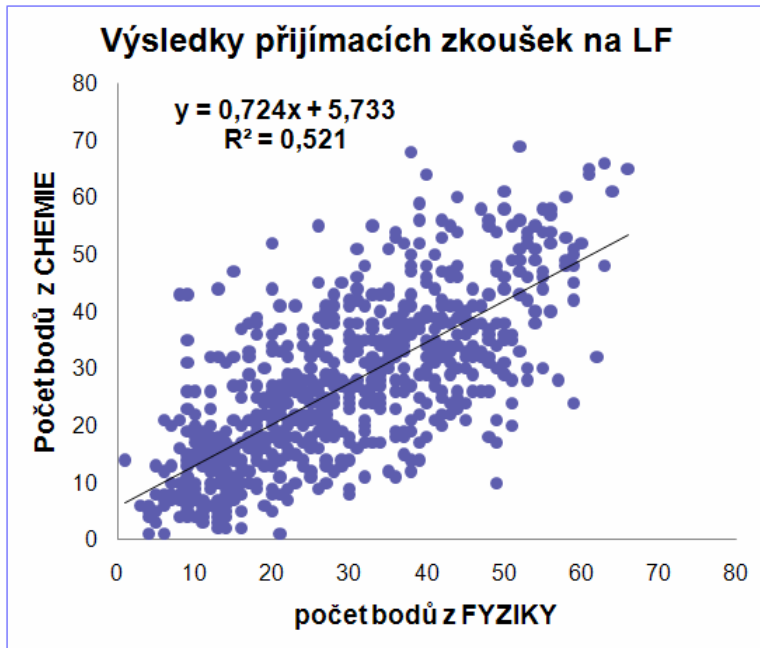
V jednoduchém lineárním modelu (s jednou nezávisle proměnnou X) je $R^2 = r^2$ (tj. druhá mocnina Pearsonova korelačního koeficientu).

16.1. Příklad

Na základě výsledků přijímacích zkoušek na LF pro obor všeobecné lékařství popište vztah mezi počtem bodů z testu z chemie a počtem bodů z testu z fyziky.

Řešení

Pro posouzení závislosti sestrojíme nejdříve bodový graf, počet bodů z chemie vynášíme na osu x a počet bodů z testu z fyziky vynášíme na osu y.



Z grafu je patrné, že mezi bodovým skóre obou testů je přímá lineární závislost (systémem bodů se dá proložit přímka s kladnou směrnici).

Pro posouzení míry lineární závislosti vypočítáme hodnotu Pearsonova korelačního koeficientu: $r = 0,722$.

Závislost mezi bodovým skóre se můžeme pokusit popsat pomocí lineárního regresního modelu, tj. pomocí rovnice přímky $Y = a + bX$, kde:

Y ... počet bodů z fyziky

X ... počet bodů z chemie

a, b ... koeficienty regresní přímky

Výsledná regresní rovnice:

body z CHEMIE = 0,724 . body FYZIKA + 5,733.

Poznámka: Řešení lineární regresní analýzy v programu MS Excel

- 1) Závislost dvou kvantitativních znaků, např. závislost viskozity na teplotě zobrazíme graficky pomocí bodového grafu (údaje o teplotě budou na ose x a hodnota viskozity na ose y).
- 2) Po zobrazení grafu klikneme pravým tlačítkem myši na body v grafu a z místní nabídky vybereme položku *Přidat spojnici trendu...*

- 3) V nabídce *Typ trendu a regrese* necháme nastavení *Lineární* a zaklikneme možnost *Zobrazit rovnici regrese* a *Zobrazit hodnotu spolehlivosti*.

Systémem bodů v grafu se proloží regresní přímka a zobrazí se její rovnice ve tvaru:
 $y = bx + a$.

K výpočtu směrnice (b) a absolutního členu (a) regresní přímky $y = bx + a$, můžeme v programu MS Excel použít také statistické funkce:

funkce INTERCEPT(pole_y;pole_x) vypočítá hodnotu absolutního členu a

funkce SLOPE(pole_y;pole_x) vypočítá hodnotu směrnice přímky.

Do okna pole_y zadáme adresu oblasti buněk s hodnotami závisle proměnné (viskozita) a do okna pole_x zadáme adresu oblasti buněk s hodnotami nezávisle proměnné (teplota). Zadáváme jen adresy buněk, které obsahují čísla.

Kontrolní otázky:

1. K čemu se používá Pearsonův korelační koeficient, jakých hodnot může nabývat?
2. V jaké situaci se používá Spearmanův korelační koeficient, jakých hodnot může nabývat?
3. Vysvětlete základní princip lineární regresní analýzy.
4. Co je to regresní koeficient, jakou metodou se odhaduje?
5. Definujte pojem koeficient determinace, jakou souvislost má s Pearsonovým korelačním koeficientem?
6. Jak se hodnotí kvalita lineárního regresního modelu?

17. Typy studií

Design studie a jeho volba je velmi komplexní problematika. Design je ovlivněn především cílem studie (co by chtěl zadavatel studií prokázat), klinickými aspekty (jakým parametrem je možné splněné cíle studie u zkoumané populace sledovat) a statistickými aspekty.

Přehled hlavních designů udává následující tabulka

Klinické experimentální studie (studie s terapeutickou intervencí)		
Studie s kontrolou	Paralelní (souběžná) kontrola	Randomizovaná (paralelní) klinická studie
		Nerandomizovaná klinická (paralelní) studie
	Sekvenční kontrola	Self-controlled design
		Cross-over design
Externí nebo historická kontrola (nebezpečí „bias“)		
Studie bez kontroly		
Observační studie (studie bez terapeutické intervence)		
Case-series (deskriptivní) studie		
Case-control studie (retrospektivní – „Co se stalo?“)		
Cross-sectional studie (prevalence – „Co se děje?“)		
Kohortová studie (prospektivní – „Co se stane?“)		
Historická kohortová studie		

17.1. Klinická studie

Klinická studie (experimentální studie v humánní medicíně) je charakterizována intervencí, tj. léčbou s použitím farmak, nebo použitím jiného léčebného procesu (např. rehabilitace). Typické dělení klinických studií je dělení na **studie s kontrolou** a **studie bez kontroly**. Studie s kontrolou musí mít alespoň dvě ramena (experimentální a kontrolní). Standardem v klinickém výzkumu je randomizovaná klinická studie s alespoň jedním experimentálním ramenem a alespoň jedním kontrolním ramenem, kdy léčba v obou ramenech probíhá paralelně (souběžně). Obecně není doporučováno použití tzv. historické kontroly (tj. kontroly z jiné studie) kvůli velmi pravděpodobnému bias (výchylce, rozdílu mezi zkoumanými populacemi a výběry). Jediná možnost, jak bias omezit, je randomizace.

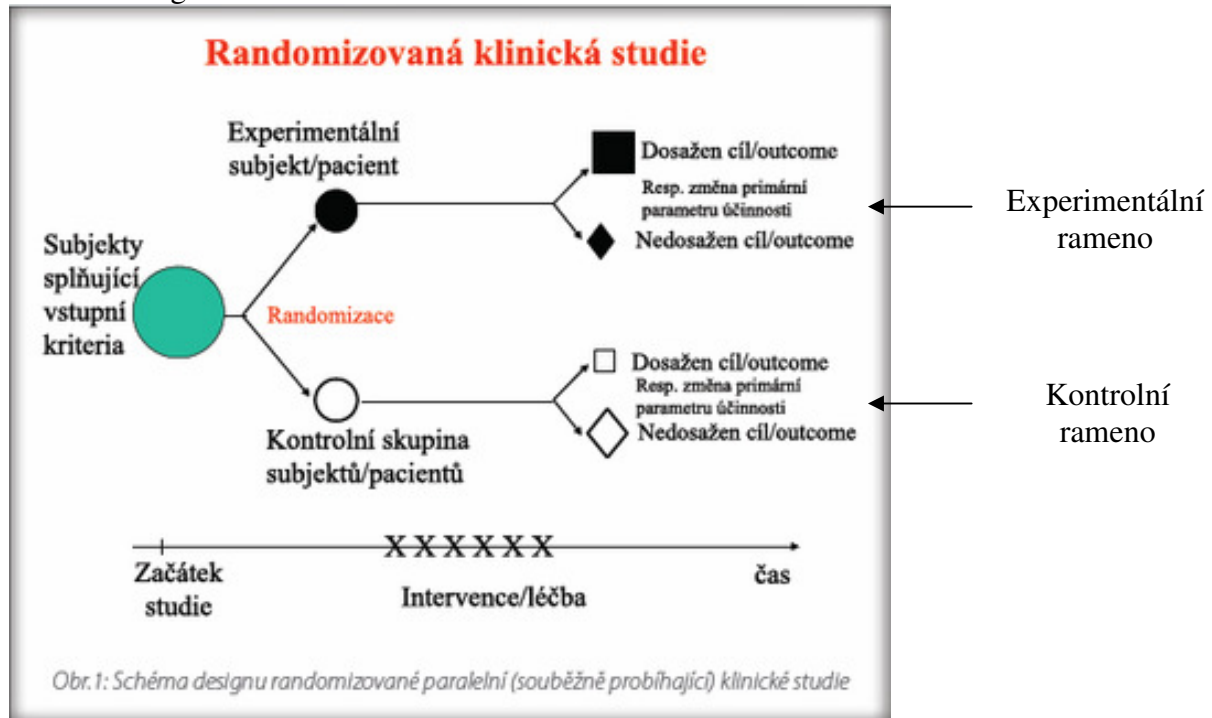
Randomizace

Randomizace znamená náhodné rozdělení výběru z jedné populace do dvou (nebo více) ramen, tj. náhodné rozdělení osob do kontrolní a experimentální skupiny. V případě randomizace je každá disproporce mezi rameny (bias) jen náhodná a každé rameno by mělo mít vyvážené zastoupení pohlaví, věku, stádií nemoci a dalších i nesledovaných rušivých faktorů.

Otevřená, jednoduše zaslepená, dvojité zaslepená a trojitě zaslepená randomizovaná klinická studie

Randomizovaná studie může být **otevřená** (pacient i lékař ví do kterého ramene studie je randomizován, tj. jaká léčba je použita), **jednoduše zaslepená** (pacient neví, jak je randomizován, ale jeho ošetřující lékař ano), **dvojité zaslepená** (neví to pacient ani lékař) nebo **trojitě zaslepená** (neví to ani statistik, který studii vyhodnocuje, ten zná jen údaj o randomizaci typu A nebo B).

Schéma designu randomizované klinické studie



Randomizovaná klinická studie je charakterizována tím, že léčba v obou ramenech probíhá paralelně a každý pacient je léčen jen jednou z randomizovaných terapií.

17.2. Zkřížená studie (cross-over design)

Často užívaný je tzv. zkřížený (cross-over) design, kdy pacient v randomizovaném pořadí dostane postupně obě terapie. Cross-over design je velmi výhodný protože umožňuje posoudit individuální odpověď každého pacienta na obě léčby a porovnat efekt v první a druhé periodě, tj. posoudit jak se efekt léčby mění s časem.

Proč se cross-over design nepoužívá častěji?

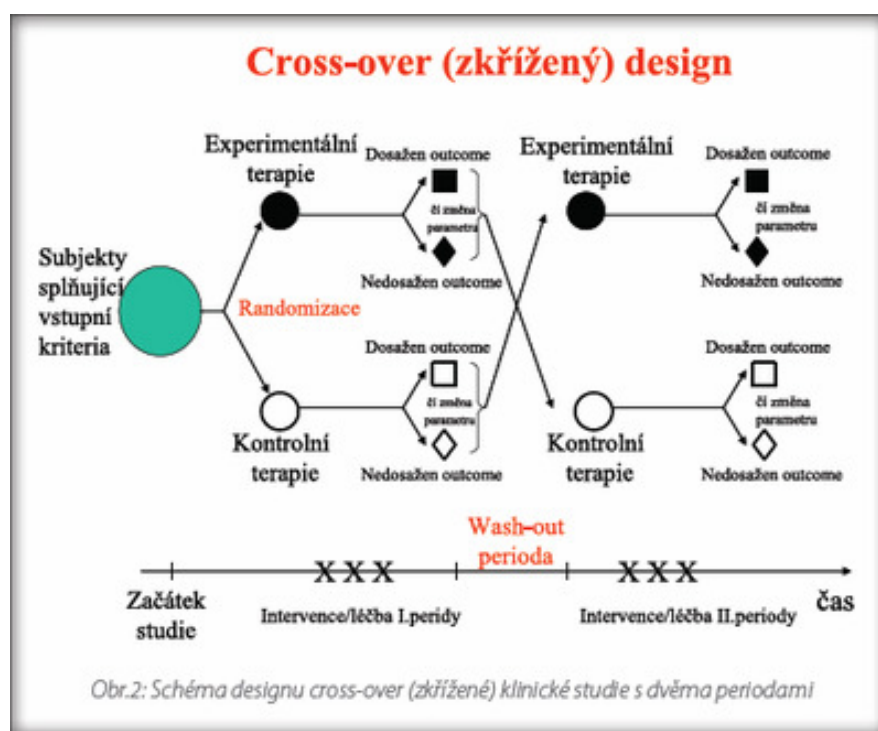
Použití tohoto designu je spojeno s problémem, který omezuje jeho použitelnost:

- 1) Pacient může být po první periodě již „vyléčen“ a pak by druhá perioda nebyla smysluplná (či etická). Optimální je, aby byl pacient na začátku každé z period v přibližně stejném stavu. To lze zaručit u farmakokinetických studií při přiměřeně dlouhé wash-out periodě k vyloučení účinné látky z organismu.
- 2) Etický důvod - může se jednat o vážnou diagnózu a pacientovi, který je zjevně responder na terapii v první periodě, by bylo neetické ji vysadit a změnit na terapii druhé periody.
- 3) Příliš dlouhý carry-over effect (přenos léčebného efektu první periody i po vysazení léčby až do druhé periody) – efekt prvního léku by se mísil s efektem léku ve druhé periodě.

Výhody a nevýhody zkřížené studie

- Porovnání účinku na stejných osobách je přesnější
- Vyžaduje menší počet respondentů
- Lze studovat jen krátkodobé účinky léků, studie jsou vhodné pouze pro studium chronických stabilních onemocnění
- Odstoupení pacientů ze studie snižuje kvalitu studie
- Možnost nežádoucího přenesení účinku léku z 1. fáze do 2. fáze studie
- Pořadí podání léků může ovlivnit jejich účinek

Schéma designu cross-over klinické studie



Wash-out perioda – je perioda bez léčby mezi léčebnými periodami I a II, musí se určit dostatečně dlouhá tak, aby nebyl ovlivněn účinek přípravku podaného ve 2. fázi.

Realizace klinických studií

Klinické studie, jejichž záměrem je testování nových léků, se realizují jako dvojitě zaslepené, placebem kontrolované pokusy. Dvojitě zaslepený pokus zabraňuje zkreslení výsledků ze strany měřeného subjektu (pacienta) i ze strany měřitele (lékaře). **Placebo** je neúčinná látka, stejného tvaru, barvy a chuti jako zkoušený lék. V případech, kdy nelze placebo použít, porovnáváme účinky nového léku oproti léku standardně používanému.

Fáze klinické studie

Po předklinických zkouškách v laboratoři.

1. První podání látky lidským subjektům - stanovení vhodné velikosti dávky nového léku, snášenlivosti a základních údajů pro farmakokinetiku.
2. Zjištění léčebné účinnosti léku, ověření vhodnosti zvažovaných indikací. Registrace výskytu případných nežádoucích účinků.
3. Fáze potvrzující účinnost a bezpečnost léčiva – s velkým počtem pacientů, často mezinárodní charakter studie. Hodnocení je srovnávací, buď s placebem, nebo standardně používaným léčivem. Fáze je podkladem pro registraci nového léku.
4. Sledování léčiva při používání v široké klinické praxi včetně nežádoucích účinků. Ověřování účinku v dalších menších studiích (často marketingového charakteru) na nových populacích.

Příklad klinické studie - testy nového léku na leukémii

V Česku bylo ukončeno první klinické testování nové účinné látky GS-9219, kterou objevil český vědec profesor Antonín Holý. Přípravek dává naději, že bude účinný při léčbě některých druhů leukémie. Testování probíhalo v několika vybraných klinikách: FN Olomouc, Všeobecná FN v Praze, FN Královské Vinohrady, Brno a Hradec Králové.

Jednalo se o první fázi testování, kdy se zkoušela toxicita, tolerance a snášenlivost léku pro pacienty a hledala se nejhodnější (netoxická) dávka pro člověka. Teprve v dalších fázích se by se testoval protinádorový účinek léku. Podle stejných zásad se přípravek testoval i ve Spojených státech na pěti špičkových pracovištích (celkem 60 pacientů).

Profesor Antonín Holý



Látka GS-9219 v testech na psech odborníky velmi překvapila, byla schopná zničit rakovinové buňky během pěti dnů od podání jediné infuze.

Unikátnost látky spočívá v tom, že likvidovala pouze nádorové buňky a zdravých si neвшímala. Pokud by byl lék úspěšný na lidech, bylo by možné s ním začít léčit nejdříve za šest až osm let.

V první fázi testování však bylo zjištěno, že je látka pro lidské subjekty příliš toxická, proto bylo testování zastaveno a další fáze se již nerealizovaly.

Studie bez kontroly

Jsou to studie, kde například chceme prokázat, že léčebná odpověď nastane u většího procenta případů, než je předem definované procento (např. 60%). Nebo chceme prokázat, že procento nežádoucích účinků je menší než nějaké předem stanovené procento. **Myslí se tím ovšem statisticky významně (větší, či menší)!**

Studie bez intervence

Jsou to studie bez jakéhokoliv terapeutického postupu, uplatňují se zde však diagnostické postupy včetně např. biopsie.

Tři základní designy těchto studií:

- **case-control** studie (český termín „případ-kontrola“ se obvykle nepoužívá),
- **kohortová** studie a
- **průřezová studie** (cross-sectional).

Case-control design

porovnává dvě populace na základě výběru z každé z nich a to případy (pacienty=cases) a kontroly (srovnávací skupinu).

Oba dva výběry by měly být vyvážené z hlediska věku či zastoupení pohlaví. I tak ale může být mezi výběry bias, který nesouvisí s tím, kdo patří mezi případy a kdo mezi kontroly.

V case-control studiích nás zajímá otázka:

„Co vedlo k tomu, že se u někoho vyvinula určitá diagnóza?“

Tento design patří mezi retrospektivní, používá se hlavně při zjišťování příčin choroby, rizikových faktorů a také při evaluaci (hodnocení) diagnostických postupů.

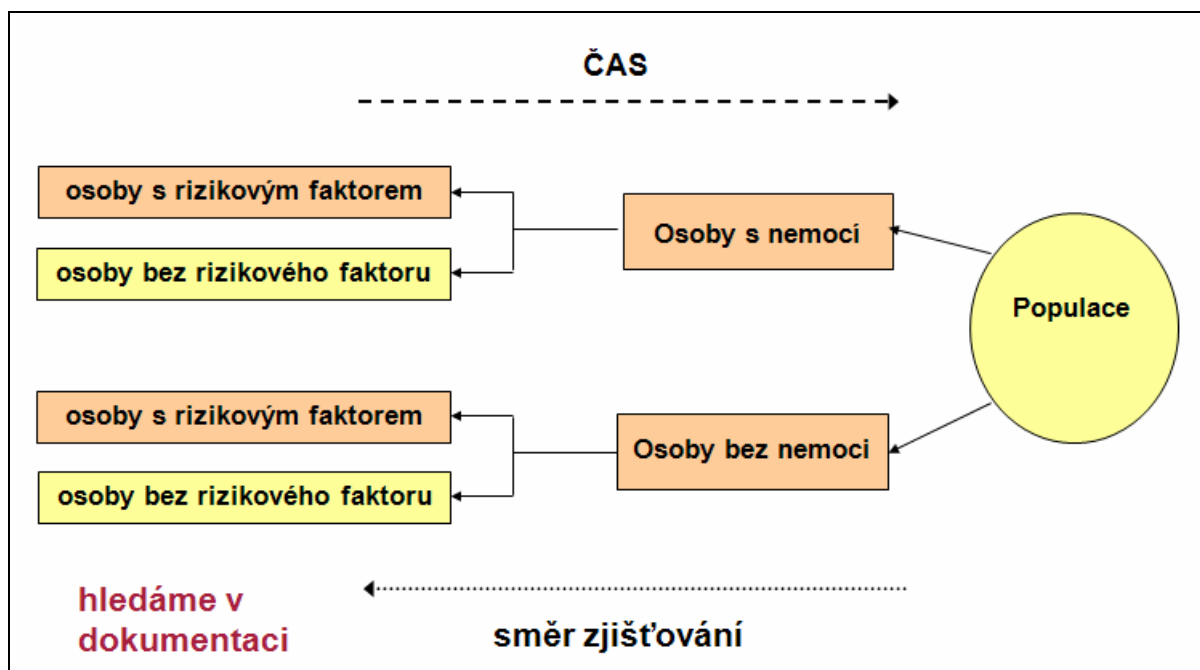
17.3. Retrospektivní studie

(také studie případů a kontrol)

V tomto typu studií testujeme hypotézu o možné příčině nemoci (tj. vycházíme od nemoci a jdeme nazpět k příčině nemoci).

- Vytvoříme dva soubory - soubor osob s nemocí a soubor osob bez nemoci, tj. soubor kontrolní.
- U obou souborů zjišťujeme (např. z dokumentace) podíl osob, které přišly do styku s daným rizikovým faktorem.

Schéma designu retrospektivní studie



Vytvoření kontrolního souboru

Kontrolní soubor, tj. soubor osob bez nemoci, by měl být výběrem ze stejné populace, jako soubor sledovaný.

Tvoří-li sledovaný soubor pacienti dané nemocnice, vybíráme velmi často do kontrolního souboru pacienty z téže nemocnice, ale s jinou diagnózou.

Důvod:

je-li sledovaný soubor zkreslen oproti základní populaci (kvalitou bydlení, typem zaměstnání, atd.), bude stejně zkreslen také soubor kontrolní.

Metoda párového vyvažování (matching)

způsob jak dosáhnout vzájemné srovnatelnosti sledovaného a kontrolního souboru

Ke každé osobě ze sledovaného souboru hledáme kontrolu se stejnými znaky (tj. pohlavím, věkem, zaměstnáním, atd.)

Použití: Je-li rozsah základní populace velký a výběry jsou malé

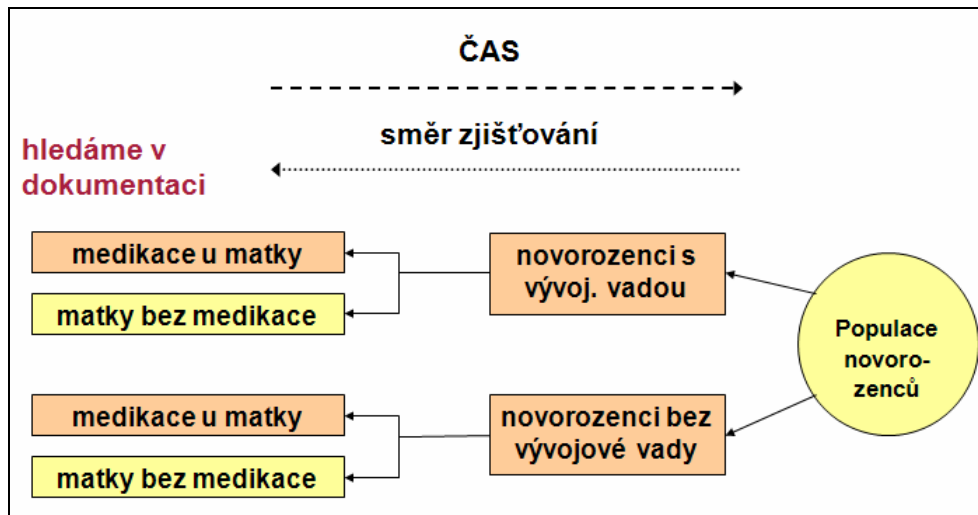
Výhody a nevýhody retrospektivní studie

- metoda je lépe technicky proveditelná a vystačí s menším rozsahem souborů
- řada potřebných údajů nemusí být v dokumentaci k dispozici

Příklady retrospektivních studií

- U souboru osob s karcinomem plic a bez karcinomu zjišťujeme kouření a přítomnost dalších rizikových faktorů
- Studium vztahu mezi vývojovými vadami u novorozenců a medikací matek v průběhu těhotenství

Příklad



17.4. Prospektivní (kohortová) studie

Na počátku máme výběr z jedné populace, u žádné z osob prozatím nebyla diagnostikována nemoc, jejíž výskyt sledujeme.

Vytvoříme dva soubory

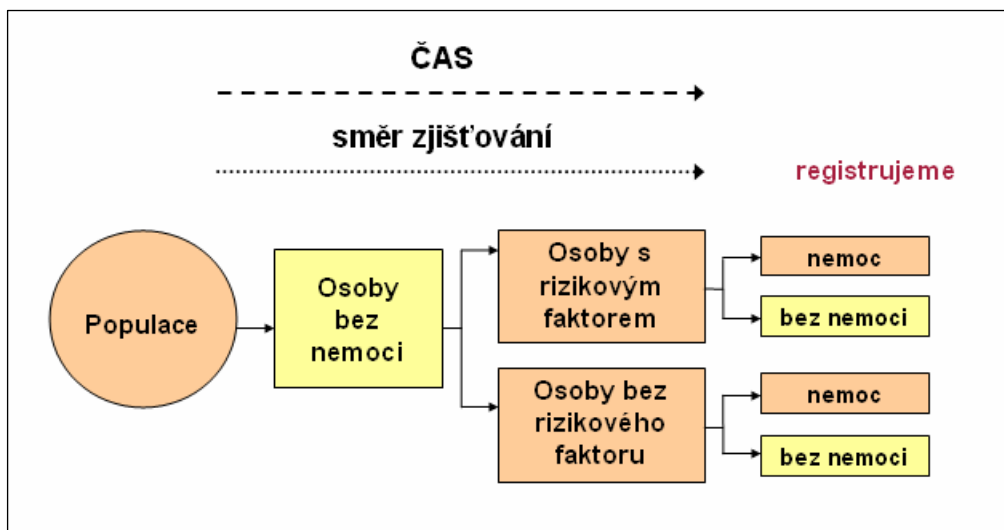
- soubor osob se sledovaným rizikovým faktorem a
- soubor osob bez rizikového faktoru.

Oba soubory sledujeme v čase a zjišťujeme, kolik osob onemocnělo. Časový horizont, kterého se týká tato otázka, může být velmi dlouhý, i několik desítek let.

Např. „Framingham Heart Study“ začala v roce 1948 a běží dosud (tato kohorta byla již dvakrát doplňována také proto, aby se ukázal vývoj v čase).

Kohortová studie se používá na zjištění příčin nemoci a nalezení prognostických (rizikových) faktorů vzniku onemocnění.

Schéma designu prospektivní studie



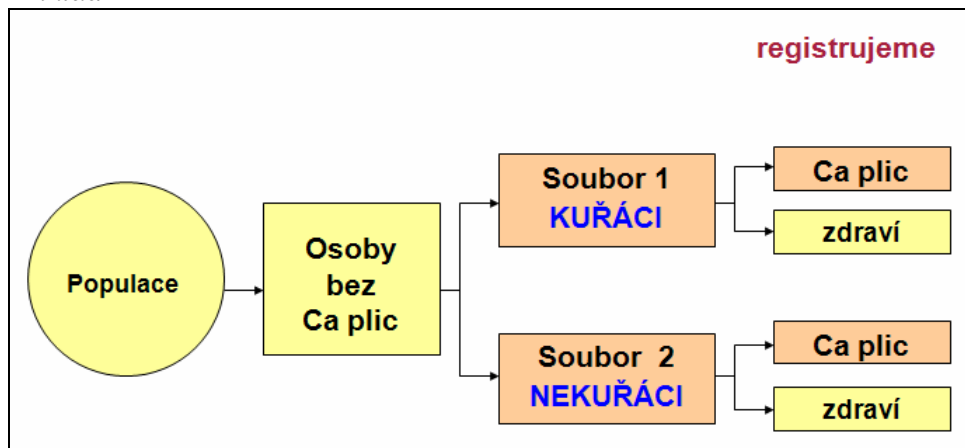
Výhody a nevýhody prospektivní studie

- studie umožňuje postupovat od příčiny k důsledku
- umožňuje plánovat a stanovit pevná kritéria
- značné nároky na čas a prostředky
- ztráta osob z evidence při déletrvajícím sledování
- změna zvyků a chování sledovaných osob

Příklady prospektivních studií

- 1) Prenatální rizika a atopie v dětském věku
 - studie souvislosti výskytu astmatu a atopie u dětí s
 - atopií v anamnéze matky
 - kouřením matky
 - počtem porodů matky
 - konzumací antioxidantů (vitamínu E) v těhotenství
- 2) Sledování vztahu mezi typem zaměstnání a vývojem varixů
- 3) Sledování vztahu mezi kouřením a výskytem rakoviny plic

Příklad



17.5. Průřezová (cross-sectional) studie

Tato studie na základě výběru z populace odpovídá na otázku: "co se děje právě teď".

Např. jaká je právě nyní incidence či prevalence nějakého onemocnění?

Často nás také zajímá zastoupení stádií a forem onemocnění v populaci v souvislosti s daným onemocněním. V libovolný okamžik provedeme náhodný výběr ze základní populace – to nám umožní učinit odhady o podílu nemocných osob a počtu osob s rizikovým faktorem.

Průřezové šetření můžeme provést na studii prospektivní i retrospektivní.

Příklad průřezové studie

Studie MONICA, epidemiologická studie zaměřená na kardiovaskulární onemocnění (probíhala ve 21 zemích Evropy).

V průběhu 10-ti let byl ve třech průřezových šetřeních zjišťován výskyt kardiovaskulárních rizikových faktorů (kouření, DM, vysoký TK, vyšší triglyceridy, porucha metabolismu lipidů) a úmrtnost na srdeční a cévní onemocnění. V každém průřezovém šetření bylo vyšetřeno kolem 30000 mužů a 30000 žen ve věku 35-64 let.

17.6. Longitudinální studie

Longitudinální studie je označení pro studie, v nichž jsou stejní respondenti zkoumáni, měření nebo dotazováni opakovaně v několika časových okamžicích – zpravidla několika let nebo několika desítek let.

Studie bývají většinou realizovány prospektivně. Typickým rysem je provádění opakovaných měření na stejných jedincích.

Multicentrická studie – studie, která probíhá ve více centrech (např. na klinikách různých nemocnic)

Chyby ve studiích

Výběrová chyba (sampling error) vzniká v důsledku toho, že neprovádíme úplné šetření, ale zkoumáme pouze výběr, tedy část cílové populace.

K **nevýběrové chybě** (non-sampling error) dochází v průběhu procesu měření, zjišťování a zpracování údajů zaviněním přístroje, výzkumníka či respondenta. Může být důsledkem neúplnosti odpovědí, chyb měření nebo ztráty již získané informace.

Podle povahy rozlišujeme **náhodné chyby** (random errors), které vznikají nepozorností respondenta či tazatele a přinášejí jak nadhodnocení, tak podhodnocení správné hodnoty znaku a **systematické chyby** (systematic errors), které jsou naopak výsledkem chybně formulované otázky nebo měřením nesprávně seřízeným přístrojem a působí ve vztahu ke správné hodnotě vždy jen jedním směrem. Za **formální chyby** v datech se považují takové chyby, kdy např. nekuřák odpovídá na otázku kolik vykouří cigaret.

Příklad chybně formulované otázky:

Otázka:

„Raději bys četl(a) zajímavou knihu, anebo by jsi jel(a) na nějaký pěkný výlet s ostatními dětmi?“

- Ano
 Ne

Na otázku není možné jednoznačně odpovědět, zvláště při nabídce odpovědí ano / ne.

Správné nabídky odpovědí při stejné formulaci otázky:

- četl(a) bych si knihu
 jel(a) bych na výlet
 nevím

Kontrolní otázky:

1. Vysvětlete základní principy klinických studií.
2. Vysvětlete princip zkřížené studie (cross-over designu), uveďte omezující důvody pro širší použití, uveďte výhody a nevýhody
3. Vysvětlete princip prospektivní studie, uveďte její výhody a nevýhody.
4. Vysvětlete princip průřezové studie, uveďte její výhody a nevýhody.
5. Vysvětlete princip retrospektivní studie, uveďte její výhody a nevýhody.
6. Definujte pojem slepý pokus, dvojité slepý pokus, trojitě slepý pokus, placebo.

Literatura

Učební text byl zpracován s podporou následujících zdrojů:

- Zvárová J.: Biomedicínská statistika I, Základy statistiky pro biomedicínské obory, KAROLINUM, 2004.
(učebnice je dostupná na webové adrese <http://ucebnice.euromise.cz>)
- Kasal P., Svačina Š. a kol.: Lékařská informatika, 1998. (kapitola Statistika).
- Hendl J.: Přehled statistických metod zpracování dat. Analýza a metaanalýza dat.
- Ladislav Pecen, Statistika nuda NENÍ! Pharm Business Magazine, červen/červenec 2008.