

OTÁZKY K TEORII ZÁKLADŮ STATISTIKY

Rozlišujeme data kvalitativní a kvantitativní. Zkuste popsat nějaké rozdíly mezi těmito typy dat: Jak vypadá základní popis takových dat a jaké (typické) testovací metody pro ně používáme?

Co to jsou charakteristiky polohy a charakteristiky variability, uveďte nějaké příklady.

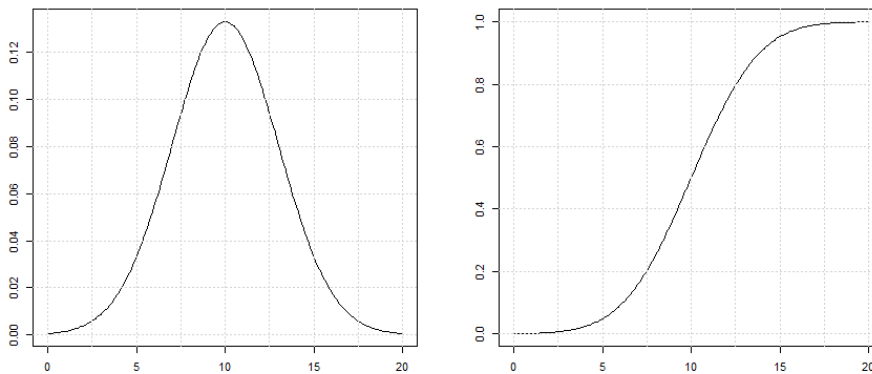
Rozdíl mezi průměrem, modem a mediánem.

Rozumět grafické prezentaci dat: histogram, boxplot. Kde je na grafu medián? Kdy mohu z grafu odečíst hodnotu průměru? Jak se na grafu projeví velká nebo malá variabilita datového souboru?

Induktivní statistika pracuje s pojmem náhodná veličina. Zkuste pojem vysvětlit, pomozte si příkladem (výška studentů biologie, barva pastelky).

Základní 2 vlastnosti dobrého výběru: které to jsou? (anglická zkratka „iid.“)

Ve statistické teorii se setkáváme s funkcí hustoty pravděpodobností a s distribuční funkcí. Přiřaďte názvy ke grafům. Popište osy obou grafů. Vyznačte na obou grafech pravděpodobnost, že náhodně vybraná veličina (subjekt) bude mít hodnotu zkoumané vlastnosti větší než 12.5 jednotky.



Co to je kvantil, kde se s ním setkáváme. Který z prvků zápisu se jmenuje kvantil: $P(X \leq x_{0,95}) = 0,95$?

Určete správné trojice:

Poissonovo rozdělení	$P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}$	Součet čtverců normovaných normálních veličin
Normální rozdělení	$P(X = 1) = p$	Zákon vzácných jevů
Binomické rozdělení	$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$	Rozložení délky zobáku v populaci kosa
Chí-kvadrát rozdělení	$N(\mu, \sigma^2)$	Pravděpodobnost, že udělám zkoušku
Alternativní rozdělení	$\sum_{i=1}^k Z_i^2 \sim \chi_k^2$	Počet infikovaných jedinců v n -členné skupině

Jakými parametry je definováno normální rozdělení?

Jakým(i) parametry jsou definována další spojitá rozdělení: t, F a χ^2 ?

Někdy je důležité poznat, že data mají binomické rozdělení, protože pro taková data máme v nabídce binomický test o pravděpodobnosti (např. $H_0: p = 0.25$). Uveďte příklad takových dat, jakou veličinu/ proměnnou měříme? A co je potom ta testovaná pravděpodobnost p ?

Poissonovo rozdělení používáme k popisu pravděpodobnosti vzácných jevů. Přibližte, o co jde.

Jaký je rozdíl mezi parametrem a odhadem parametru?

Co vyjadřuje střední hodnota náhodné veličiny? Jak mohu zjistit její hodnotu?

Co vyjadřuje rozptyl (variabilita) náhodné veličiny? Jak mohu zjistit jeho hodnotu?

K čemu se vztahuje pojem kovariance, kdy nás zajímá její hodnota (odhad)?

Co vyjadřuje, popisuje korelace?

Znát vzorečky pro výpočet aritmetického průměru a rozptylu.

Co se stane s náhodnou veličinou během normování? Tedy proč to děláme a které důležité charakteristiky se změň?

Když odhaduji hodnotu skutečného parametru, jaké různé charakteristiky mohu předložit k posouzení kvality (spolehlivosti) odhadu?

Jaký je rozdíl mezi směrodatnou odchylkou a střední chybou odhadu? Jak se značí? Vzorečky pro výpočet.

Co říká bodový odhad parametru a co intervalový odhad parametru? Co leží uvnitř intervalu?

Které charakteristiky náhodného výběru ovlivňují šířku intervalu spolehlivosti (konfidenčního intervalu)?

Umíte přečíst zápis pro výpočet konfidenčního intervalu? Která čísla potřebuji pro výpočet?

$$\left\{ \bar{X} - t_{(n-1)}(1 - \alpha/2) \cdot \frac{S}{\sqrt{n}} ; \bar{X} + t_{(n-1)}(1 - \alpha/2) \cdot \frac{S}{\sqrt{n}} \right\}$$

Když neznám skutečnou populační střední hodnotu a chci ji popsat pomocí výběrového průměru, použiju k tomu

- bodový odhad parametru
- intervalový odhad parametru
- test hypotézy o střední hodnotě.

V každé možnosti se skrývá nějaká informace, ale která z nich řekne o poloze populační střední hodnoty nejlíp?

Formulace nulové a alternativní hypotézy. Rozumíte všem formulacím z domácích úkolů? Pro jaké hodnoty je hypotéza formulována, pro odhady nebo pro skutečné parametry? Vidíte souvislost mezi matematickou formulací a biologickou interpretací, tedy mezi hypotetickou hodnotou testovaného parametru a vlastností, která z této hodnoty vyplývá (např. závislost X nezávislost proměnných, srovnatelnost X odlišnost průměrů, srovnatelnost X odlišnost pravděpodobností).

Podle čeho se rozhodujeme o zamítnutí či nezamítnutí nulové hypotézy?

Jakou metodou budete vyhodnocovat data

- o počtu lidí s černýma, hnědýma, šedýma a modrýma očima
- o výšce lipnice na loukách s různým substrátem

- c) o výšce lipnice na gradientu vlhkosti
- d) o frekvenci barevných variet papoušků v populaci
- e) o hmotnosti netopýrů druhu XY, když známe typickou průměrnou hmotnost tohoto druhu?

Kde se berou stupně volnosti u t -rozdělení?

Co je to párový test? Vymyslete situaci, kdy byste použili párový test.

Co je to pravděpodobnostní diagram (quantile-quantile plot)? K čemu ho používáme (a jak funguje)?

Yatesova oprava na spojitost: kdy se používá, ke kterému testu patří, co oprava způsobí?

Při složitějších testech se často využívá aproximace normálním rozdělením. Potom testovou statistiku porovnávám s normovaným normálním rozdělením. Tušíte, co znamená „aproximace“, případně co nás opravňuje aproximaci použít?

Když studujeme (analyzujeme) dva a více výběrů najednou, často kontrolujeme předpoklad o shodnosti rozptylů (variancí). Které testy k tomu využíváme? Odhady čeho potřebujeme k testu dvou výběrů?

Co testuje t -test a jaký je rozdíl mezi párovým a dvouvýběrovým t -testem? Jaké předpoklady musí být splněny, abychom mohli použít t -test?

Neparametrický Mannův–Whitneyův test: co testuje, jak zní nulová hypotéza?

Co jsou to shody v pořadí (tied values)?

Vlastními slovy popište nezávislost náhodných veličin.

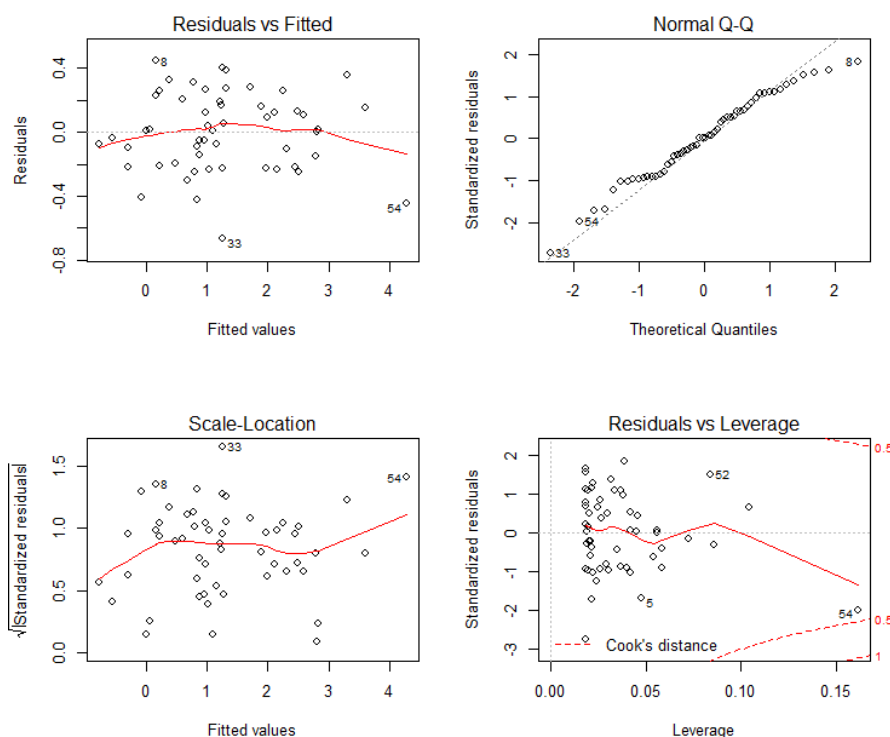
Korelační koeficient může mít hodnotu jen v nějakém intervalu. V jakém a tato čísla interpretujeme?

Jaká je nulová hypotéza Pearsonova testu?

Transformace: co to je, kdy to používáme?

Regresní analýza:

Zapište teoretickou rovnici jednoduchého lineárního modelu a vysvětlete. Jaké jsou předpoklady?



Parametry b0 a b1: co znamenají, jak je interpretujeme? Jak je testujeme?

Co to je reziduum?

Test významnosti celého modelu: který výsledek to je? Co to znamená, když je výsledek neprůkazný? A co znamená, když je průkazný?

Koeficient determinace: které číslo ze summary(model) to je a co znamená?

Kontingenční tabulky:

Jaký typ dat analyzujeme touto metodou?

Jaké hypotézy /otázky dokážeme testovat?

Jaká je pravděpodobnost, že student s hnědými vlasy bude mít modré oči? (data_vlasy.RData)
Demonstrujte pojmy četnost a relativní četnost a odhad pravděpodobnosti.

Jak se jmenuje rozdělení a test, který používáme při analýze kontingenčních tabulek? Rozumíte vzorečku $\sum_{j=1}^k \frac{(n_j - o_j)^2}{o_j}$?

Co říká míra těsnosti vazby v kontingenční tabulce?

Kdy můžeme použít Fisherův test?

Jak analyzujeme párová nominální data?

Analýza rozptylu:

Jak zní nulová hypotéza?

Jaké jsou předpoklady pro parametrickou analýzu?

Co testuje Tukeyho test?

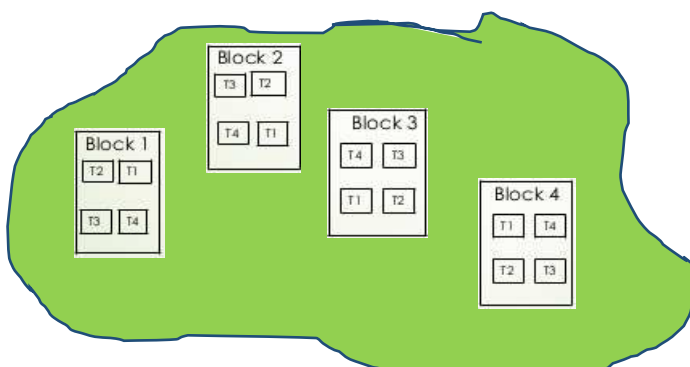
Jak se jmenuje neparametrický test pro porovnání více průměrů?

Design experimentů:

CO je to pevný a náhodný efekt?

Jak funguje latinský čtverec?

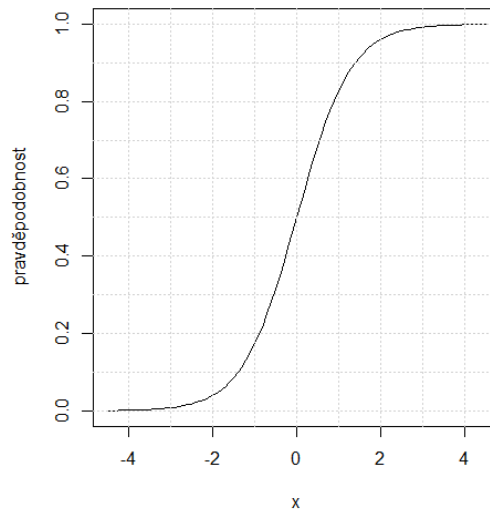
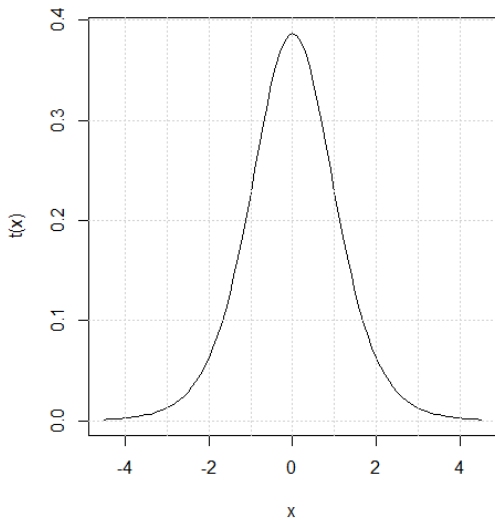
Co lze říci k tomuto nákresu?



==== Otázky z extraligy ☺ =====

Co to je rozdělení pravděpodobnosti? Jaký je rozdíl mezi diskretním a spojitým rozdělením?

Co to je distribuční funkce náhodné veličiny? A jak s ní souvisí hustotní funkce?



Co říká centrální limitní věta, k čemu je dobrá?

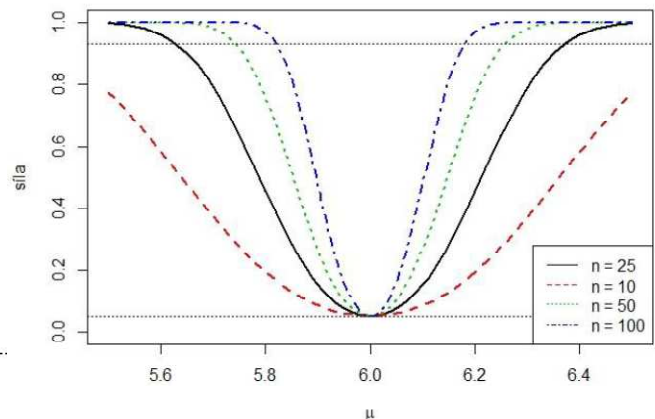
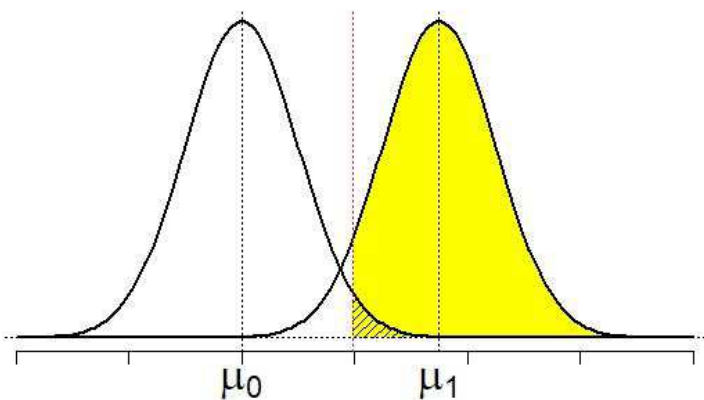
V t -testu hypotézy $H_0: \mu_X = 0$ ($df = 10$) jsem dostala testovou statistiku $T = 2.03$ a p -hodnotu 0.07 . Jenže hodnota odpovídá kvantilu $t_{10}(1 - 0.035)$. Je v tom rozpor?

Když $X_i \sim N(\mu, \sigma^2)$, potom výběrový průměr \bar{X} má rozdělení $N(\mu, \frac{\sigma^2}{n})$. Zkuste vysvětlit, proč to tak je. Této vlastnosti využíváme při testování hypotézy o skutečné hodnotě populačního průměru. Víte, jak tento test vypadá (jak je konstruován – myšlenka)?

Vysvětlete myšlenku Wilcoxonova neparametrického párového testu.

F -test používáme k porovnání dvou odhadů variancí. Kdy je ideální hodnota (odpovídající nulové hypotéze) jednička a kdy malé číslo blízko nuly?

Zkuste vysvětlit, co to je síla testu.



Vysvětlete metodu nejmenších čtverců.

Zapište rozklad variability v regresním modelu na sumy čtverců. Popište slovy.

Test podmodelu (regrese nebo ANOVA): jak zní nulová hypotéza? A jak ji testujeme?

Jaká je myšlenka analýzy rozptylu (ANOVA)?