

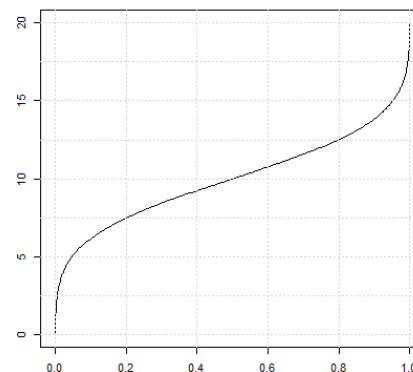
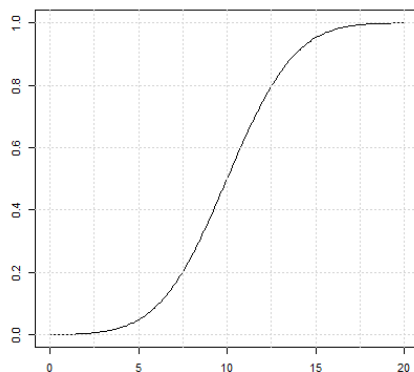
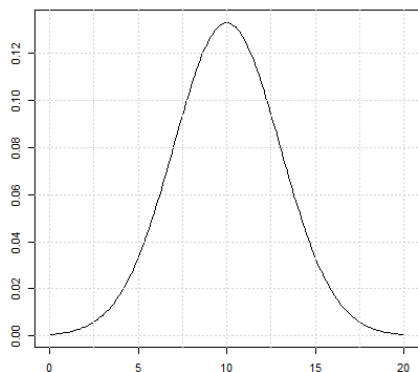
Rozlišujeme data kvalitativní a kvantitativní. Zkuste popsat nějaké rozdíly mezi těmito typy dat: Jak vypadá základní popis takových dat a jaké (typické) testovací metody pro ně používáme?

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

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?

Jakou délku má křídlo dospělého kosa? Měřím 10 náhodně chyčených jedinců. Na tomto příkladu zkuste vysvětlit pojem náhodná veličina.

V teorii testování hypotéz se setkáváme s touto trojicí funkcí: kvantilová funkce, hustota a distribuční funkce. Přiřaďte názvy ke grafům a vyberte správné popisky:  $F(x)$ ,  $f(x)$ , a jednu z dvojice  $F^{-1}(p)$  a  $\Phi^{-1}(p)$ .



Řekněme, že funkce popisují pravděpodobnosti hmotnosti vrbců. Vyznačte na všech grafech pravděpodobnost, že náhodně vybraný jedinec bude lehčí než 7.5 g.

Který z prvků zápisu se jmenuje kvantil:  $P(X \leq x_{0,95}) = 0,95$ ? Umíte tento zápis správně přečíst?

Co to je kvantil, kde se s ním setkáváme.

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

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

Jaký je rozdíl mezi parametrem a odhadem parametru?

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?

Co říká bodový odhad parametru a co intervalový odhad parametru?

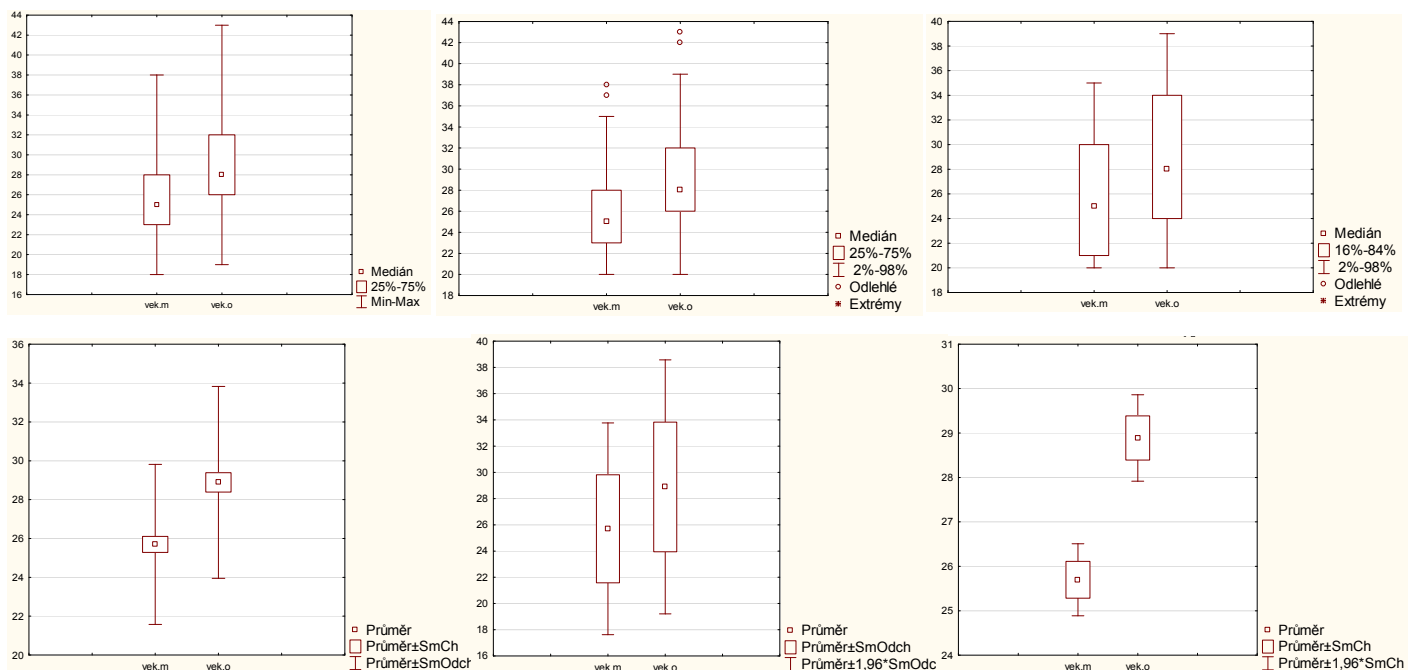
Které charakteristiky náhodného výběru ovlivňují šířku intervalu spolehlivosti (konfidenčního intervalu)?

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 nejmíc?

Ve kterém z grafů najdete informaci o: 50 % „středních“ hodnot, 68 % „středních“ hodnot, intervalu pokrývajícím zhruba 96 % pozorování datového souboru, celém rozsahu hodnot v datovém souboru, intervalu spolehlivosti pro parametr  $\mu$ , přesnosti odhadu střední hodnoty  $\mu$ ?



Na obrázku je rozhodovací schéma při testování nějakého „statistického“ tvrzení. Kde je znázorněno špatné rozhodnutí? Kterou pravděpodobnost volím na začátku testu? Na co má tato pravděpodobnost vliv (ve kterém vzorečku/výpočtu se objeví)? Zkuste říct, co to je síla testu.

NAŠE ROZHODNUTÍ	SKUTEČNOST	
	$H_0$ platí	$H_0$ neplatí (platí $H_1$ )
$H_0$ zamítáme	<b>Chyba 1. druhu: <math>\alpha</math></b> Prst. chyby $\leq \alpha$	<b>SPRÁVNÉ ROZHODNUTÍ</b> $P = 1 - \beta$ <b>síla testu</b>
$H_0$ nezamítáme	<b>SPRÁVNÉ ROZHODNUTÍ</b> $(P \geq 1 - \alpha)$	<b>Chyba 2. druhu: <math>\beta</math></b> $\beta$ většinou neznáme

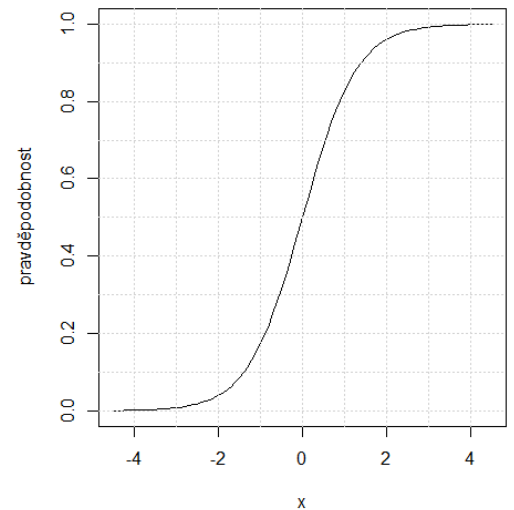
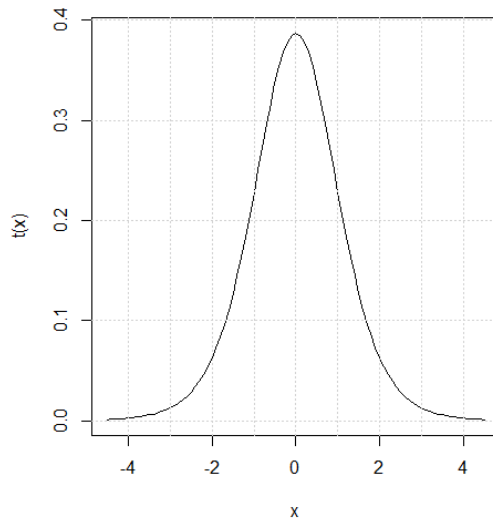
Podle čeho se rozhodujeme o zamítnutí či nezamítnutí nulové hypotézy?

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  $2.03$  odpovídá kvantilu  $t_{10}(1 - 0.035)$ . Je v tom rozpor?

Na obou grafech vyznačte zjištěnou  $p$ -hodnotu  $0.07$ . Jak tuto pravděpodobnost interpretujeme? (Uvědomte si, že jde o oboustrannou alternativu.)

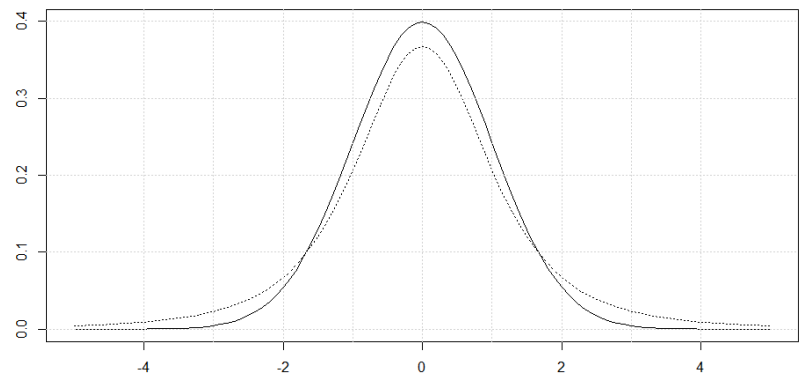
Co říká číslo  $t_{n-1}(1 - \alpha/2)$ ?

Jaké nejčastější hodnoty pravděpodobnosti prvního druhu používáme?



V grafu jsou vykresleny hustoty normálního a  $t$ -rozdělení. Která hustota je která a proč?

Testovou statistiku  $t$ -testu při oboustranné alternativě dáváme do absolutní hodnoty. Proč?  
kritérium:  $|T| \geq t_{n-1}(1 - \alpha/2)$



Jakou metodou budete vyhodnocovat data

- a) o počtu lidí s černýma, hnědýma, šedýma a modrýma očima
- b) 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?

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)?

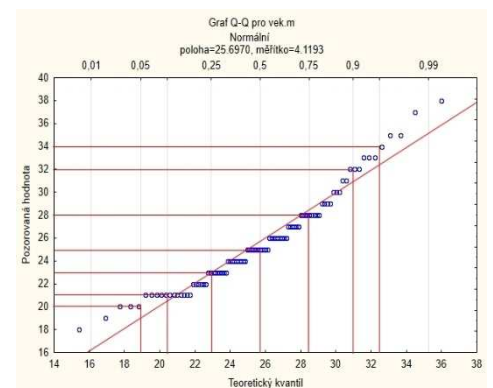
Kde se berou stupně volnosti u  $t$ -rozdělení?

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

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

Wilcoxonův test je neparametrický párový test. Znáte jeho myšlenku?

V čem spočívá Yatesova korekce na spojitost? Kdy se používá? (vizte 5. přednáška, Test o binom. pravděpodobnosti.)



Při složitějších testech se často využívá aproximace normálním rozdělením. Potom testovou statistiku porovnáváme 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). Jak se jmenuje test pro dva výběry? Odhady čeho k testu potřebujeme?

Ve výpisu výsledků F testu píše STATISTICA: „F-poměr rozptyly“ a Rko: „alternative hypothesis: true ratio of variances is not equal to 1“. Víte proč, k čemu tato formulace odkazuje?

K testu shody variancí dvou výběrů dostanu jediný výsledek – konfidenční interval. Podle čeho poznám, zda jsme hypotézu o shodě zamítnuli nebo nezamítnuli? Ukažte na příkladu z přednášky: „95 percent confidence interval (1.054853 , 3.295444). Víte, proč je tento konfidenční interval asymetrický?

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

Na grafu jsou histogramy pro délku chlapců a dívek ve 24. týdnu života. Zároveň jsou proloženy hustoty normálního rozdělení. Co můžete říct o odhadech průměrů a variability obou výběrů na základě tohoto grafu?

Dokážete okomentovat všechny výsledky, které k *t*-testu vytiskne software (STAT nebo Rko)? Mám na mysli, co které číslo znamená.

Neparametrický Mannův-Whitneyův (Wilcoxonův) test: co přesně testuje? A když ho použijeme na porovnání charakteristik polohy, co vlastně navíc předpokládáme?

Co jsou to shody v pořadí (tied values)?

Analýzu rozptylu, regresní analýzu i kontingenční tabulky procvičíme na konkrétních příkladech. Zaměřte se na pořadí kroků: formulovat problém, ověřit předpoklady, aplikovat test, výsledky zformulovat do věty, přidat ilustrační grafy.

