

ÚVOD DO TESTOVÁNÍ HYPOTÉZ

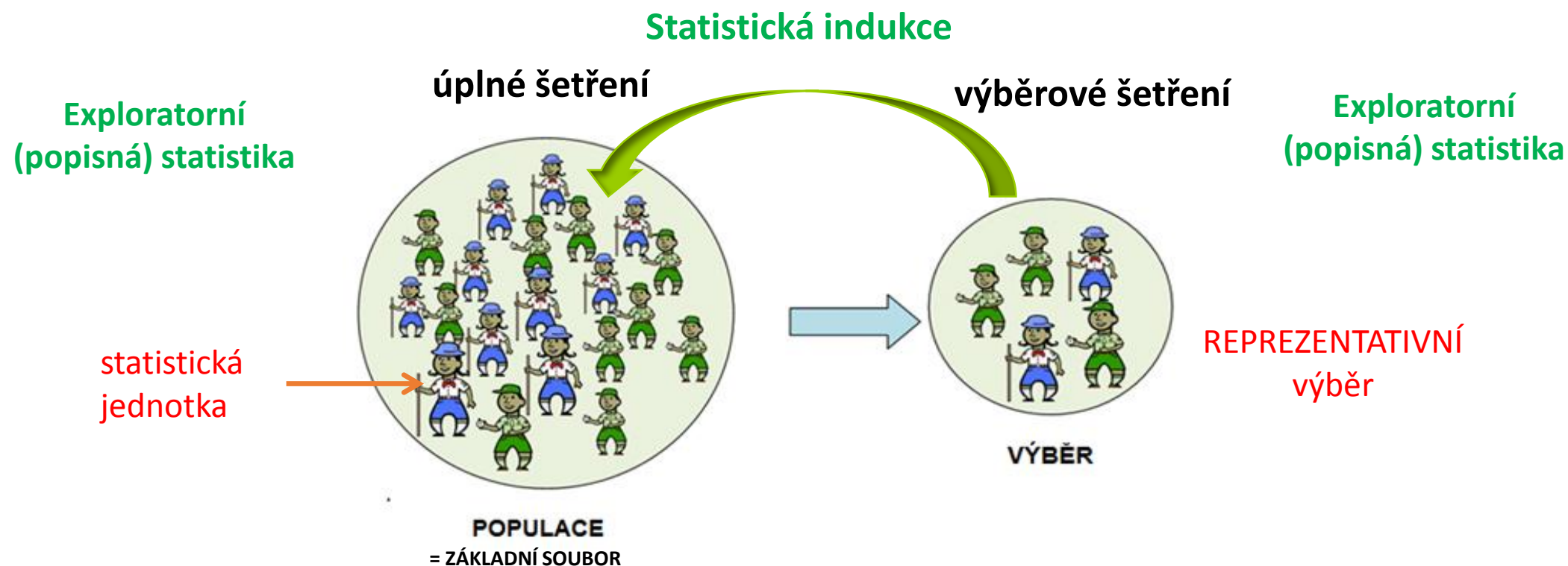
Martina Litschmannová, Adéla Vrtková



Obsah

- **Opakování**
 - Základní pojmy spojené se statistickou indukcí
 - Výběrové charakteristiky a populační parametry
 - Intervalové odhady
- **Úvod do testování hypotéz**
 - Co je to statistická hypotéza
 - Základní typy statistických hypotéz
 - Co je to testování hypotéz? (zavedení pojmu nulová a alternativní hypotéza pro parametrické hypotézy, definování pojmu efekt, chyby při testování hypotéz)
 - Jak rozhodnout o pravdivosti statistické hypotézy?

Jaké pojmy je nutno znát?



statistické znaky – údaje, které u statistických znaků sledujeme (např. váha, výška, IQ, ...)

Výběrové charakteristiky vs. parametry populace

- **Parametry populace** (obvykle pro jejich značení používáme symboly řecké abecedy) jsou **konstanty**.
- **Charakteristiky výběru** (obvykle značíme latinkou) jsou obvykle různé – v závislosti na pořazeném výběru. Jsou to **náhodné veličiny**.

Základní soubor (populace)	střední hodnota μ	medián $x_{0,5}$	rozptyl σ^2	směr. odchylka σ	pravděpodobnost π
Výběrový soubor (výběr)	průměr $\bar{X}$	výběrový medián $\tilde{X}_{0,5}$	výběrový rozptyl S^2	výběrová směr. odchylka S	rel. četnost $\hat{\pi}$

Poznámka: $\bar{X}, \hat{\mu}$... průměr (obecně, tj. průměr nějakého výběru, náhodná veličina),
 $\bar{x}$... průměr (konkrétní realizace, tj. průměr daného výběru, jedno číslo),
 p ... relativní četnost (konkrétní realizace zjištěna z daného výběru)

Proč používáme metody statistické indukce?



Lze určit střední hodnotu životnosti el. součástek?



Lze určit účinnost léku?



Lze určit, který výrobce vyrábí kvalitněji?

Neznáme-li rozdělení náhodné veličiny X , pak
parametry náhodné veličiny X nelze většinou přesně určit, lze je jen odhadnout.

Proč používáme metody statistické indukce?



Lze určit střední hodnotu životnosti el. součástek?



Lze určit účinnost léku?



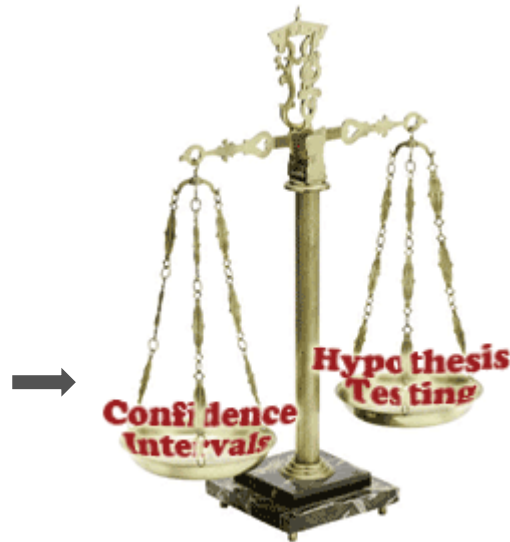
Lze určit, který výrobce vyrábí kvalitněji?

Neznáme-li rozdělení náhodné veličiny X , pak
**parametry náhodné veličiny X nelze většinou přesně určit, lze je jen odhadnout
nebo lze ověřovat tvrzení, která se týkají jejich rozdělení.**

Základní metody statistické indukce

- **Intervalové odhady** (angl. confidence intervals) – umožňují odhadnout nejistotu v odhadu parametru náhodné veličiny
- **Testování hypotéz** (angl. hypothesis testing) - umožňuje posoudit, zda experimentálně získaná data nepopírají předpoklad, který jsme **před** provedením testování učinili.

Používáme, chceme-li určit velikost parametru NV, resp. velikost efektu (rozdílu, resp. poměru parametrů dvou NV).



Používáme, chceme-li ověřit platnost předem definované hypotézy (s předem danou hladinou významnosti).

Co je to statistická hypotéza?

Statistická hypotéza – předpoklad (tvrzení) o rozdělení náhodné veličiny

Co je zdrojem statistických hypotéz?

- předchozí zkušenosti,
- teorie, kterou je třeba doložit,
- požadavky na kvalitu produktu,
- dohady založené na náhodném pozorování...

Co je to statistická hypotéza?

Statistická hypotéza – předpoklad (tvrzení) o rozdělení náhodné veličiny

Příklady statistických hypotéz:

- Střední životnost žárovek Ed je nižší než výrobcem udávaných 5 let.
- Mortalita je u laparoskopických operací nižší než u operací konvenčních.
- Průměrné výsledky srovnávacích testů závisí na typu absolvované střední školy.
- Pořízený datový soubor je výběrem z populace mající normální rozdělení.

Poznámka: Rozdíl (resp. poměr) parametru náhodné veličiny a jeho očekávané hodnoty, popřípadě rozdíl (resp. poměr) parametrů náhodných veličin nazýváme **efekt**.

Jaké typy statistických hypotéz rozlišujeme?

- **Parametrická statistická hypotéza** – tvrzení ohledně efektu
 - Hypotézy o parametru jedné populace (o střední hodnotě, rozptylu, mediánu, parametru binomického rozdělení, ...)
 - Hypotézy o parametrech dvou populací (srovnávací testy)
 - Hypotézy o parametrech více než dvou populací (ANOVA, Kruskalův-Wallisův test, ...)
- **Neparametrická statistická hypotéza** – tvrzení o jiné vlastnosti rozdělení náhodné veličiny než o jejím parametru (např. hypotézy o typu rozdělení NV, hypotézy o závislosti NV, ...)

Jak ověřit, zda je statistická hypotéza pravdivá?

Příklad: Domníváme se, že střední hodnota obsahu cholesterolu v krvi je u české populace 4,7 mmol/l.

$$H_0: \mu = 4,7$$

$$H_A: \mu \neq 4,7$$

Jak tento předpoklad ověřit?

- Zjistíme údaje o obsahu cholesterolu v krvi u 100 náhodně vybraných Čechů.
- Průměrný obsah cholesterolu v krvi probandů (tj. jedinců, kteří jsou předmětem zkoumání) byl 5,4 mmol/l.

Jsou tyto výsledky v souladu s naší hypotézou?

- I kdyby byla testovaná hypotéza pravdivá, nelze očekávat, že průměrná hodnota pozorovaná ve výběru bude přesně 4,7 mmol/l.
- **Nulovou hypotézu zamítneme, pokud získané uspořádání výběru bude za předpokladu platnosti nulové hypotézy velmi nepravděpodobné.**

Jak ověřit, zda je statistická hypotéza pravdivá?

Pravdivost nulové hypotézy nelze na základě dat dokázat!!!

Pravdivost nulové hypotézy lze na základě dat pouze vyvrátit.

Nulová hypotéza
(obžalovaný je nevinen)

Alternativní hypotéza
(obžalovaný je vinen)

Data (výběrový soubor)
(svědci)



Testové kritérium
(soudce)

Princip presumpce nevinny

Neodsoudí-li soudce obžalovaného, nemusí to znamenat, že je obžalovaný nevinný. Může to znamenat, že neexistuje dostatek důkazů pro jeho odsouzení!

Terminologie v praxi (I)

Zadání problému: Ověřte, zda použití bezpečnostních pásů ovlivňuje úmrtnost při dopravních nehodách.

Populace 1 (základní soubor 1): účastníci dopravních nehod, kteří seděli na místech, na nichž je možno používat bezpečnostní pásy a byli připoutáni.

Populace 2 (základní soubor 2): účastníci dopravních nehod, kteří seděli na místech, na nichž je možno používat bezpečnostní pásy a nebyli připoutáni.

Sledovaný statistický znak (náhodná veličina): úmrtnost (relativní četnost zemřelých)

Nulová hypotéza H_0 : $\pi_A = \pi_N$, kde π_A , resp. π_N označuje úmrtnost účastníků dopravních nehod, kteří byli, resp. nebyli, připoutáni

Alternativní hypotéza H_A : $\pi_A \neq \pi_N$ (zadání problému neobsahuje jednostrannou nerovnost).

Terminologie v praxi (II)

Zadání problému: Ověřte, zda průměrný plat v ČR je větší než 24 000,- Kč.

Populace (základní soubor): všichni občané ČR pobírající mzdu

Sledovaný statistický znak (náhodná veličina): mzda

Nulová hypotéza $H_0: \mu = 24\,000$

Alternativní hypotéza $H_A: \mu > 24\,000$

(zadání obsahuje nerovnost v tomto tvaru)

Poznámka: *Průměrný plat zjištěný z výběrového souboru by měl být větší než 24 000,- Kč. Pokud by tomu tak nebylo, měli bychom použít oboustrannou alternativní hypotézu.*

Možné chyby při testování hypotéz

Při testování hypotéz mohou nastat čtyři situace:

		Rozhodnutí	
		Nezamítáme H_0	Zamítáme H_0
Skutečnost	Platí H_0	Správné rozhodnutí Pravděpodobnost: $1 - \alpha$	Chyba I. druhu Pravděpodobnost: α
	Platí H_A	Chyba II. druhu Pravděpodobnost: β	Správné rozhodnutí Pravděpodobnost: $1 - \beta$

(zvolená) hladina významnosti

síla testu

- Jelikož výběr, na jehož základě rozhodujeme, je náhodný, chybám I. a II. druhu se nelze vyhnout.
- Chtěli bychom mít k dispozici testy s nízkou hladinou významnosti a vysokou silou testu.

Možné chyby při testování hypotéz

Závěry:

- Chtěli bychom mít k dispozici testy s nízkou hladinou významnosti a vysokou silou testu - to jsou, bohužel, dva protichůdné požadavky.
- S klesající hladinou významnosti roste pravděpodobnost chyby II. druhu!
- Existuje jediný způsob jak snížit α i β – zvýšení rozsahu výběru.
- Hladinu významnosti α volíme obvykle 0,05 (resp. 0,01).
- Sílu testu lze poté ovlivnit volbou testové statistiky (pro ověření určité hypotézy lze často použít několik různých testových statistik) a dostatečného počtu pozorování.

Jak postupovat při testování parametrických hypotéz? (čistý test významnosti)

1. Formulujeme **nulovou a alternativní hypotézu**.
2. Zvolíme tzv. **testovou statistiku**, tj. výběr. charakteristiku, jejíž rozdělení závisí na test. parametru θ .
(Rozdělení test. statistiky za předpokladu platnosti H_0 nazýváme **nulové rozdělení**.)
3. Ověříme předpoklady testu!
4. Výpočet pozorované hodnoty x_{OBS} testové statistiky $T(\mathbf{X})$.
5. Výpočet p – hodnoty (angl. „ p – value“).

Tvar alternativní hypotézy H_A	p -hodnota
$\theta < \theta_0$	$p\text{-hodnota} = F_0(x_{OBS})$
$\theta > \theta_0$	$p\text{-hodnota} = 1 - F_0(x_{OBS})$
$\theta \neq \theta_0$	$p\text{-hodnota} = 2\min\{F_0(x_{OBS}); 1 - F_0(x_{OBS})\}$

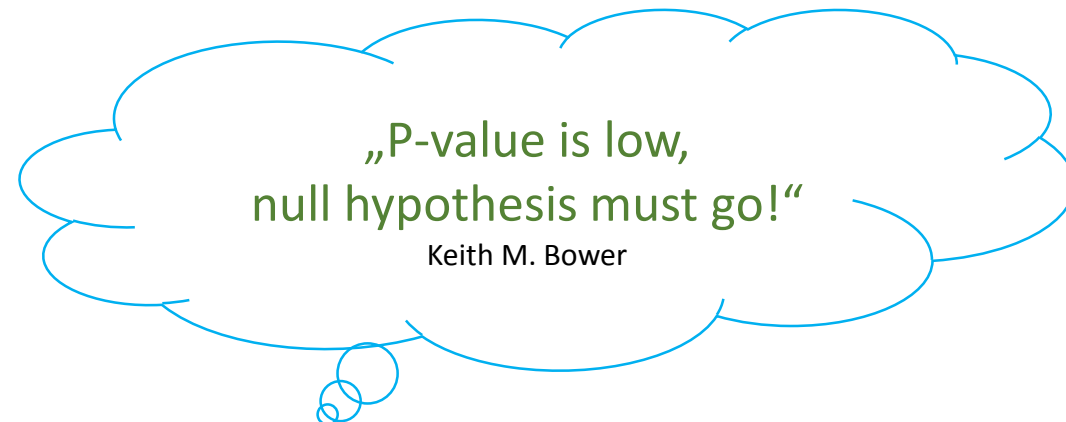
6. Rozhodnutí o výsledku testu:

p -hodnota	Rozhodnutí
$p\text{-hodnota} < \alpha$	Na hladině významnosti α zamítáme H_0 ve prospěch H_A .
$p\text{-hodnota} \geq \alpha$	Na hladině významnosti α nelze H_0 zamítnout.

Co je to p-hodnota?

- Zjednodušeně:

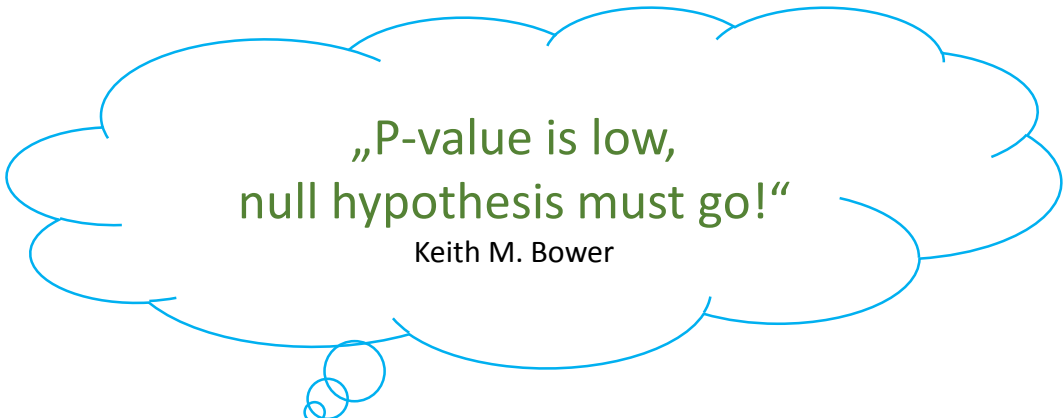
Je to pravděpodobnost,
že náhodným výběrem získáme data,
která svědčí v neprospěch nulové hypotézy
silněji (extrémněji) než náš výběrový soubor.



Jak postupovat při testování parametrických hypotéz? (s využitím statistického software)

1. Formulujeme **nulovou a alternativní hypotézu**.
2. Zvolíme vhodný test.
3. Ověříme předpoklady testu!
4. Určíme p – *hodnotu* (angl. „ p – *value*“).
5. Rozhodnutí o výsledku testu:

p -hodnota	Rozhodnutí
$p\text{-hodnota} < \alpha$	Na hladině významnosti α zamítáme H_0 ve prospěch H_A .
$p\text{-hodnota} \geq \alpha$	Na hladině významnosti α nelze H_0 zamítnout.



„P-value is low,
null hypothesis must go!“

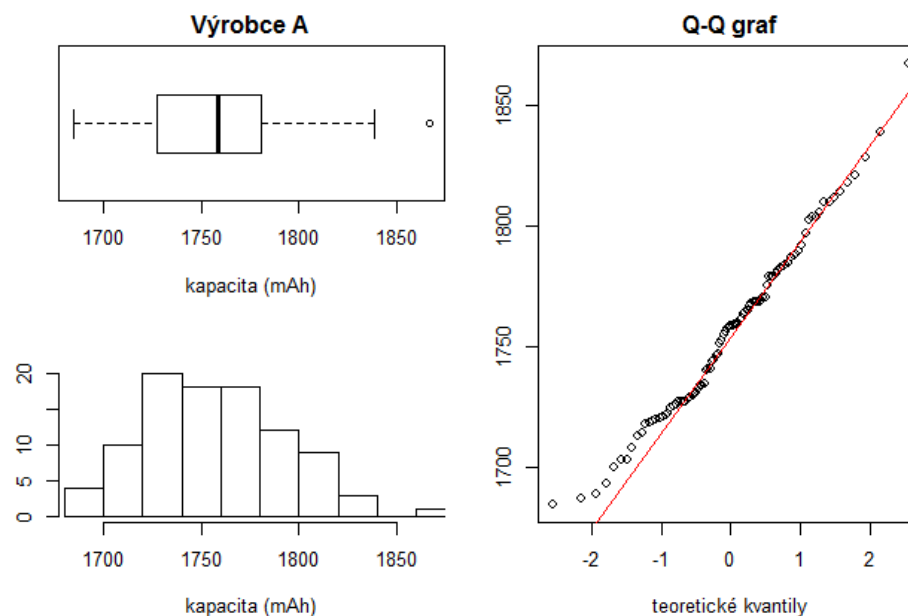
Keith M. Bower

Trocha praxe neuškodí 😊

Výrobce A tvrdí, že jeho akumulátory vykazují vyšší průměrnou kapacitu než 1 750 mAh (což je údaj udávaný konkurenčními výrobci). Pro ověření tohoto tvrzení byly zjištěny kapacity 95 akumulátorů tohoto výrobce – viz `aku_new.csv`. Ověřte na 5% hladině významnosti, zda lze s tvrzením výrobce A souhlasit.

Řešení:

1. $H_0: \mu = 1750, H_A: \mu > 1750$
2. **t- test** – jsou-li data výběrem z normálního rozdělení × **Wilcoxonův test** (o mediánu) – nejsou-li data výběrem z normálního rozdělení, avšak jsou výběrem z nějakého symetrického rozdělení
3. Ověření předpokladů testu:
 - Vypořádání se s odlehlými pozorováními – odstraněno
– z technologického hlediska je kapacita větší než 1 850 mAh nepravděpodobná.



Výrobce A tvrdí, že jeho akumulátory vykazují vyšší průměrnou kapacitu než 1 750 mAh (což je údaj udávaný konkurenčními výrobci). Pro ověření tohoto tvrzení byly zjištěny kapacity 95 akumulátorů tohoto výrobce – viz `aku_new.csv`. Ověřte na 5% hladině významnosti, zda lze s tvrzením výrobce A souhlasit.

Řešení:

1. $H_0: \mu = 1750, H_A: \mu > 1750$
2. **t- test** – jsou-li data výběrem z normálního rozdělení × **Wilcoxonův test** (o mediánu) – nejsou-li data výběrem z normálního rozdělení, avšak jsou výběrem z nějakého symetrického rozdělení
3. Ověření předpokladů testu:

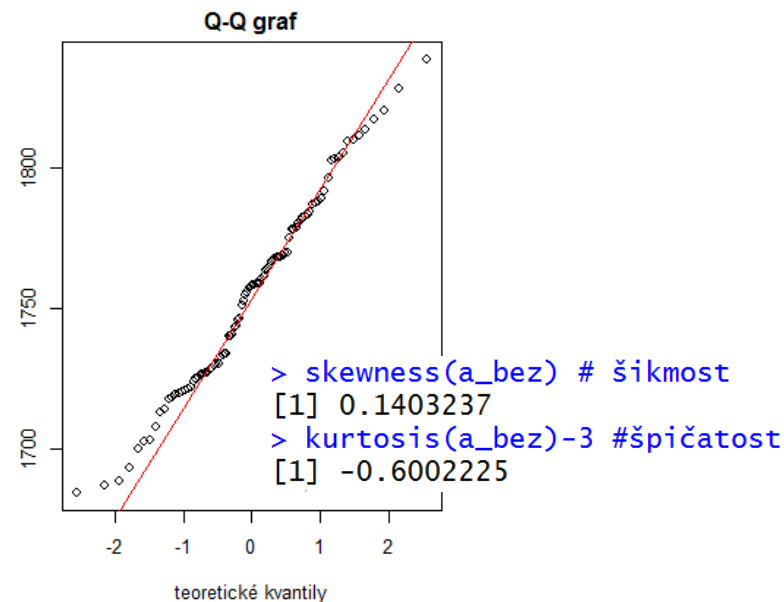
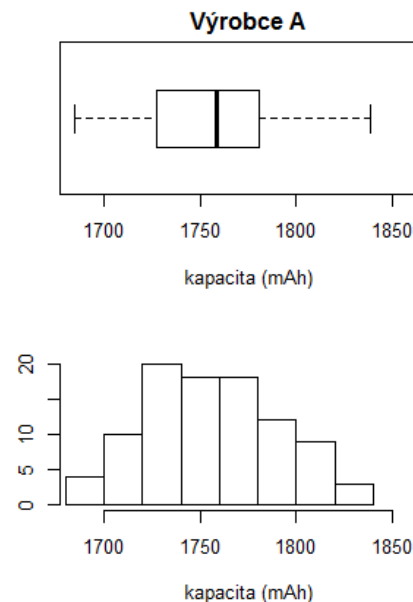
Na základě explorační analýzy považujeme data za výběr z normálního rozdělení. $\Rightarrow$ **t-test**

4. Určení p-hodnoty:

```
> t.test(a_bez,  
+       mu=1750,          # definování nulové hypotézy  
+       alternative="greater") # upřesnění alternativy
```

One sample t-test

```
data: a_bez  
t = 1.4699, df = 93, p-value = 0.07248  
alternative hypothesis: true mean is greater than 1750  
95 percent confidence interval:  
 1749.313      Inf  
sample estimates:  
mean of x  
1755.271
```



Výrobce A tvrdí, že jeho akumulátory vykazují vyšší průměrnou kapacitu než 1 750 mAh (což je údaj udávaný konkurenčními výrobci). Pro ověření tohoto tvrzení byly zjištěny kapacity 95 akumulátorů tohoto výrobce – viz `aku_new.csv`. Ověřte na 5% hladině významnosti, zda lze s tvrzením výrobce A souhlasit.

Řešení:

1. $H_0: \mu = 1750, H_A: \mu > 1750$
2. **t- test** – jsou-li data výběrem z normálního rozdělení × **Wilcoxonův test** (o mediánu) – nejsou-li data výběrem z normálního rozdělení, avšak jsou výběrem z nějakého symetrického rozdělení

3. Ověření předpokladů testu:

Na základě explorační analýzy považujeme data za výběr z normálního rozdělení. $\Rightarrow$ **t-test**

4. Určení p-hodnoty:

$p - \text{hodnota} = 0,072$

5. Rozhodnutí:

Na hladině významnosti 5 % nelze nulovou hypotézu ve prospěch dané alternativy zamítnout (t – test, $p - \text{hodnota} = 0,072$).

Výrobce A tvrdí, že jeho akumulátory vykazují vyšší průměrnou kapacitu než 1 750 mAh (což je údaj udávaný konkurenčními výrobci). Pro ověření tohoto tvrzení byly zjištěny kapacity 95 akumulátorů tohoto výrobce – viz `aku_new.csv`. Ověřte na 5% hladině významnosti, zda lze s tvrzením výrobce A souhlasit.

Řešení:

1. $H_0: \mu = 1750, H_A: \mu > 1750$
2. **t- test** – jsou-li data výběrem z normálního rozdělení × **Wilcoxonův test** (o mediánu) – nejsou-li data výběrem z normálního rozdělení, avšak jsou výběrem z nějakého symetrického rozdělení

3. Ověření předpokladů testu:

Na základě explorační analýzy považujeme data za výběr z normálního rozdělení. $\Rightarrow$ **t-test**

4. Určení p-hodnoty:

$p - \text{hodnota} = 0,072$

5. Rozhodnutí:

Na hladině významnosti 5 % nelze nulovou hypotézu ve prospěch dané alternativy zamítnout (t – test, $p - \text{hodnota} = 0,072$). Tj. odchylka zjištěné průměrné kapacity (1 755 mAh) od testované hodnoty není statisticky významná.

Jaké základní testy parametrických hypotéz používáme?

viz http://am-nas.vsb.cz/lit40/PRASTA/Statisticka_indukce.pdf

Výrobce A tvrdí, že jeho akumulátory vykazují vyšší průměrnou kapacitu akumulátory výrobce B. Pro ověření tohoto tvrzení byly zjištěny kapacity 95 akumulátorů výrobce A a 89 akumulátorů výrobce B – viz `aku_new.csv`. Ověřte na 5% hladině významnosti, zda lze s tvrzením výrobce A souhlasit.

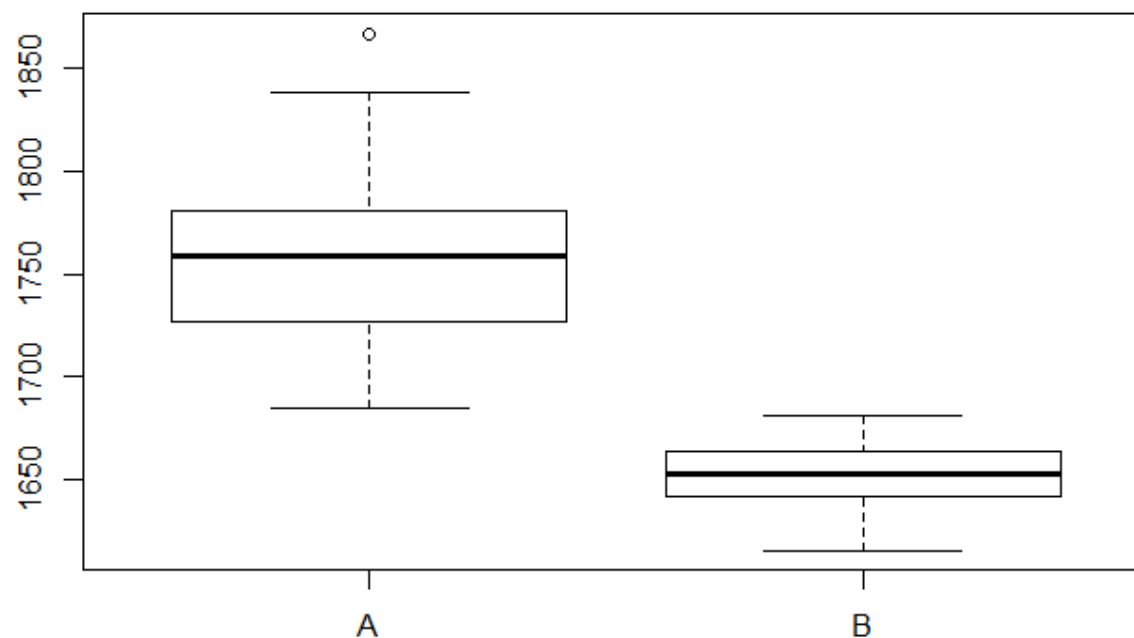
Řešení:

- **Pre-procesing dat:** vytvoření souboru, v němž jsou údaje pouze o výrobcích A a B + identifikace odlehlých pozorování a jejich zpracování
- **Explorační analýza:** vizualizace dat (vícenásobný boxplot, histogramy), výpočet výběrových charakteristik, posouzení normality (QQ grafy), posouzení shody rozptylů
- **Statistická indukce** (nezapomínat na volbu hladiny významnosti a ověření předpokladů):
 - Intervalový odhad rozdílu středních kapacit baterií výrobců A a B.
 - Ověření hypotézy, zda je průměrná kapacita akumulátorů výrobce A statisticky významně větší než průměrná kapacita akumulátorů výrobce B.

Výrobce A tvrdí, že jeho akumulátory vykazují vyšší průměrnou kapacitu akumulátory výrobce B. Pro ověření tohoto tvrzení byly zjištěny kapacity 95 akumulátorů výrobce A a 89 akumulátorů výrobce B – viz `aku_new.csv`. Ověřte na 5% hladině významnosti, zda lze s tvrzením výrobce A souhlasit.

Řešení:

Pre-processing dat:

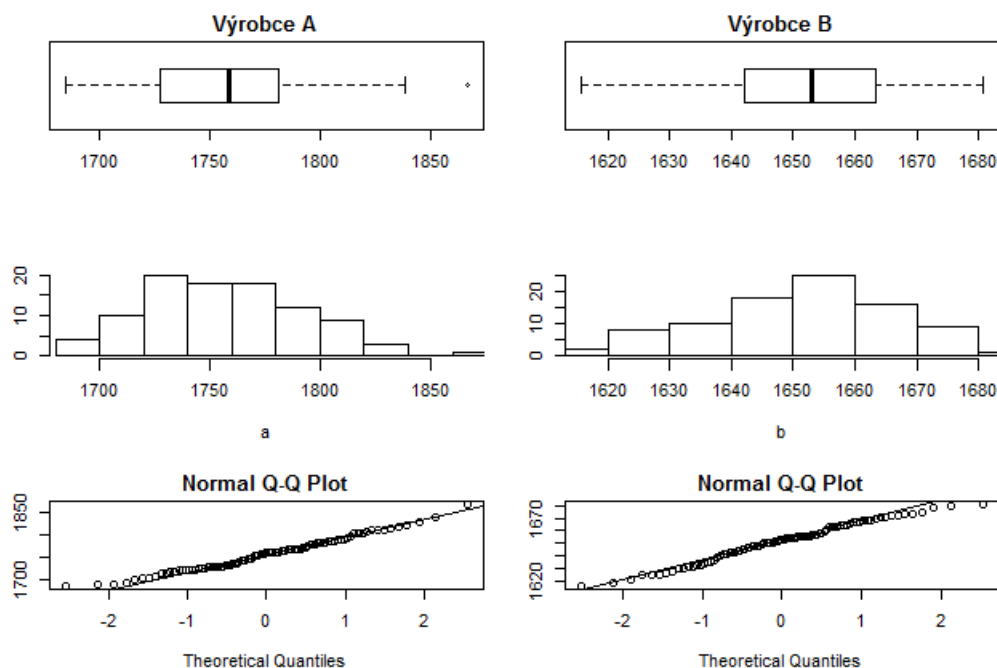


V datech bylo identifikováno jedno odlehlé pozorování, rozhodli jsme se jej v datovém souboru ponechat.

Výrobce A tvrdí, že jeho akumulátory vykazují vyšší průměrnou kapacitu akumulátory výrobce B. Pro ověření tohoto tvrzení byly zjištěny kapacity 95 akumulátorů výrobce A a 89 akumulátorů výrobce B – viz `aku_new.csv`. Ověřte na 5% hladině významnosti, zda lze s tvrzením výrobce A souhlasit.

Řešení:

Explorační analýza:



```
> skewness(a)
[1] 0.3310414
> kurtosis(a)-3
[1] -0.1318987
```

```
> skewness(b)
[1] -0.2836238
> kurtosis(b)-3
[1] -0.5496731
```

Zdá se, že oba výběry pocházejí z normálního rozdělení.

```
> sd(a)
[1] 36.43102
> sd(b)
[1] 15.19928
> max(var(a),var(b))/min(var(a),var(b))
[1] 5.745087
```

Zdá se, že rozptyly jsou různé.

Volba testu dle http://am-nas.vsb.cz/lit40/PRASTA/Statisticka_indukce.pdf)

- Data jsou nezávislá (každý údaj byl změřen na jiné statistické jednotce).
- Ověření normality (exaktní):

H_0 : Data jsou výběrem z normálního rozdělení.

H_A : Data nejsou výběrem z normálního rozdělení. ($\neg H_0$)

výrobce	p-hodnota (Shapirův – Wilkův test)
A	0,330
B	0,198

Na hladině významnosti 5% nelze zamítnout předpoklad normality (viz tabulka).

- Ověření shody rozptylů

$$H_0: \sigma_A^2 = \sigma_B^2$$

$$H_A: \sigma_A^2 \neq \sigma_B^2$$

F test to compare two variances

data: a and b

F = 5.7451, num df = 94, denom df = 88, p-value = 6.661e-15

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

3.793094 8.676799

sample estimates:

ratio of variances

5.745087

Volba testu dle http://am-nas.vsb.cz/lit40/PRASTA/Statisticka_indukce.pdf)

- Data jsou nezávislá (každý údaj byl změřen na jiné statistické jednotce).
- Ověření normality (exaktní):

H_0 : Data jsou výběrem z normálního rozdělení.

H_A : Data nejsou výběrem z normálního rozdělení. ($\neg H_0$)

výrobce	p-hodnota (Shapirův – Wilkův test)
A	0,330
B	0,198

Na hladině významnosti 5 % nelze zamítnout předpoklad normality (viz tabulka).

- Ověření shody rozptylů

$$H_0: \sigma_A^2 = \sigma_B^2$$

$$H_A: \sigma_A^2 \neq \sigma_B^2$$

Na hladině významnosti 5 % zamítáme hypotézu o shodě rozptylů ($p - hodnota < 0,001$, F test).

Na základě ověření předpokladů byl pro analýzu vybrán Aspinové – Welchův test.

$$H_0: \mu_A = \mu_B$$

$$H_A: \mu_A > \mu_B$$

Na hladině významnosti 5% zamítáme hypotézu o shodě středních kapacit akumulátorů výrobců A a B vůči alternativě, že baterie výrobce A vykazují větší průměrnou kapacitu než baterie výrobce B ($p - \text{hodnota} \ll 0,001$, Aspinové – Welchův test). Baterie výrobce A lze v tomto ohledu označit za statisticky významně kvalitnější než baterie výrobce B.

- Doplnění bodového a 95% intervalového odhadu

$$\hat{\mu}_A - \hat{\mu}_B \cong 105 \text{ mAh}$$

Průměrná kapacita akumulátorů výrobce A je o cca 105 mAh větší než průměrná kapacita akumulátorů výrobce B.

S 95% spolehlivostí lze očekávat, že rozdíl průměrná kapacita akumulátorů výrobce A bude nejméně o 98 mAh větší než průměrná kapacita akumulátorů výrobce B.

DĚKUJEME ZA
POZORNOST!

