

PRAVDĚPODOBNOST A STATISTIKA

Vzorové domácí úkoly 1S – 4S

Zadání 1

JMÉNO STUDENTKY/STUDENTA:

SUPERMAN SNAŽIVÝ

OSOBNÍ ČÍSLO:

SNA007

JMÉNO CVIČÍCÍ/CVIČÍCÍHO:

ANALFABETKA HODNÁ

	DATUM ODEVZDÁNÍ	HODNOCENÍ
DOMÁCÍ ÚKOL 1:		
DOMÁCÍ ÚKOL 2:		
DOMÁCÍ ÚKOL 3:		
DOMÁCÍ ÚKOL 4:		
CELKEM:	-----	

Popis datového souboru

Pro dlouhodobý test tužkových NiMH akumulátorů od čtyř různých výrobců (A, B, C, D) bylo od každého výrobce pořízeno 100 akumulátorů. Všechny akumulátory měly deklarovanou kapacitu 2 000 mAh. Kapacita každého akumulátoru byla změřena nejprve po 5 nabíjecích cyklech a následně po 100 nabíjecích cyklech.

V souboru [akumulatory.xlsx](#) jsou pro každý z akumulátorů uvedeny následující údaje: číslo akumulátoru, výrobce (A, B, C, D), změřená kapacita po 5 nabíjecích cyklech a změřená kapacita po 100 nabíjecích cyklech.

Obecné pokyny:

- Úkoly zpracujte dle obecně známých typografických pravidel.
- Všechny tabulky i obrázky musí být opatřeny titulkem.
- Do úkolů ne vkládejte tabulky a obrázky, na něž se v doprovodném textu nebudete odkazovat.
- Bude-li to potřeba, citujte zdroje dle mezinárodně platné citační normy ČSN ISO 690.

Poznámky k preprocessingu:

- Dříve než se pustíme do vlastní analýzy, je vhodné udělat si představu, co je cílem naší analýzy (přečíst si zadání všech úkolů) a s ohledem na ně si připravit data do standardního datového formátu. To, zda si analytik pro tento krok zvolí statistický software (např. R), nějaký tabulkový procesor (např. MS Excel) nebo kombinaci obojího, je zcela na něm.
- V případě tohoto úkolu byla původní datová tabulka

ID	po 5 cyklech				po 100 cyklech			
	Výrobce A	Výrobce B	Výrobce C	Výrobce D	Výrobce A	Výrobce B	Výrobce C	Výrobce D
1	1946,5	2006,5	1881,8	1806,9	1780,4	1654,2	1663,3	1668,4
2	1963,5	1991,5	1890,4	1788,1	1751,4	1663,1	1641,1	1641,9
3	1934,3	1988,8	1865,7	1775	1743,5	1633,3	1621,5	1620

upravena do standardního datového formátu

ID	kapacita po 5 cyklech	kapacita po 100 cyklech	výrobce
1	1946,5	1780,4	A
2	1963,5	1751,4	A
3	1934,3	1743,5	A

a následně doplněna o proměnné, které dle zadaných úkolů budeme dále potřebovat (DÚ 2S – pokles kapacity, DÚ 4S – pokles o více než 10 %):

ID	kapacita po 5 cyklech	kapacita po 100 cyklech	výrobce	pokles kapacity	pokles o více než 10 %
1	1946,5	1780,4	A	166,1	NE
2	1963,5	1751,4	A	212,1	ANO
3	1934,3	1743,5	A	190,8	NE

- V dalším kroku byla u numerických proměnných, které budeme pro analýzu používat, identifikována a odstraněna odlehlá pozorování (v závislosti na výrobcí, tj. **v souladu se zadáním domácích úkolů 1S – 3S**). Nezapomeňte, že byste si neměli přepsat původní data, ale definovat nové proměnné (bez odlehlých pozorování), zde značeno *puvodni_nazev_w*. Pro tento krok je již výhodné použít statistický software.

id	kap5	kap100	vyrobce	pokles	pokles_10	kap5_w	kap100_w	pokles_w
1	1946,5	1780,4	A	166,1	NE	1946,5	1780,4	166,1
2	1963,5	1751,4	A	212,1	ANO	1963,5	1751,4	212,1
3	1934,3	1743,5	A	190,8	NE	1934,3	1743,5	190,8

- Posledním krokem základního preprocessingu je to, že kvalitativní proměnné, které budeme pro analýzu používat, nastavíme jako typ factor, u něž definujeme i požadované pořadí variant proměnných. Tento krok provádíme ve statistickém software. (V tomto úkolu jsme tento krok použili u proměnných *vyrobce* a *pokles_10*).
- V tuto chvíli máme připravena data pro vlastní analýzu. Pro jednotlivé úkoly si z nich můžeme vybírat menší logické celky v závislosti na tématu dílčích úkolů.
- V **domácím úkolu 1S** se máme zabývat srovnáním kapacity akumulátorů výrobce A v závislosti na počtu cyklů. Proto si vyfiltrujeme pouze záznamy o kapacitách akumulátorů výrobce A (jak *kap5*, tak *kap100* – původní data i data po odstranění odlehlých pozorování):

id	kap5	kap100	kap5_w	kap100_w
1	1946,5	1780,4	1946,5	1780,4
2	1963,5	1751,4	1963,5	1751,4
3	1934,3	1743,5	1934,3	1743,5

Takto vyfiltrovaná data si upravíme do formátu vhodného pro zkoumání dané závislosti:

id	kap	cykly	kap_w
1	1946,5	5 cyklů	1946,5
2	1963,5	5 cyklů	1963,5
3	1934,3	5 cyklů	1934,3
1	1780,4	100 cyklů	1780,4
2	1751,4	100 cyklů	1751,4
3	1743,5	100 cyklů	1743,5

- V **domácím úkolu 2S** se máme zabývat srovnáním poklesů kapacit akumulátorů po 100 nabíjecích cyklech pro výrobce A a B, tj. z tabulky

id	kap5	kap100	vyrobce	pokles	pokles_10	kap5_w	kap100_w	pokles_w
1	1946,5	1780,4	A	166,1	NE	1946,5	1780,4	166,1
2	1963,5	1751,4	A	212,1	ANO	1963,5	1751,4	212,1
3	1934,3	1743,5	A	190,8	NE	1934,3	1743,5	190,8

si vybereme pouze proměnné *pokles_w* (budeme analyzovat data po odstranění odlehlých pozorování) a *vyrobce*, přičemž vyfiltrujeme záznamy pouze za výrobce A a B. Takto připravená data jsou již ve standardním datovém formátu.

- V **domácím úkolu 3S** máme srovnávat kapacity akumulátorů po 100 nabíjecích cyklech pro všechny výrobce. Tj. ze základní tabulky ve standardním datovém formátu vybereme (podobně jako v úkolu 2S) pouze proměnné *kap100_w* a *vyrobce*. Nyní máme srovnávat všechny výrobce, tzn. že další filtrování již neprovádíme.
- V **domácím úkolu 4S** se máme věnovat poklesu kapacity akumulátorů o více než 10 % pro výrobce A a D. Ze základní tabulky ve standardním datovém formátu si proto vybereme pouze proměnné *pokles_10* a *vyrobce* a vyfiltrujeme pouze záznamy pro výrobce A a D. Následně se můžeme pustit do vlastní analýzy.

Úkol 1

Pomocí nástrojů explorační analýzy analyzujte kapacity akumulátorů výrobce A po 5 a po 100 nabíjecích cyklech. Data vhodně graficky prezentujte (krabicový graf, histogram, q-q graf) a doplňte následující tabulky a text.

Pro veškeré analýzy byly použity pouze úplné záznamy o jednotlivých akumulátorech.

Poznámka:

V úvodu analýzy byste měli uvést veškeré informace o případné manipulaci s datovým souborem (příčímž manipulací je myšleno případné vyloučení některých záznamů, manipulací se nemyslí úpravy dat do různých struktur). V tomto případě úvodní věta upozorňuje na to, že do analýzy byly zahrnuty pouze ty akumulátory, u nichž jsme zjistili všechny údaje (kapacitu po 5 nabíjecích cyklech, kapacitu po 100 nabíjecích cyklech, výrobce).

Tab. 1: Kapacita akumulátorů (mAh) výrobce A v závislosti na počtu nabíjecích cyklů (souhrnné statistiky)

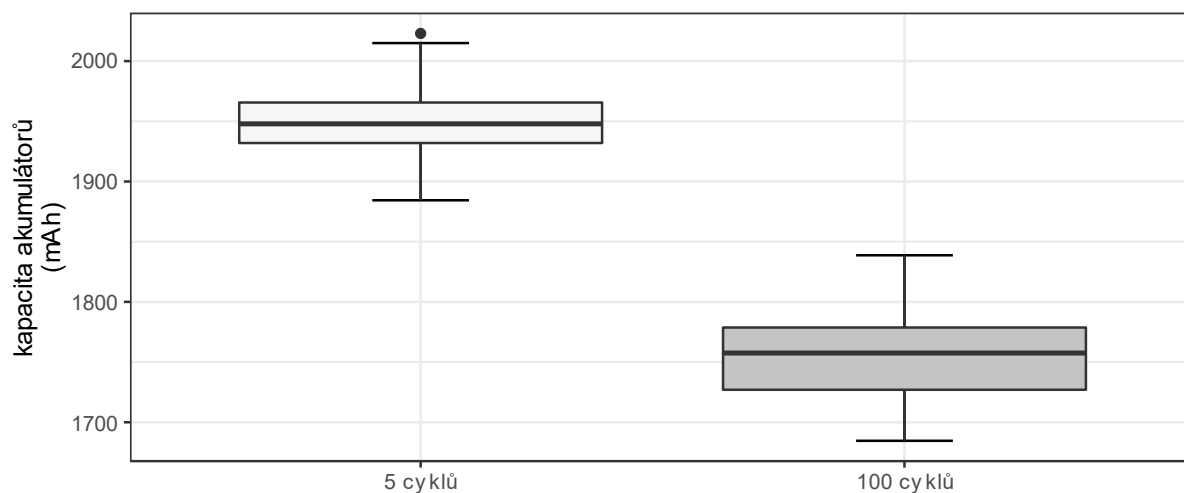
Kapacita akumulátorů, výrobce A (mAh)			Po odstranění odlehlých pozorování	
	po 5 cyklech	po 100 cyklech	po 5 cyklech	po 100 cyklech
rozsah souboru	79	79	78	-
minimum	1 884,4	1 684,6	1 884,4	-
dolní kvartil	1 932,0	1 727,0	1 931,8	-
medián	1 947,9	1 757,5	1 947,3	-
průměr	1 948,4	1 753,2	1 947,5	-
horní kvartil	1 965,6	1 778,7	1 964,9	-
maximum	2 023,0	1 838,7	2 015,0	-
směrodatná odchylka	27,9	34,0	26,7	-
variační koeficient (%)	1,4	1,9	1,4	-
šikmost	0,1	0,2	0,0	-
špičatost	0,0	-0,4	-0,2	-
Identifikace odlehlých pozorování – vnitřní hrady				
dolní mez	1 881,48	1 649,45		
horní mez	2 016,08	1 856,25		

Poznámky:

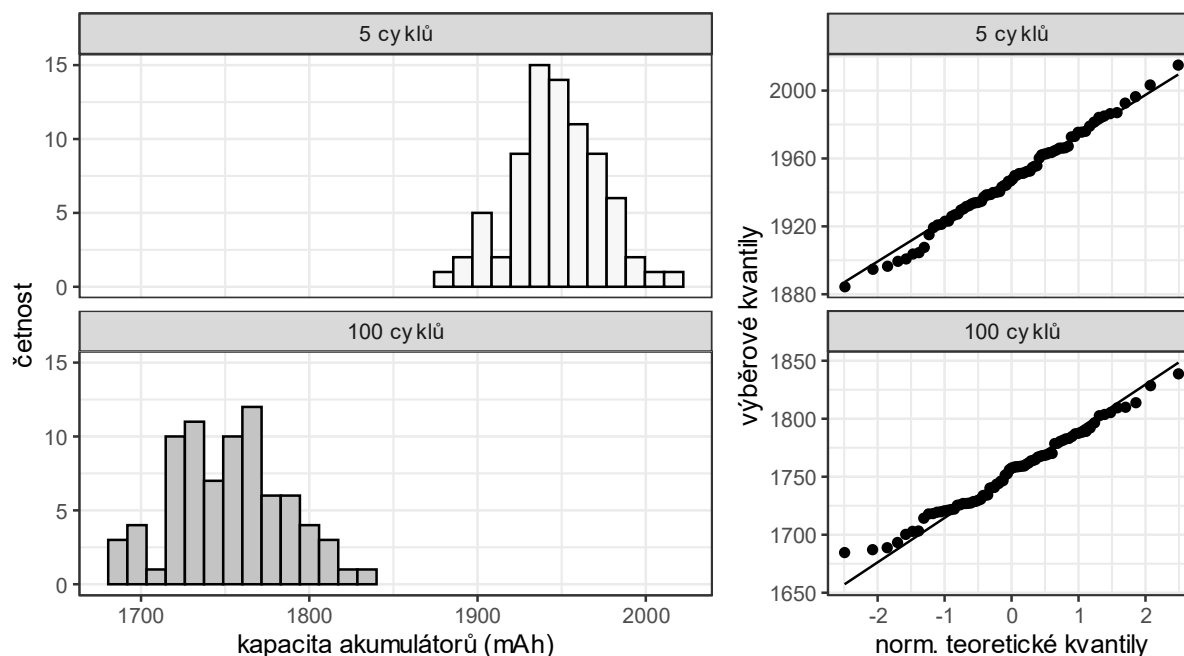
- Tabulky vkládáme zásadně v editovatelném formátu, nikoliv jako obrázky (printscreeny, ...)!
- Do tabulky doplňujeme hodnoty zaokrouhlené dle Manuálu pro zaokrouhlování! Viz <http://homel.vsb.cz/~lit40/PRASTA/Materialy/zaokrouhlovani.pdf>
- Chceme-li číselné charakteristiky dvou proměnných srovnávat (kapacita po 5 cyklech, kapacita po 100 cyklech), preferujeme zaokrouhlení charakteristik na stejný řád.
- Titulky k objektům se skládají z návěští (např. Obr., Tab., ...), čísla objektu, oddělovače (např. :) a vlastního textu titulku.
- Vlastní text titulku musí mít dostatečnou vypovídací schopnost proto, abychom v seznamu objektů, který lze na závěr automaticky vygenerovat, dle titulku snadno identifikovali, co je v objektu znázorněno.

- Návěští titulku volíme, pokud možno nesklonné, abychom se na objekt mohli v textu práce snadno odkazovat prostřednictvím křížových odkazů.
- Titulky tabulek se umísťují nad tabulky.
- Titulky obrázků se umísťují pod obrázky.

Grafická prezentace (krabicový graf, histogram, q-q graf):



Obr. 1: Kapacita akumulátorů (mAh) výrobce A v závislosti na počtu nabíjecích cyklů (krabicový graf, původní data)



Obr. 2: Kapacita akumulátorů (mAh) výrobce A v závislosti na počtu nabíjecích cyklů (histogram, Q-Q graf, data po odstranění odlehlých pozorování)

Poznámky:

- Prioritně se snažíme zařadit vizualizaci původních dat (alespoň v podobě vícenásobných krabicových grafů) i výstupy prezentující data, která následně analyzujeme (po odstranění odlehlých pozorování – vícenásobný krabicový graf, histogramy, Q-Q grafy).
- Pokud odstraněním odlehlých pozorování nedošlo k výrazné změně vzhledu vícenásobného krabicového grafu prezentujícího původní data (odlehlá pozorování vzhled grafu příliš nezkracovala), není nutno tento graf opětovně zařazovat do vizualizace dat po odstranění odlehlých pozorování.
- Po odstranění odlehlých pozorování můžeme někdy při vizualizaci vidět „nová odlehlá pozorování“. To však již nejsou odlehlá pozorování z původních dat, proto se jimi dále nezabýváme.
- Pro větší přehlednost se snažíme využít kombinace grafických výstupů.
- Velikost grafických výstupů by měla odpovídat tomu, co chceme vizualizovat. (Příliš velké obrázky působí neprofesionálně, příliš malé obrázky znemožňují interpretaci výsledků.)
- Vizualizujeme-li kvantitativní proměnnou (kapacitu) dle proměnné kvalitativní (počet nabíjecích cyklů), pokud je to alespoň trochu možné, snažíme se o stejné rozsahy os odpovídajících grafů.
- Grafy stejného typu, které mají společnou osu x (např. histogramy) se snažíme umístit pod sebe.
- Grafy stejného typu, které mají spol. osu y (např. krabicové grafy) se snažíme umístit vedle sebe.
- Pro odlišení výstupů pro jednotlivé varianty kvalitativní proměnné využíváme (pokud možno) odstíny šedi, popř. pokud to není nutno, barevné rozlišení vůbec nevyužíváme.

Výsledky popisné statistiky lze vidět v Tab. 1 a na Obr. 2.

Analýza kapacity akumulátorů výrobce A po 5 nabíjecích cyklech

Během testu byla měřena kapacita 79 kusů akumulátorů výrobce A. Naměřená kapacita po 5 nabíjecích cyklech se pohybovala v rozmezí 1 884,4 mAh až 2 023,0 mAh. Kapacity ležící mimo interval (1 881,48; 2 016,08) mAh byly identifikovány jako odlehlá pozorování (viz Obr. 1) a nebudou zahrnuty do dalšího zpracování. Za příčinu vzniku odlehlých pozorování je možno považovat nestandardně kvalitní akumulátor. Dále uvedené výsledky tedy pocházejí z analýzy kapacit 78 kusů akumulátorů. Jejich průměrná kapacita byla 1 947,5 mAh, směrodatná odchylka pak 26,7 mAh. U poloviny testovaných akumulátorů kapacita nepřekročila 1 947,3 mAh. V polovině měření se kapacita pohybovala v rozmezí 1 931,8 mAh až 1 964,9 mAh. Vzhledem k hodnotě variačního koeficientu (1,4 %) lze analyzovaný soubor považovat za homogenní.

Analýza kapacity akumulátorů výrobce A po 100 nabíjecích cyklech

Během testu byla měřena kapacita 79 kusů akumulátorů výrobce A. Naměřená kapacita po 100 nabíjecích cyklech se pohybovala v rozmezí 1 684,6 mAh až 1 838,7 mAh. Žádné z měření nebylo identifikováno jako odlehlé pozorování. Dále uvedené výsledky tedy pocházejí z analýzy kapacit 79 kusů akumulátorů. Jejich průměrná kapacita byla 1 753,2 mAh, směrodatná odchylka pak 34,0 mAh. U poloviny testovaných akumulátorů kapacita nepřekročila 1 757,5 mAh. V polovině měření se kapacita pohybovala v rozmezí 1 727,0 mAh až 1 778,7 mAh. Vzhledem k hodnotě variačního koeficientu (1,9 %) lze analyzovaný soubor považovat za homogenní.

Poznámky:

- Na každý objekt (tabulku, obrázek) se v práci musíme odkazovat. V opačném případě to je posuzováno tak, že objekt je v práci nadbytečný.
- Čísla uvedená v textu musí přesně korespondovat s údaji uvedenými v Tab. 1.
- Za homogenní považujeme proměnnou, jejíž variační koeficient je menší než 50 %.

Ověření normality kapacity akumulátorů výrobce A po 5 nabíjecích cyklech na základě explorační analýzy

Na základě grafického zobrazení (viz Obr. 2) a výběrové šikmosti a špičatosti (viz Tab. 1, výběrová šikmost i špičatost leží v intervalu $(-2; 2)$) lze předpokládat, že kapacita akumulátorů výrobce A po 5 nabíjecích cyklech má normální rozdělení. Dle pravidla 3σ lze tedy očekávat, že přibližně 95 % akumulátorů bude mít kapacitu v rozmezí 1 894,1 mAh až 2 000,9 mAh.

Ověření normality kapacity akumulátorů výrobce A po 100 nabíjecích cyklech na základě explorační analýzy

Na základě grafického zobrazení (viz Obr. 2) a výběrové šikmosti a špičatosti (viz Tab. 1, výběrová šikmost i špičatost leží v intervalu $(-2; 2)$) lze předpokládat, že kapacita akumulátorů výrobce A po 100 nabíjecích cyklech má normální rozdělení. Dle pravidla 3σ lze tedy očekávat, že přibližně 95 % akumulátorů bude mít kapacitu v rozmezí 1 685,2 mAh až 1 821,2 mAh.

Poznámky:

- Předpokládáme-li, že proměnná **nemá** normální rozdělení, používáme pro odhad rozsahu proměnné Čebyševovu nerovnost. (Dle uváděné p-stí dosadíme příslušné k , viz níže uvedená tabulka.)
- Předpokládáme-li, že proměnná **má** normální rozdělení, používáme pro odhad rozsahu proměnné pravidlo 3σ . (Dle uváděné p-stí dosadíme příslušné k , viz níže uvedená tabulka.)
- V obou případech jako odhad μ použijeme průměr a jako odhad σ směrodatnou odchylku.

k	$P(\mu - k\sigma < X < \mu + k\sigma)$	
	Čebyševova nerovnost	Pravidlo 3σ
1	>0	0,682
2	$>0,75$	0,954
3	$>0,89$	0,997

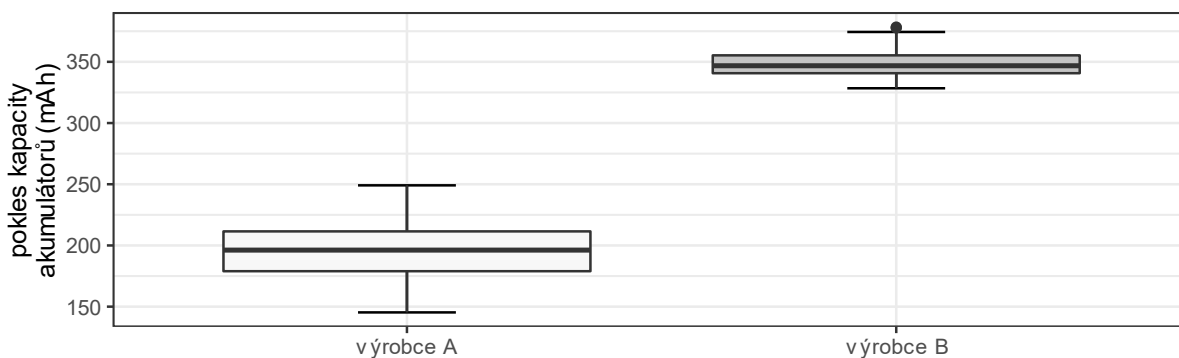
Úkol 2

Porovnejte poklesy kapacit akumulátorů výrobců A a B po 100 nabíjecích cyklech. Nezapomeňte, že použité metody mohou vyžadovat splnění určitých předpokladů. Pokud tomu tak bude, okomentujte splnění/nesplnění těchto předpokladů jak na základě explorační analýzy (např. s odkazem na histogram apod.), tak exaktně pomocí metod statistické indukce.

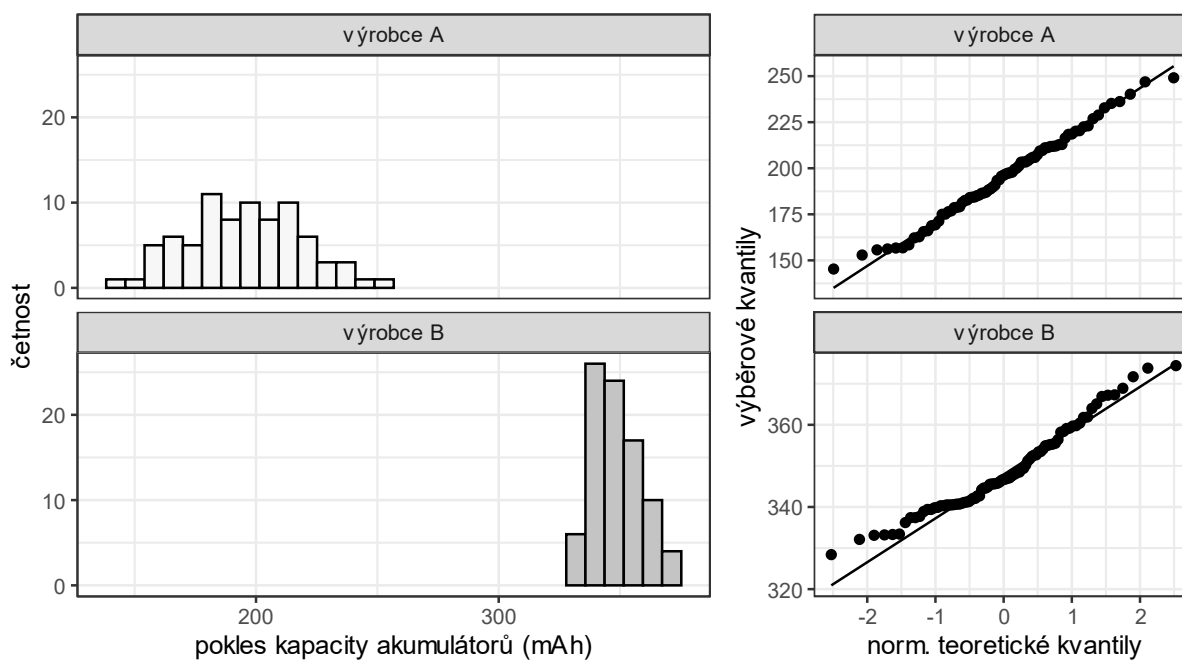
- a) Graficky prezentujte srovnání poklesů kapacit akumulátorů výrobců A a B po 100 nabíjecích cyklech (vícenásobný krabicový graf, histogramy, q-q grafy). Srovnání okomentujte (včetně informace o případné manipulaci s datovým souborem).

U výrobce B bylo pozorováno jedno odlehlé pozorování (viz Obr. 3), toto jsme se rozhodli z dalšího zpracování vypustit. Dle vizualizace srovnání poklesů kapacit akumulátorů výrobců A a B (viz Obr. 3, Obr. 4) lze usoudit, že u výrobce B došlo po 100 nabíjecích cyklech k výrazně vyššímu poklesu kapacity akumulátorů než u výrobce A.

Poznámka: Komentujeme: Manipulaci s datovým souborem a předběžný úsudek o řešené úloze (srovnání poklesu kapacit u výrobců A a B).



Obr. 3: Srovnání poklesů kapacit akumulátorů (mAh) výrobců A a B (krabicový graf, původní data)



Obr. 4: Srovnání poklesů kapacit akumulátorů (mAh) výrobců A a B (histogramy, Q-Q grafy, data po odstranění odlehlých pozorování)

- b) Na hladině významnosti 5 % rozhodněte, zda jsou střední poklesy (popř. mediány poklesů) kapacit akumulátorů výrobců A a B statisticky významné. K řešení využijte bodové a intervalové odhady i testování hypotéz. Výsledky okomentujte.

Dle prezentovaných Q-Q grafů (viz Obr. 4) lze usuzovat že poklesy kapacit u výrobce A lze modelovat normálním rozdělením, o normalitě poklesů kapacit akumulátorů výrobce B je na základě Q-Q grafu obtížné rozhodnout. Šikmost i špičatost (viz Tab. 2) odpovídají normálnímu rozdělení.

Dle Shapirova-Wilkova testu nelze na hladině významnosti 0,05 pokles kapacit akumulátorů výrobce B modelovat normálním rozdělením (viz Tab. 2), proto budou pro popis poklesu kapacit použity neparametrické metody.

Rozdělení poklesu kapacit akumulátorů lze u obou výrobců považovat za symetrické (viz šikmosti, výsledky testu symetrie (Tab. 2) a tvar histogramů (Obr. 4)), proto lze pro intervalové odhady a test významnosti mediánu v případě obou výrobců použít Wilcoxonovu testovou statistiku.

Tab. 2: Ověření normality a symetrie poklesu kapacity akumulátorů dle výrobce

	šikmost	stand. špičatost	Shapirův-Wilkův test (p-hodnota)	Test symetrie (p-hodnota)
výrobce A	0,1	-0,5	0,789	0,630
výrobce B	0,6	-0,2	0,013	0,053

Poznámky:

- Chystáme-li se použít metody statistické indukce (intervalové odhady, testování hypotéz), měli bychom nejdříve ověřit předpoklady pro jejich použití).
- Normalitu dat je vhodné posoudit jak pomocí explorační analýzy (tvar histogramu, Q-Q graf, hodnoty šikmosti a stand. špičatosti (u norm. rozdělení očekáváme hodnoty v intervalu $(-2;2)$), tak exaktně na základě vhodného testu (např. Shapirův – Wilkův test)).
- Test normality:
 H_0 : Data jsou výběrem z normálního rozdělení.
 H_A : Data nejsou výběrem z normálního rozdělení. ($\neg H_0$)
 Tj. je-li $p - \text{hodnota} < \alpha$ (zvolená hladina významnosti), zamítáme předpoklad normality.
- Také symetrii rozdělení je vhodné posoudit jak pomocí explorační analýzy (na základě tvaru histogramu a šikmosti), tak exaktně na základě testu (test symetrie).
- Test symetrie:
 H_0 : Data jsou výběrem ze symetrického rozdělení.
 H_A : Data nejsou výběrem ze symetrického rozdělení. ($\neg H_0$)
 Tj. je-li $p - \text{hodnota} < \alpha$ (zvolená hladina významnosti), zamítáme předpoklad symetrie.
- Není-li hladina významnosti součástí zadání, volíme ji před začátkem analýzy. Naši volbu uvedeme buď v popisu metodiky (např. „Veškeré vyhodnocení statistické indukce je prováděno na hladině významnosti 0,05.“), nebo ji uvádíme u každého výstupu.

- Nezapomínejme, že p-hodnotu zaokrouhlujeme o 1 desetinné místo více než je zvolená hladina významnosti. (Tj. volíme-li hladinu významnosti 0,05, p-hodnotu zaokrouhlujeme na 3 des. místa.)
- Vychází-li po zaokrouhlení p-hodnota rovna nule, zapisujeme při hladině významnosti 0,05: p-hodnota < 0,001 (Obecně: p-hodnota je menší než nejmenší číslo s danou přesností).

Vzhledem k tomu, že očekáváme kladné poklesy kapacit (s rostoucím počtem nabíjecích cyklů se kapacita akumulátorů snižuje), volíme levostranné intervalové odhady / levostranné testy.

U výrobce A lze očekávat, že polovina akumulátorů bude vykazovat pokles kapacity menší než cca 196,2 mAh. 95% levostranný intervalový odhad mediánu poklesu kapacity u výrobce A je (190,4; ∞) mAh. Intervalový odhad, stejně jako p-hodnota Wilcoxonova testu, ukazují, že medián poklesu kapacity je statisticky významně větší než nula. Tj. na hladině významnosti 0,05 lze pokles kapacity akumulátorů výrobce A považovat za statisticky významný. Odhady pro výrobce B (viz Tab. 3) lze interpretovat obdobně.

Tab. 3: Odhad mediánu poklesu kapacit akumulátorů (mAh) dle výrobce a test významnosti poklesu kapacit

	bodový odhad (mAh)	95% levostranný intervalový odhad (mAh)	Wilcoxonův levostranný test (p-hodnota)
výrobce A	196,2	(190,4; ∞)	<0,001
výrobce B	346,7	(346,1; ∞)	<0,001

Poznámky:

- Při analýze vícerozměrných dat (pokles kapacity v závislosti na výrobci) preferujeme jednotný přístup k analýze jednotlivých podskupin (pokles kapacity akumulátorů výrobce A, pokles kapacity akumulátorů výrobce B). U výrobce A bychom mohli modelovat střední hodnotu poklesu kapacity (předpoklad normality nebyl zamítnut) i medián poklesu kapacity (horší volba vzhledem k možnosti modelování střední hodnoty), u výrobce B musíme přistoupit k modelování mediánu poklesu kapacity (předpoklad normality byl zamítnut). Vzhledem k tomu, že výstupy analýzy chceme vzájemně srovnávat, zvolili jsme pro modelování u obou výrobců medián poklesu kapacity.
- Je-li to možné, používáme pro ověření statistické hypotézy, intervalové odhady i čistý test významnosti (p-hodnotu) a závěry vyhodnocujeme najednou.
- Posuzujeme-li statistickou významnost nějaké změny (např. pokles kapacity), ověřujeme, zda se střední hodnota μ (popř. medián $x_{0,5}$) této změny statisticky významně liší od nuly, popř. zda je střední hodnota μ (popř. medián $x_{0,5}$) této změny statisticky významně větší / menší (dle definice změny) než nula.

	$H_0: \mu = 0 \ (x_{0,5} = 0)$		
	$H_A: \mu \neq 0 \ (x_{0,5} \neq 0)$	$H_A: \mu > 0 \ (x_{0,5} > 0)$	$H_A: \mu < 0 \ (x_{0,5} < 0)$
Typ IO	oboustranný IO stř. hodnoty (oboustranný IO mediánu)	levostranný IO stř. hodnoty (levostranný IO mediánu)	pravostranný IO stř. hodnoty (pravostranný IO mediánu)
Zápis IO	(dolní mez; horní mez)	(dolní mez; ∞)	($-\infty$; horní mez)
Vhodný test	oboustranný t-test (oboustranný Wilcoxonův test)	levostranný t-test (levostranný Wilcoxonův test)	pravostranný t-test (pravostranný Wilcoxonův test)

- Tj. neobsahuje-li nalezený intervalový odhad nulu, zamítáme zvolenou H_0 . Jinými slovy, příslušná změna je (na dané hladině významnosti) statisticky významná.

- Nezapomínejme, že u intervalových odhadů zaokrouhlujeme dolní mez dolů a horní mez nahoru na požadovaný řád.
 - Je-li $p - \text{hodnota} < \alpha$ (zvolená hladina významnosti), zamítáme zvolenou H_0 , tj. příslušná změna je (na dané hladině významnosti) statisticky významná.
 - Závěr plynoucí z čistého testu významnosti musí být v souladu se závěrem stanoveným pomocí intervalového odhadu.
 - Chceme-li použít Wilcoxonův test (test mediánu), nezapomínejme alespoň na základě popisné statistiky posoudit symetrii dat.
- c) Na hladině významnosti 5 % rozhodněte, zda je rozdíl středních hodnot (popř. mediánů) poklesů kapacit akumulátorů (po 100 cyklech) výrobců A a B statisticky významný. K řešení využijte bodový a intervalový odhad i čistý test významnosti. Výsledky okomentujte.

Vzhledem k zamítnutí předpokladu normality poklesu kapacit akumulátorů výrobce B (viz Tab. 2) budeme i nadále pokračovat v aplikaci neparametrických metod.

Tvar rozdělení poklesů kapacit akumulátorů výrobců A i B je srovnatelný (viz Obr. 4), pro odhad a test významnosti rozdílu mediánů proto použijeme metody založené na Mannově-Whitneyho statistice.

Tab. 4: Srovnání mediánu poklesu kapacit akumulátorů (mAh) výrobce A ($x_{0,5}^A$) a výrobce B ($x_{0,5}^B$)

Bodový odhad $x_{0,5}^B - x_{0,5}^A$ (mAh)	150,5
95% levostranný intervalový odhad $x_{0,5}^B - x_{0,5}^A$ (mAh)	(148,4; ∞)
Mannův-Whitneyho levostranný test (p-hodnota)	<0,001

U akumulátorů výrobce B lze očekávat medián poklesu kapacit akumulátorů o cca 150,5 mAh vyšší než u výrobce A. Odpovídající 95% levostranný intervalový odhad tohoto rozdílu je (148,4; ∞) mAh. Intervalový odhad, stejně jako Mannův-Whitneyho test ($x_{\text{OBS}} = 5,3$, $df_1 = 78$, $df_2 = 86$, $p\text{-hodnota} < 0,001$) ukazují, že medián poklesu kapacit akumulátorů výrobce B je statisticky významně vyšší než u výrobce A (viz Tab. 4).

Poznámky:

- Posuzujeme-li, zda nějaká proměnná (např. pokles kapacity) statisticky významně závisí na nějakém dichotomickém faktoru (např. výrobci (A/B)), ověřujeme, zda se statisticky významně liší příslušné střední hodnoty μ_A, μ_B (popř. mediány $x_{0,5}^A, x_{0,5}^B$), tj. zda se rozdíl těchto parametrů statisticky významně liší od nuly, popř. zda je větší / menší (dle toho, co nás zajímá) než nula.

	$H_0: \mu_A - \mu_B = 0, \text{ tj. } \mu_A = \mu_B \quad (x_{0,5}^A - x_{0,5}^B = 0, \text{ tj. } x_{0,5}^A = x_{0,5}^B)$		
	$H_A: \mu_A - \mu_B \neq 0, \text{ tj. } \mu_A \neq \mu_B$ ($x_{0,5}^A - x_{0,5}^B \neq 0, \text{ tj. } x_{0,5}^A \neq x_{0,5}^B$)	$H_A: \mu_A - \mu_B > 0, \text{ tj. } \mu_A > \mu_B$ ($x_{0,5}^A - x_{0,5}^B > 0, \text{ tj. } x_{0,5}^A > x_{0,5}^B$)	$H_A: \mu_A - \mu_B < 0, \text{ tj. } \mu_A < \mu_B$ ($x_{0,5}^A - x_{0,5}^B < 0, \text{ tj. } x_{0,5}^A < x_{0,5}^B$)
Typ IO	oboustranný IO rozdílu stř. hodnot (oboustranný IO rozdílu mediánů)	levostranný IO rozdílu stř. hodnot (levostranný IO rozdílu mediánů)	pravostranný IO rozdílu stř. hodnot (pravostranný IO rozdílu mediánů)
Zápis IO	(dolní mez; horní mez)	(dolní mez; ∞)	($-\infty$; horní mez)
Vhodný test	oboustranný t-test / Aspinové-Welchův test (oboustranný Mannův-Whitneyho test)	levostranný t-test / Aspinové-Welchův test (levostranný Mannův-Whitneyho test)	pravostranný t-test / Aspinové-Welchův test (pravostranný Mannův-Whitneyho test)

- Zvolíme-li ověřovaný rozdíl (např.: $\mu_A - \mu_B$, popř. $\mu_B - \mu_A$) tak, aby jeho bodový odhad vycházel kladný, bude pro nás snazší výsledky slovně interpretovat.
- Vzhledem k formulaci zadání bychom mohli pro ověření hypotézy použít oboustranný intervalový odhad, popř. oboustranný test. Pro praktické využití nás však spíše zajímá, zda jsou poklesy kapacit výrobce B statisticky významně větší než u výrobce A, proto volíme jednostranný intervalový odhad / jednostranný test.
- Chceme-li použít Mannův-Whitneyho test, měli bychom na základě popisné statistiky posoudit, zda je tvar rozdělení sledované proměnné pro obě varianty faktoru srovnatelný (např. srovnání tvaru obálek histogramů).
- Chceme-li posuzovat shodu středních hodnot, musíme se rozhodnout, zda použít testovou statistiku t (předpoklad shody rozptylů), nebo Aspinové-Welchovu testovou statistiku. O tom rozhodneme na základě ověření shody rozptylů.

Test shody rozptylů:

$H_0: \sigma_A^2 = \sigma_B^2$, tj. $\frac{\sigma_A^2}{\sigma_B^2} = 1$ (Pro snadnou interpretaci bodového a intervalového odhadu zadáváme jako první (na levé straně rovnosti) ten rozptyl, jehož bodový odhad je větší.)

$H_A: \sigma_A^2 \neq \sigma_B^2$, tj. $\frac{\sigma_A^2}{\sigma_B^2} \neq 1$

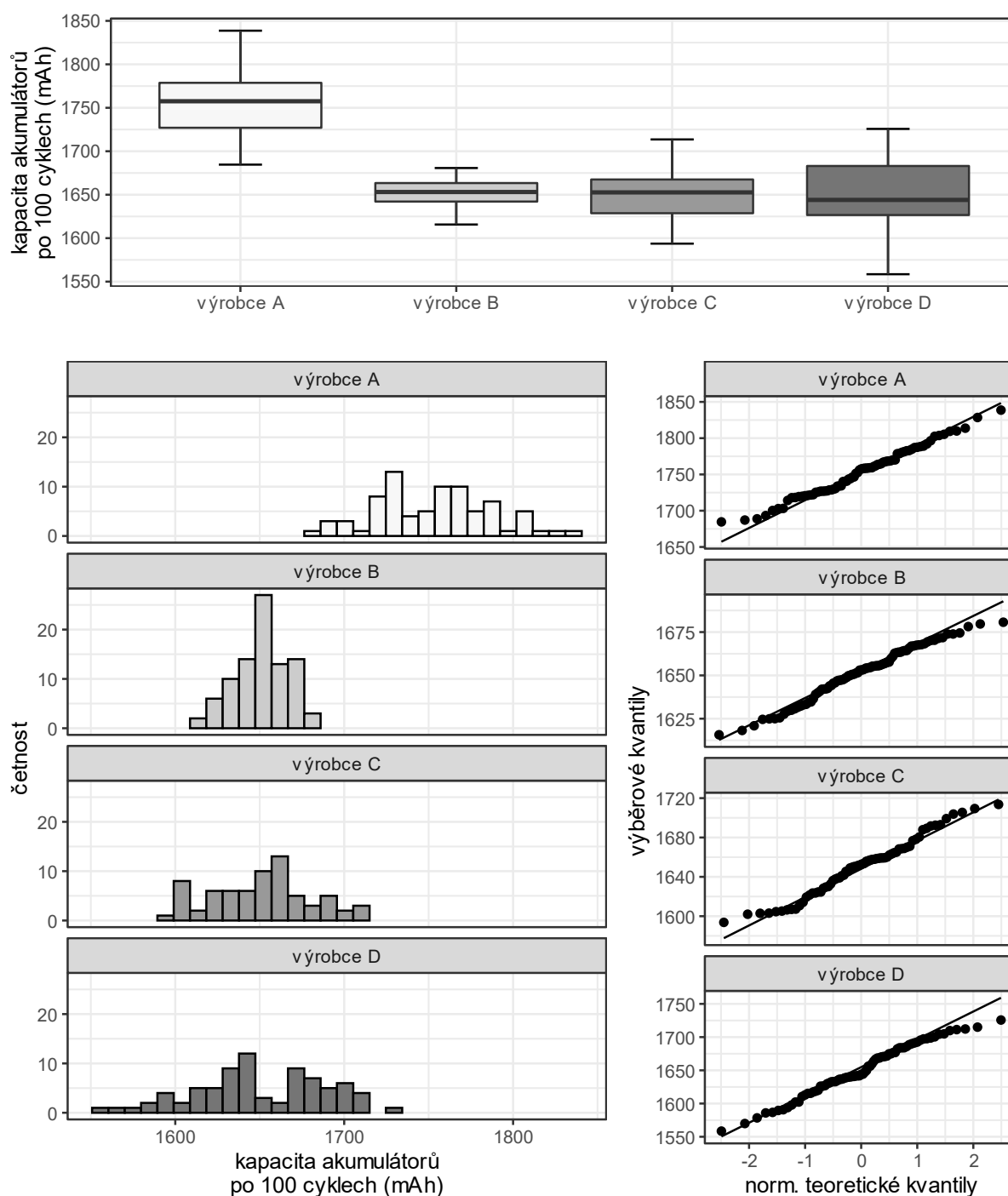
Tj. je-li $p - \text{hodnota} < \alpha$ (zvolená hladina významnosti), zamítáme předpoklad o shodě rozptylů.

Jak toto komentovat? Např. Z důvodů zamítnutí shody rozptylů (F-test, $p\text{-hodnota} = 0,034$) budeme pro úsudky o rozdílu středních hodnot využívat Aspinové-Welchovu statistiku.

Úkol 3

Na hladině významnosti 5 % rozhodněte, zda se kapacity akumulátorů po 100 nabíjecích cyklech liší v závislosti na tom, od kterého výrobce pocházejí. Posouzení proveďte nejprve na základě explorační analýzy a následně pomocí vhodného statistického testu včetně ověření potřebných předpokladů. V případě, že se kapacity statisticky významně liší, určete, či akumulátory se odlišují od ostatních.

a) Daný problém vhodným způsobem graficky prezentujte (vícenásobný krabicový graf, histogramy, q-q grafy).



Obr. 5: Srovnání kapacit akumulátorů (mAh) po 100 nabíjecích cyklech dle výrobců (původní data)

Analyzovaná data neobsahují žádná odlehlá pozorování. Dle Obr. 5 se zdá, že akumulátory výrobce A vykazují po 100 nabíjecích cyklech vyšší kapacitu než akumulátory ostatních tří výrobců.

Poznámka: Komentujeme: Případnou manipulaci s daty a úsudek o řešené úloze (srovnání kapacit po 100 cyklech dle výrobců).

b) Ověřte normalitu a symetrii kapacit akumulátorů po 100 nabíjecích cyklech všech čtyř výrobců (empiricky i exaktně).

Dle prezentovaných Q-Q grafů (viz Obr. 5) lze usuzovat, že kapacity po 100 nabíjecích cyklech lze u všech výrobců modelovat normálním rozdělením. Rovněž šikmost i špičatost (viz Tab. 5) odpovídají normálnímu rozdělení.

Dle Shapirova-Wilkova testu nelze na hladině významnosti 0,05 předpoklad normality pro kapacity akumulátorů po 100 cyklech u žádného z výrobců zamítnout (viz Tab. 5).

Rozdělení kapacit akumulátorů po 100 nabíjecích cyklech lze u všech výrobců považovat za symetrické (viz šikmosti, výsledky testu symetrie (Tab. 5) a tvar histogramů (Obr. 5)).

Tab. 5: Ověření normality a symetrie kapacity akumulátorů (mAh) po 100 nabíjecích cyklech dle výrobce

	šikmost	stand. špičatost	Shapirův-Wilkův test (p-hodnota)	Test symetrie (p-hodnota)
výrobce A	0,2	-0,4	0,582	0,148
výrobce B	-0,3	-0,6	0,197	0,134
výrobce C	0,1	-0,6	0,197	0,352
výrobce D	-0,2	-0,7	0,271	0,058

c) Ověřte homoskedasticitu (shodu rozptylů) kapacit akumulátorů po 100 nabíjecích cyklech všech čtyř výrobců (empiricky i exaktně).

Dle Obr. 5 a poměru největšího a nejmenšího rozptylu kapacit akumulátorů po 100 nabíjecích cyklech ($\frac{s_D^2}{s_B^2} \cong 6,4$) se zdá, že srovnávané rozptyly nelze považovat za srovnatelné.

Na základě Bartlettova testu ($\chi_{OBS} = 68,1$, $df = 3$, $p\text{-hodnota} < 0,001$) zamítáme předpoklad o shodě rozptylů.

Poznámky:

- Empirické posouzení shody rozptylů děláme na základě vícenásobného krabicového grafu (srovnáváme šířku krabic) a na základě poměru největšího a nejmenšího výběrového rozptylu. Je-li tento poměr větší než 2, usuzujeme na nesrovnatelné rozptyly.
- Test shody rozptylů (test homoskedasticity):

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

$$H_A: \neg H_0 \text{ (Alespoň jedna dvojice rozptylů se statisticky významně liší.)}$$

Tj. je-li $p\text{-hodnota} < \alpha$ (zvolená hladina významnosti), zamítáme předpoklad o shodě rozptylů.

Pro ověření shody rozptylů ve vícevýběrovém testu volíme mezi Bartlettovým testem (předpoklad normality) a Leveneho testem.

- d) Určete bodové a 95% intervalové odhady střední hodnoty (resp. mediánu) kapacit akumulátorů po 100 nabíjecích cyklech pro všechny srovnávané výrobce. (Nezapomeňte na ověření předpokladů pro použití příslušných intervalových odhadů.)

Předpoklad o symetrii rozdělení nelze zamítnout (viz bod a). Vzhledem k zamítnutí předpokladu o shodě rozptylů proto pro srovnání kapacit akumulátorů po 100 nabíjecích cyklech použijeme Kruskalův-Wallisův test, který ověřuje shodu mediánů. Z tohoto důvodu využijeme mediány rovněž v rámci odhadu měř polohy. Pro určení intervalového odhadu byla, s ohledem na nezamítnutí předpokladu symetrie, použita Wilcoxonova statistika (viz Tab. 6).

Tab. 6: Odhad mediánů kapacit akumulátorů (mAh) po 100 nabíjecích cyklech dle výrobce

	bodový odhad (mAh)	95% intervalový odhad (mAh)
výrobce A	1 757,5	(1 744,5;1 761,2)
výrobce B	1 653, 1	(1 648,3;1 655,1)
výrobce C	1 652, 7	(1 642,3;1 657,6)
výrobce D	1 644,0	(1 641,0;1 660,3)

U poloviny akumulátorů výrobce A očekáváme kapacitu po 100 nabíjecích cyklech nižší než cca 1 757,5 mAh. 95% intervalový odhad tohoto mediánu je (1 744,5;1 761,2) mAh. Výsledky výrobců B, C a D lze interpretovat obdobně. Výrobce A po 100 nabíjecích cyklech vykazuje kapacitu akumulátorů o více než 100 mAh vyšší než ostatní výrobci. Výrobci B, C a D se z hlediska tohoto parametru liší v řádu 10 mAh.

Poznámka: V komentáři ukazujeme, jak číst výstupy uváděné v Tab. 6 a zároveň popisujeme, čeho bychom si na těchto výstupech v rámci srovnání jednotlivých výrobců měli všimnout.

- e) Čistým testem významnosti ověřte, zda je pozorovaný rozdíl středních hodnot (resp. mediánů) kapacit akumulátorů po 100 nabíjecích cyklech statisticky významný na hladině významnosti 5 %. Pokud ano, zjistěte, zda lze některé skupiny výrobců označit (z hlediska kapacit akumulátorů po 100 cyklech) za homogenní. (Nezapomeňte na ověření předpokladů pro použití zvoleného testu.)

Na hladině významnosti 0,05 zamítáme hypotézu o shodě mediánů (Kruskalův-Wallisův test, $\chi_{OBS} = 172$, $df = 3$, $p\text{-hodnota} < 0,001$), tj. mediány kapacit akumulátorů po 100 nabíjecích cyklech se statisticky významně liší.

Dle Dunnové post hoc analýzy můžeme konstatovat, že akumulátory výrobce A vykazují statisticky významně vyšší kapacity po 100 nabíjecích cyklech než akumulátory zbylých tří výrobců. Výrobci B, C a D se mezi sebou z hlediska kapacity akumulátorů po 100 nabíjecích cyklech statisticky významně neliší.

Poznámky:

- Pro srovnání měr polohy ve vícevýběrových testech používáme **v případě nezávislých výběrů** (každý údaj je měřen na jiné statistické jednotce) test ANOVA nebo Kruskalův-Wallisův test. Vhodný test volíme na základě ověření předpokladů, přičemž standardně preferujeme užití parametrického testu, tj. testu ANOVA.

	ANOVA	Kruskalův-Wallisův test
	$H_0: \mu_1 = \mu_2 = \dots = \mu_k$ $H_A: \neg H_0$	$H_0: x_{0,5}^1 = x_{0,5}^2 = \dots = x_{0,5}^k$ $H_A: \neg H_0$
Předpoklady testu	<ul style="list-style-type: none"> normalita dat ve všech výběrech shoda rozptylů 	<ul style="list-style-type: none"> symetrie rozdělení ve všech výběrech
Metoda pro post hoc analýzu	Tukeyho test	Dunnové test

- Tj. je-li $p - \text{hodnota} < \alpha$ (zvolená hladina významnosti), zamítáme předpoklad o shodě středních hodnot / mediánů.
- Pokud předpoklad o shodě středních hodnot / mediánů nezamítneme, test je ukončen konstatováním, že střední hodnoty / mediány sledované proměnné (kapacity akumulátorů) se na příslušné hladině významnosti v závislosti na daném faktoru (výrobci) statisticky významně neliší.
- Pokud předpoklad o shodě středních hodnot / mediánů zamítneme, měli bychom pomocí post hoc analýzy (vícenásobného porovnávání) zjistit, zda mezi faktory nelze najít podskupiny, v rámci nichž se střední hodnoty / mediány sledované proměnné statisticky významně neliší. K tomu využijeme příslušnou metodu pro post hoc analýzu a výsledky interpretujeme například na základě písmenkového schéma.

Úkol 4

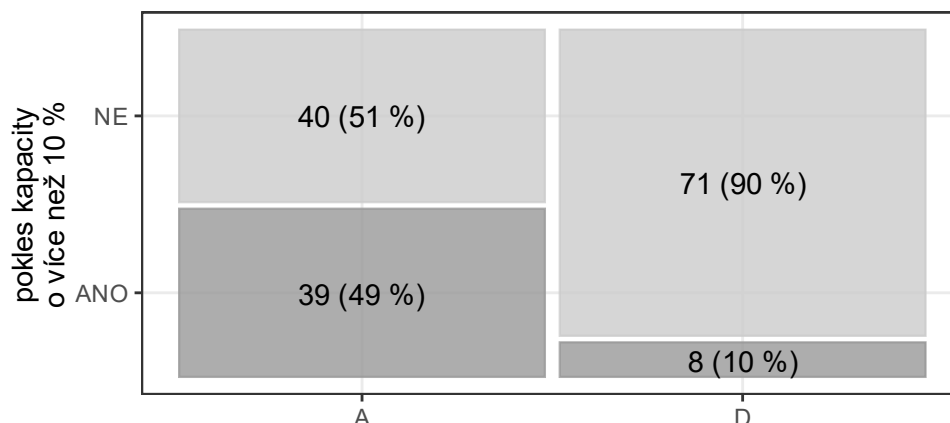
U každého z akumulátorů výrobců A a D zjistěte, zda kapacita po 100 nabíjecích cyklech poklesla o více než 10 % původní kapacity (po 5 cyklech). (Definujte si novou dichotomickou proměnnou, která bude nabývat hodnot {ANO, NE}.)

- a) Srovnajte četnosti poklesu kapacity akumulátorů o více než 10 % pro výrobce A a D. Výsledky prezentujte pomocí kontingenční tabulky, vhodného grafu a vhodné míry kontingence. Úsudky o pozorované závislosti komentujte.

Do analýzy bylo zařazeno 79 (50 %) akumulátorů výrobce A a 79 (50 %) akumulátorů výrobce D. Pokles kapacity o více než 10 % vykazovalo 39 (83 %) akumulátorů výrobce A a pouze 8 (17 %) výrobce D (viz Tab. 7, Obr. 6). Akumulátory výrobce D se z tohoto hlediska jeví jako kvalitnější. Na základě Obr. 6 a Cramerova $V(0,415)$ lze závislost mezi výskytem poklesu kapacity o více než 10 % a výrobcem akumulátorů hodnotit jako středně silnou.

Tab. 7: Srovnání výskytu poklesu kapacity o více než 10 % pro výrobce A a D
(v závorce jsou uvedeny řádkové relativní četnosti, ve sloupci Celkem sloupcové relativní četnosti)

výrobce \ pokles kapacity o více než 10 %	ANO	NE	Celkem
A	39 (49 %)	40 (51 %)	79 (50 %)
D	8 (10 %)	71 (90 %)	79 (50 %)
Celkem	47 (30 %)	111 (70 %)	158 (100 %)



Obr. 6: Srovnání výskytu poklesu kapacity o více než 10 % pro výrobce A a D (mozaikový graf)

- b) Určete bodový i 95% intervalový odhad pravděpodobnosti poklesu kapacity akumulátorů o více než 10 % pro výrobce A.

Jak již bylo uvedeno, pokles kapacity akumulátorů o více než 10 % vykazovalo 39 (49 %) ze 79 akumulátorů výrobce A. 95% Clopperův-Pearsonův intervalový odhad této pravděpodobnosti je (37; 61) %. (Předpoklady pro použití intervalového odhadu byly splněny ($\frac{9}{p \cdot (1-p)} = \frac{9}{0,49 \cdot 0,51} = 36 < 79 (n)$).)

Poznámka:

- Informace o ověřování předpokladu pro použití Clopperova-Pearsonova odhadu se běžně v publikacích neuvádí, nicméně předpokládá se, že výzkumník toto ověření provedl. V domácím úkolu z výukových důvodů tuto informaci uvádějte.

- c) Určete bodový i 95% intervalový odhad relativního rizika poklesu kapacity akumulátoru o více než 10 % pro „horšího“ výrobce (vzhledem k „lepšímu“ výrobcu). Výsledky slovně interpretnete.

U výrobce A lze očekávat cca 4,9 krát větší riziko poklesu kapacit o více než 10 % než u výrobce B. Dle 95% intervalového odhadu lze toto relativní riziko očekávat v rozmezí (2,4; 9,8). Na základě intervalového odhadu lze konstatovat, že na hladině významnosti 0,05 lze vliv výrobce na výskyt poklesu kapacity o více než 10 % považovat za statisticky významný.

Poznámky:

- Relativní riziko je definováno jako poměr rizika (pravděpodobnosti) výskytu sledované události (v tomto případě poklesu kapacity o více než 10 %) u exponované populace (populace s vyšším rizikem, zde výrobce A) a rizika výskytu události u neexponované populace (výrobce B), tj. $\frac{49\%}{10\%} = 4,9$, viz Tab. 7). Toto číslo udává kolikrát je riziko (pravděpodobnost) výskytu události u exponované populace větší než u populace neexponované.
- Pro nalezení 95% intervalového odhadu lze využít vztah uvedený ve Vzorcích a tabulkách, popř. statistický software (v Rku např. funkce `epi.2by2()` z balíčku `epiR`). Při dosazování do vzorce, stejně jako při využití statistického softwaru je **nutné**, aby asociační tabulka byla zadána ve správném tvaru:

	výskyt sledované události	„nevýskyt“ sledované události
exponovaná populace	<i>a</i>	<i>b</i>
neexponovaná populace	<i>c</i>	<i>d</i>

- Neobsahuje-li 95% intervalový odhad relativního rizika číslo 1 (tj. nepřipouštíme-li s 95% spolehlivostí, že rizika výskytu události jsou u exponované i neexponované populace stejné), lze na hladině významnosti 0,05 konstatovat, že výskyt sledované události statisticky významně závisí na typu populace.
- d) Určete bodový i 95% intervalový odhad poměru šancí poklesu kapacity akumulátoru o více než 10 % pro „horšího“ výrobce (vzhledem k „lepšímu“ výrobcu). Výsledky slovně interpretnete.

U akumulátoru výrobce A je šance na pokles kapacity větší než 10 % cca 975:1 000, tj. na 1 975 akumulátorů výrobce A lze očekávat 975 akumulátorů s poklesem kapacity o více než 10 %.

U akumulátoru výrobce D je šance na pokles kapacity větší než 10 % cca 113:1 000, tj. na 1 113 akumulátorů výrobce D lze očekávat 113 akumulátorů s poklesem kapacity o více než 10 %.

Z výše uvedeného je zřejmé, že u výrobce A je šance na pokles kapacity o více než 10 % cca 8,7 krát větší než u výrobce D. Dle 95% intervalového odhadu lze tento poměr šancí očekávat v rozmezí (3,6; 20,4). Na základě intervalového odhadu lze konstatovat, že na hladině významnosti 0,05 lze vliv výrobce na výskyt poklesu kapacity o více než 10 % považovat za statisticky významný.

Poznámky:

- Šance je poměr $\frac{\text{počet výskytu sledované události}}{\text{počet "nevýskytu" sledované události}}$. Šance na výskyt poklesu kapacit o více než 10 % u výrobce A proto byla stanovena jako $\frac{39}{40} = 0,975 = \frac{975}{1\,000} \approx 975:1\,000$. Obdobně jsme určili šanci na výskyt poklesu kapacit o více než 10 % u neexponované populace, tj. u výrobce D.
- Poměr šancí je definován jako poměr šance na výskyt sledované události u exponované populace (populace s vyšším rizikem, zde výrobce A) a šance na výskyt sledované události u neexponované

populace (výrobce B), tj. $\frac{0,975}{0,113} \cong 8,7$. Toto číslo udává kolikrát je šance na výskyt události u exponované populace větší než u populace neexponované.

- Pro nalezení 95% intervalového odhadu lze využít vztah uvedený ve Vzorcích a tabulkách, popř. statistický software (v Rku např. funkce `epi.2by2()` z balíčku `epiR`). Při dosazování do vzorce, stejně jako při využití statistického softwaru je i zde **nutné**, aby asociační tabulka byla zadána ve správném tvaru (viz Poznámky k úloze c).
- Neobsahuje-li 95% intervalový odhad poměru šancí číslo 1 (tj. nepřipouštíme-li s 95% spolehlivostí, že šance na výskyt události jsou u exponované i neexponované populace stejné), lze na hladině významnosti 0,05 konstatovat, že výskyt sledované události statisticky významně závisí na typu populace.

e) Pomocí chí-kvadrát testu nezávislosti rozhodněte, jestli pravděpodobnost poklesu kapacity o více než 10 % závisí statisticky významně na tom, od kterého výrobce akumulátor pochází.

Všechny očekávané četnosti jsou větší než 23, tj. pro ověření uvedené hypotézy lze použít χ^2 test nezávislosti.

Výskyt poklesu kapacit o více než 10 % statisticky významně závisí na výrobci akumulátoru (χ^2 test nezávislosti, $x_{OBS} = 27,3$, $df = 1$, $p\text{-hodnota} < 0,001$).

Poznámky:

- χ^2 test nezávislosti v asociační tabulce
 H_0 : Výskyt události statisticky významně **nesouvisí** s typem populace.
 H_A : Výskyt události statisticky významně **souvisí** s typem populace.
 Tj. je-li $p\text{-hodnota} < \alpha$ (zvolená hladina významnosti), zamítáme předpoklad o nezávislosti výskytu události a typu populace.
 Předpokladem pro použití testu je to, že
 - ✓ všechny **očekávané četnosti** musí být rovny alespoň 2 a
 - ✓ alespoň 80 % **očekávaných četností** musí být větších než 5.
- Informace o ověřování předpokladu pro použití χ^2 testu nezávislosti se běžně v publikacích neuvádí, nicméně předpokládá se, že výzkumník toto ověření provedl. V domácím úkolu z výukových důvodů tuto informaci uvádějte.
- Závěry plynoucí z posouzení relativního rizika, poměru šancí a χ^2 testu nezávislosti musí být stejné. Jde o tři metody sloužící k obdobnému účelu. V praxi si většinou pro posouzení nezávislosti v kontingenční tabulce vyberete, kterou z uvedených metod použijete.

Jak identifikovat, zda jsou v datech odlehlá pozorování?

Emiprické posouzení:

- použití vnitřních (vnějších) hradeb,
- vizuální posouzení krabicového grafu.

Jak naložit s odlehlými hodnotami by měl definovat hlavně zadavatel analýzy (expert na danou problematiku).

Jak ověřit normalitu dat?

Emiprické posouzení:

- vizuální posouzení histogramu,
- vizuální posouzení grafu odhadu hustoty pravděpodobnosti,
- Q-Q graf,
- posouzení výběrové šikmosti a výběrové špičatosti.

Exaktní posouzení:

- testy normality (např. Shapirův – Wilkův test, Andersonův-Darlingův test, Lillieforsův test, ...)

Jak ověřit homoskedasticitu (shodu rozptylů)?

Emiprické posouzení:

- poměr největšího a nejmenšího rozptylu,
- vizuální posouzení krabicového grafu.

Exaktní posouzení:

- *F – test (parametrický dvouvýběrový test),*
- *Bartlettův test (parametrický vícevýběrový test),*
- *Leveneův test (neparametrický test).*