

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:	-----	

Ostrava, AR 2020/2021

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 nekládejte 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. 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 budeme 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 opět upravíme do standardního datového formátu 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 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 *kap100_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 (nezávislá proměnná (příčina, zde výrobce) i závislá proměnná (následek, zde *kap100_w*) je v samostatném sloupci).

- 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, B, C, D. Ze základní tabulky ve standardním datovém formátu si proto vybereme pouze proměnné *pokles_10* a *vyrobce* a můžeme se pustit do vlastní analýzy.

Úkol 1

- a) 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.

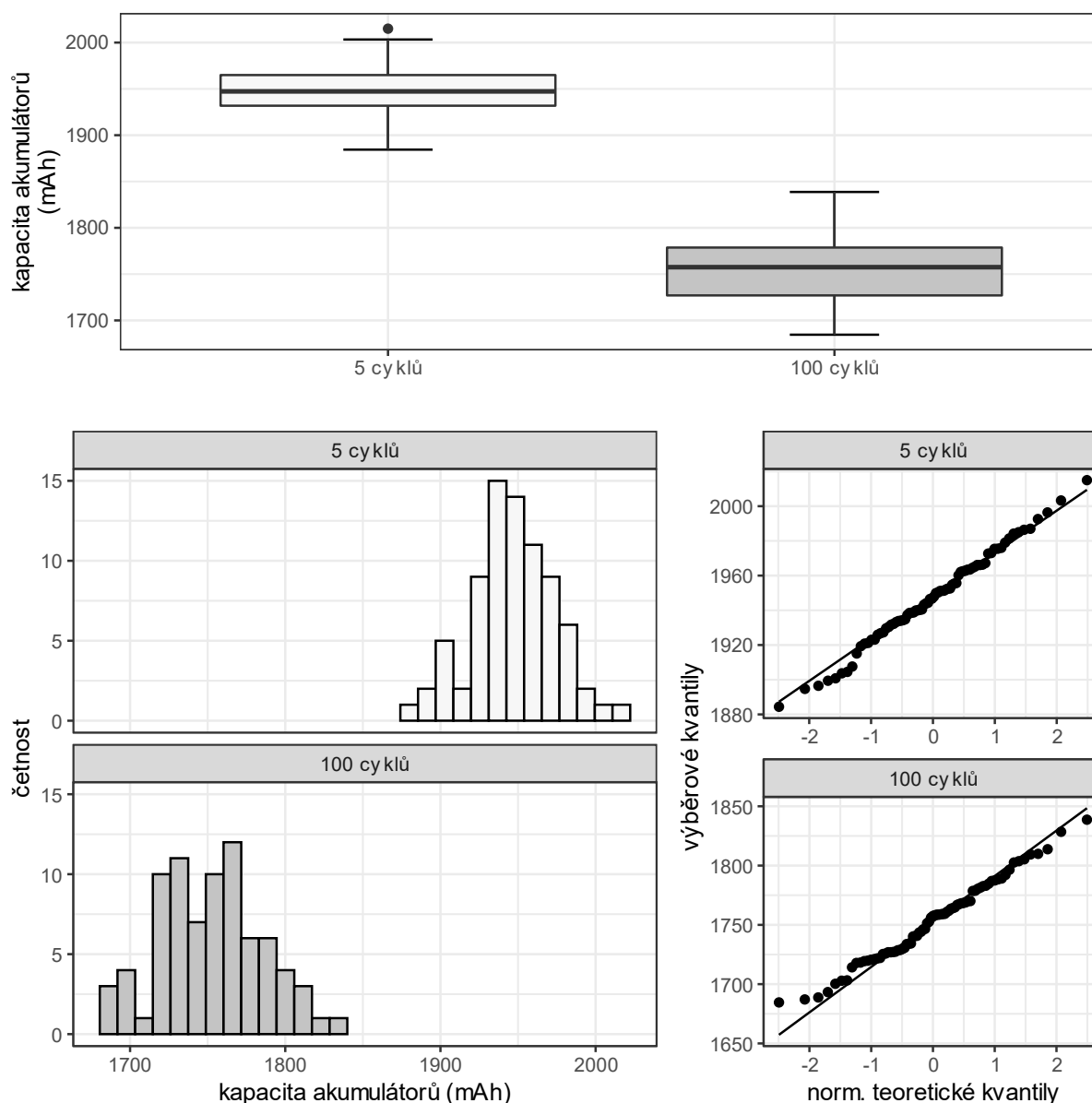
Tab. 1: Kapacita akumulátorů (mAh) výrobce A v závislosti na počtu nabíjecích cyklů (souhrnné statistiky)

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 hradby		
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 (printscreensy, ...)!
- Do tabulky doplňujeme hodnoty zaokrouhlené dle Manuálu pro zaokrouhlování! Viz <http://homel.vsb.cz/~lit40/PRASTA/Materialy/zaokrouhlovani.pdf>
- 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 (data po odstranění odl. pozorování)

Poznámky:

- 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. 1.

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 a nebudou zahrnuty do dalšího zpracování. Možné příčiny vzniku odlehlých pozorování jsou: výrobní vada. 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. 1) 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. 1) 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