

VŠB – Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky

SMAD

Cvičení

JMÉNO STUDENTKY/STUDENTA:

OSOBNÍ ČÍSLO:

JMÉNO CVIČÍCÍ/CVIČÍCÍHO:

Ostrava, AR 2016/2017

Jméno:

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 2000 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.

Jméno:

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.

Kapacita akumulátorů, výrobce A			Po odstranění odlehlých pozorování	
	po 5 cyklech	po 100 cyklech	po 5 cyklech	po 100 cyklech
rozsah souboru				
Míry polohy				
minimum				
dolní kvartil				
medián				
průměr				
horní kvartil				
maximum				
Míry variability				
směrodatná odchylka				
variační koeficient (%)				
Míry šikmosti a špičatosti				
šikmost				
špičatost				

Identifikace odlehlých pozorování – vnitřní hranice		
dolní mez		
horní mez		

Jméno:

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

Jméno:

Analyza kapacity akumulátorů výrobce A po 5 nabíjecích cyklech

Během testu byla měřena kapacita kusů akumulátorů výrobce A. Naměřená kapacita po 5 nabíjecích cyklech se pohybovala v rozmezí až mAh. Kapacity byly identifikovány jako odlehlá pozorování a nebudou zahrnuty do dalšího zpracování. Možné příčiny vzniku odlehlých pozorování jsou: / Žádné z měření nebylo identifikováno jako odlehlé pozorování. Dále uvedené výsledky tedy pocházejí z analýzy kapacit kusů akumulátorů. Jejich průměrná kapacita byla mAh, směrodatná odchylka pak mAh. U poloviny testovaných akumulátorů kapacita nepřekročila mAh. V polovině měření se kapacita pohybovala v rozmezí až mAh. Vzhledem k hodnotě variačního koeficientu (.....%) lze / nelze analyzovaný soubor považovat za homogenní.

Analyza kapacity akumulátorů výrobce A po 100 nabíjecích cyklech

Během testu byla měřena kapacita kusů akumulátorů výrobce A. Naměřená kapacita po 100 nabíjecích cyklech se pohybovala v rozmezí až mAh. Kapacity byly identifikovány jako odlehlá pozorování a nebudou zahrnuty do dalšího zpracování. Možné příčiny vzniku odlehlých pozorování jsou: / Žádné z měření nebylo identifikováno jako odlehlé pozorování. Dále uvedené výsledky tedy pocházejí z analýzy kapacit kusů akumulátorů. Jejich průměrná kapacita byla mAh, směrodatná odchylka pak mAh. U poloviny testovaných akumulátorů kapacita nepřekročila mAh. V polovině měření se kapacita pohybovala v rozmezí až mAh. Vzhledem k hodnotě variačního koeficientu (.....%) lze / nelze analyzovaný soubor považovat za homogenní.

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

Na základě grafického zobrazení (viz) a výběrové šikmosti a špičatosti (výběrová šikmost i špičatost leží / neleží v intervalu $(-2; 2)$) lze / nelze předpokládat, že kapacita akumulátorů výrobce A po 5 nabíjecích cyklech má normální rozdělení. Dle pravidla 3σ / Čebyševovy nerovnosti lze tedy očekávat, že přibližně 95 % / více než 75 % akumulátorů bude mít kapacitu v rozmezí až mAh.

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

Na základě grafického zobrazení (viz) a výběrové šikmosti a špičatosti (výběrová šikmost i špičatost leží / neleží v intervalu $(-2; 2)$) lze / nelze předpokládat, že kapacita akumulátorů výrobce A po 100 nabíjecích cyklech má normální rozdělení. Dle pravidla 3σ / Čebyševovy nerovnosti lze tedy očekávat, že přibližně 95 % / více než 75 % akumulátorů bude mít kapacitu v rozmezí až mAh.

Jméno:

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 (více násobný krabicový graf, histogramy, q-q grafy).

b) Určete bodové a 95% intervalové odhady pro střední hodnoty poklesu kapacit po 100 nabíjecích cyklech pro akumulátory výrobců A i B.

Jméno:

c) Na hladině významnosti 5 % rozhodněte, zda jsou poklesy kapacit akumulátorů výrobců A a B statisticky významné.

d) Určete bodový a 95% intervalový odhad rozdílu středních hodnot poklesů kapacit akumulátorů (po 100 cyklech) výrobců A a B.

e) Na hladině významnosti 5% rozhodněte, zda je rozdíl středních hodnot poklesů kapacit akumulátorů (po 100 cyklech) výrobců A a B statisticky významný.

Jméno:

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 provedte 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).

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

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

Jméno:

- 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.)
- e) Čistým testem významnosti ověřte, zda je pozorovaný rozdíl středních hodnot (resp. mediánů) kapacit akumulátorů po 100 nabíjecích cyklech a všech výrobců 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.)

Jméno:

Úkol 4

U každého z akumulátorů 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, B, C, D. Výsledky prezentujte pomocí kontingenční tabulky, vhodného grafu a vhodné míry kontingence.

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

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

Jméno:

d) Určete bodový i 95% intervalový odhad poměru šancí poklesu kapacity akumulátoru o více než 10% pro „nejhoršího“ výrobce (vzhledem k „nejlepšímu“ výrobcí). Výsledky slovně interpretujte.

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.

Jméno:

Jak identifikovat, zda jsou v datech odlehlá pozorování?

Empirické posouzení:

- použití vnitřních (vnějších) hradeb, resp. z – souřadnice, resp. mediánová suřadnice,
- vizuální posouzení krabicového grafu.

Exaktní posouzení:

- Grubbsův test (parametrický test - vyžaduje normalitu dat)
- Deanův - Dixonův test (neparametrický test)

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

Jak ověřit normalitu dat?

Empirické posouzení:

- vizuální posouzení histogramu,
- vizuální posouzení grafu odhadu hustoty pravděpodobnosti,
- Q-Q graf,
- P-P 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)?

Empirické posouzení:

- poměr největší a nejmenší směrodatné odchylky,
- vizuální posouzení krabicového grafu.

Exaktní posouzení:

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