

PRAVDĚPODOBNOST A STATISTIKA

Domácí úkoly 1S – 4S

Zadání X

JMÉNO STUDENTKY/STUDENTA:

OSOBNÍ ČÍSLO:

JMÉNO CVIČÍCÍ/CVIČÍCÍHO:

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

Ostrava, AR 2020/2021

Popis datového souboru

Běžné zářivky trpí efektem pomalého nabíhání, tedy plného výkonu dosáhnou až po jisté době provozu. Toto chování je ovlivněno okolní teplotou, což v praxi znamená, že v chladném prostředí může zářivkám trvat výrazně déle než dosáhnou maximálního výkonu.

Pro test náběhu zářivek na plný světelný výkon bylo vybráno celkem 350 zářivek od čtyř různých výrobců (Amber, Bright, Clear, Dim). Všechny zářivky měly deklarovaný maximální světelný tok 1000 lm. U každé zářivky byl změřen světelný tok po 30 sekundách od zapnutí, nejprve při teplotě 22 °C a poté při teplotě 5°C.

V souboru [ukol_X.xlsx](#) jsou pro každou z testovaných zářivek uvedeny následující údaje:

- pořadové číslo zářivky,
- výrobce – Amber (A), Bright (B), Clear (C), Dim (D),
- naměřený světelný tok v lumenech při okolní teplotě 5°C,
- naměřený světelný tok v lumenech při okolní teplotě 22°C.

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.

Úkol 1

- Pomocí nástrojů explorační analýzy analyzujte světelný tok zářivek výrobce Amber po 30 sekundách od zapnutí při teplotách 5°C a 22°C. Data vhodně graficky prezentujte (krabicový graf, histogram, q-q graf) a doplňte následující tabulky a text.

Výsledky popisné statistiky lze vidět v Tab. 1 a na

Tab. 1: Světelný tok (lm) zářivek Amber v závislosti na teplotě (souhrnné statistiky)

Světelný tok zářivek Amber (lm)		Po odstranění odlehlých pozorování	
5°C	22°C	5°C	22°C
rozsah souboru			
minimum			
dolní kvartil			
medián			
průměr			
horní kvartil			
maximum			
směrodatná odchylka			
variační koeficient (%)			
šikmost			
špičatost			
Identifikace odlehlých pozorování – vnitřní hradby			
dolní mez			
horní mez			

Jméno:

Číslo zadání: X

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

Analýza světelného toku zářivek výrobce Amber (po 30 sekundách od zapnutí, při teplotě 5°C)

Během testu byl měřen světelný tok kusů zářivek výrobce Amber. Naměřená světelný tok při teplotě 5°C se pohyboval v rozmezí lm až lm. Světelný tok zářivek č. byl na základě metody vnitřních hradeb identifikován jako odlehle pozorování a nebude zahrnut do dalšího zpracování. Možné příčiny vzniku odlehlých pozorování jsou: / Žádné z měření nebylo identifikováno jako odlehle pozorování. Dále uvedené výsledky tedy pocházejí z analýzy světelný toku kusů zářivek. Jejich průměrný světelný tok byl lm, směrodatná odchylka pak lm. U poloviny testovaných zářivek světelný tok nepřekročil lm. V polovině měření se světelný tok pohyboval v rozmezí lm až lm. Vzhledem k hodnotě variačního koeficientu (.....%) lze / nelze analyzovaný soubor považovat za homogenní.

Analýza světelného toku zářivek výrobce Amber (po 30 sekundách od zapnutí, při teplotě 22°C)

Během testu byl měřen světelný tok kusů zářivek výrobce Amber. Naměřená světelný tok při teplotě 22°C se pohyboval v rozmezí lm až lm. Světelný tok zářivek č. byl na základě metody vnitřních hradeb identifikován jako odlehle pozorování a nebude zahrnut do dalšího zpracování. Možné příčiny vzniku odlehlých pozorování jsou: / Žádné z měření nebylo identifikováno jako odlehle pozorování. Dále uvedené výsledky tedy pocházejí z analýzy světelný toku kusů zářivek. Jejich průměrný světelný tok byl lm, směrodatná odchylka pak lm. U poloviny testovaných zářivek světelný tok nepřekročil lm. V polovině měření se světelný tok pohyboval v rozmezí lm až lm. Vzhledem k hodnotě variačního koeficientu (.....%) lze / nelze analyzovaný soubor považovat za homogenní.

Ověření normality světelného toku zářivek výrobce Amber po 30 sekundách od zapnutí při teplotě 5°C 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 světelný tok zářivek výrobce Amber při teplotě 5°C má normální rozdělení. Dle pravidla 3σ / Čebyševovy nerovnosti lze tedy očekávat, že přibližně 95 % / více než 75 % zářivek bude mít světelný tok v rozmezí lm až lm.

Ověření normality světelný toku zářivek výrobce Amber po 30 sekundách od zapnutí při teplotě 22°C 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 světelný tok zářivek výrobce Amber při teplotě 22°C má normální rozdělení. Dle pravidla 3σ / Čebyševovy nerovnosti lze tedy očekávat, že přibližně 95 % / více než 75 % zářivek bude mít světelný tok v rozmezí lm až lm.

Úkol 2

Porovnejte pokles světelného toku po 30 sekundách od zapnutí při snížení okolní teploty z 22°C na 5°C u zářivek od výrobců Amber a Bright. 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 světelného toku zářivek výrobců Amber a Bright při snížení okolní teploty (vícenásobný krabicový graf, histogramy, q-q grafy). Srovnání okomentujte (včetně informace o případné manipulaci s datovým souborem).*

b) *Na hladině významnosti 5 % rozhodněte, zda jsou střední poklesy (popř. mediány poklesů) světelného toku zářivek výrobců Amber a Bright statisticky významné. K řešení využijte bodové a intervalové odhady i testování hypotéz. Výsledky okomentujte.*

c) *Na hladině významnosti 5 % rozhodněte, zda je rozdíl středních hodnot (mediánů) poklesů světelných toků zářivek výrobců Amber a Bright (při snížení okolní teploty) statisticky významný. K řešení využijte bodový a intervalový odhad i čistý test významnosti. Výsledky okomentujte.*

Úkol 3

Na hladině významnosti 5 % rozhodněte, zda se světelný tok zářivek při teplotě 5 °C 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 světelný tok zářivek jednotlivých výrobců statisticky významně liší, určete pořadí výrobců dle středního světelného toku (popř. mediánu světelného toku) zářivek při 5°C.

a) Daný problém vhodným způsobem graficky prezentujte (vícenásobný krabicový graf, histogramy, q-q grafy). Srovnání okomentujte (včetně informace o případné manipulaci s datovým souborem).

b) Ověřte normalitu a symetrii světelného toku zářivek při teplotě 5°C u všech čtyř výrobců (empiricky i exaktně).

c) Ověřte homoskedasticitu (shodu rozptylů) světelného toku zářivek při teplotě 5 °C jednotlivých výrobců (empiricky i exaktně).

- d) *Určete bodové a 95% intervalové odhady střední hodnoty (popř. mediánu) světelného toku zářivek při teplotě 5°C pro všechny srovnávané výrobce. (Nezapomeňte na ověření předpokladů pro použití příslušných intervalových odhadů.)*
- e) *Čistým testem významnosti ověřte, zda je pozorovaný rozdíl středních hodnot (popř. mediánů) světelného toku zářivek při teplotě 5°C statisticky významný na hladině významnosti 5 %. Pokud ano, zjistěte, zda lze některé skupiny výrobců označit (z hlediska světelného toku zářivek po 30 sekundách od zapnutí, při teplotě 5°C) za homogenní, tj. určete pořadí výrobců dle středních hodnot (popř. mediánů) světelného toku zářivek při 5°C. (Nezapomeňte na ověření předpokladů pro použití zvoleného testu.)*

Úkol 4

Všichni čtyři výrobci udávají, že jejich zářivky dosáhnou při 5°C po 30 sekundách od zapnutí alespoň osmdesátiprocent deklarovaného maximálního světelného toku (tj. 80 % z 1 000 lm). Definujte si novou dichotomickou proměnnou Splnění požadavku na deklarovaný světelný tok po 30 s (při 5°C), která bude nabývat hodnot {ANO, NE}. Poznámka: Pracujte s původními daty, nikoliv s daty po odstranění odlehlých pozorování.

a) Srovnajte zářivky jednotlivých výrobců dle toho, zda při teplotě 5°C splňují deklarovaný světelný tok po 30 s od zapnutí pro jednotlivé výrobce (Amber, Bright, Clear, Dim). Výsledky prezentujte pomocí kontingenční tabulky, vhodného grafu a vhodné míry kontingence. Vaše úsudky komentujte.

b) V případě výrobce Bright určete bodový i 95% intervalový odhad pravděpodobnosti, že při teplotě 5°C zářivka nedosáhne po 30 sekundách požadovaného světelného toku (80 % deklarovaného maximálního světelného toku). Nezapomeňte na ověření předpokladů pro použití intervalového odhadu.

- c) *Určete bodový i 95% intervalový odhad relativního rizika, že zářivka při teplotě 5°C nedosáhne po 30 sekundách požadovaného světelného toku (80 % deklarovaného maximálního světelného toku), pro „nejhoršího“ výrobce (vzhledem k „nejlepšímu“ výrobcí). Výsledky slovně interpretujte.*
- d) *Určete bodový i 95% intervalový odhad poměru šancí, že zářivka při teplotě 5°C nedosáhne po 30 sekundách požadovaného světelného toku (80 % deklarovaného maximálního světelného toku), pro „nejhoršího“ výrobce (vzhledem k „nejlepšímu“ výrobcí). Výsledky slovně interpretujte.*
- e) *Pomocí chí-kvadrát testu nezávislosti rozhodněte, jestli to, že zářivka při teplotě 5°C nedosáhne po 30 sekundách požadovaného světelného toku (80 % deklarovaného maximálního světelného toku), závisí statisticky významně na tom, od kterého výrobce zářivka pochází. Výsledky okomentujte.*

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).