VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY

PRAVDĚPODOBNOST A STATISTIKA

Domácí úkoly 1S – 4S

Zadání 85

JMÉNO STUDENTKY/STUDENTA: MARTIN ŠMÍDL

OSOBNÍ ČÍSLO: SMI0116

JMÉNO CVIČÍCÍ/CVIČÍCÍHO: MGR. ADÉLA VRTKOVÁ

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

Ostrava, AR 2022/2023

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 nevkládejte 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

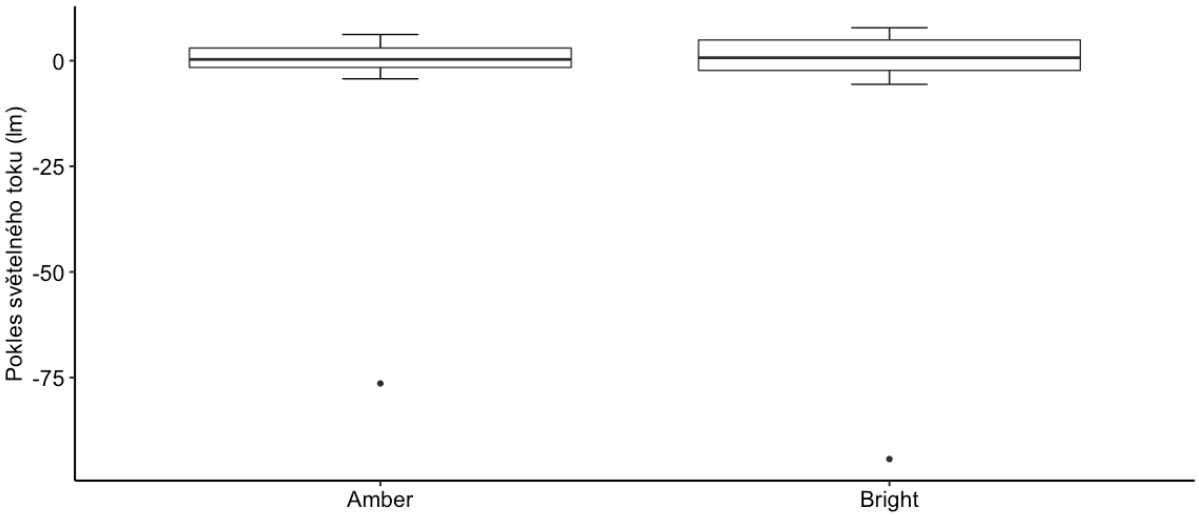
- *Pomocí nástrojů explorační analýzy porovnejte pokles světelného toku po 30 sekundách od zapnutí při snížení okolní teploty z 22°C na 5°C u zářivek od výrobců Amber a Bright. Data vhodně graficky prezentujte (krabicový graf, histogram, q-q graf) a doplňte následující tabulky a text.*

Pro veškeré analýzy byla použita původní datová sada bez úprav, záznamy v datové sadě jsou kompletní.

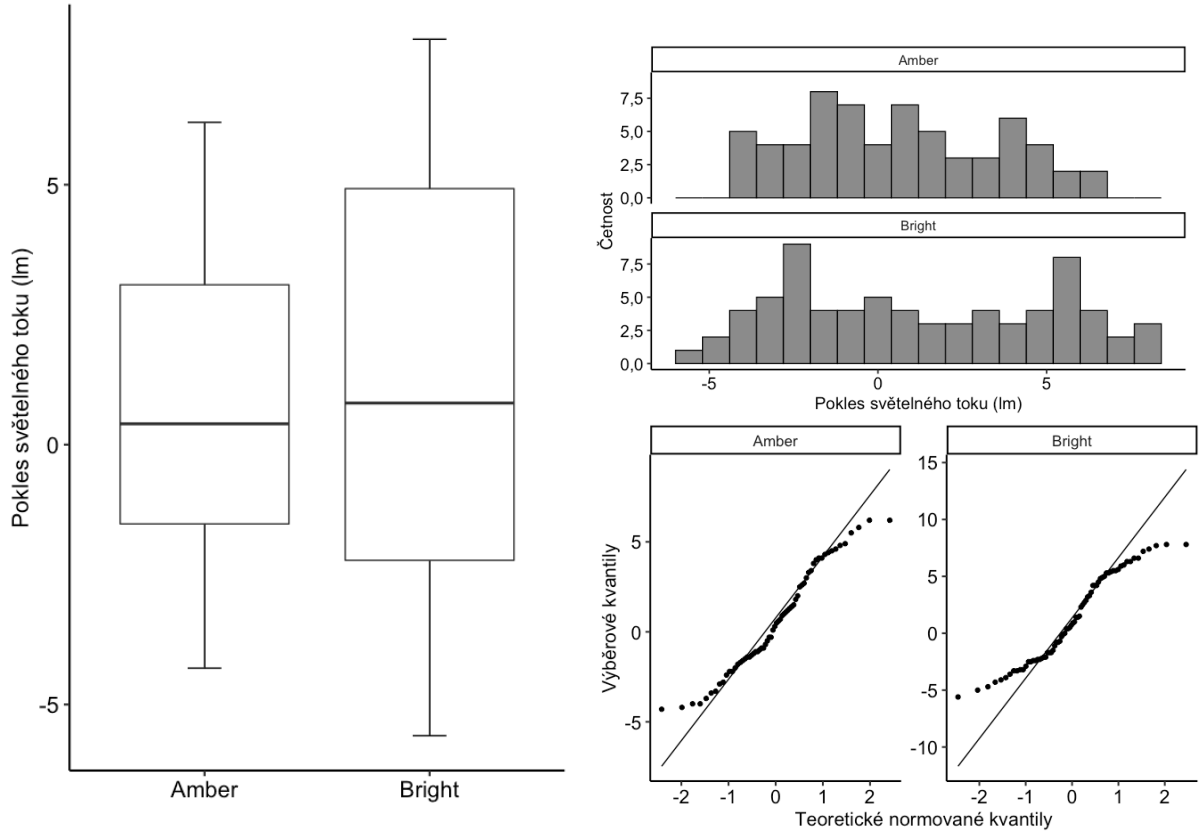
Výsledky popisné statistiky lze vidět v Tab 1 a na Obr 1 a Obr 2.

Tab 1 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 (souhrnné statistiky)

	Původní data		**Data po odstranění odlehlých pozorování**	
	Amber	**Bright**	**Amber**	**Bright**
rozsah souboru	**65**	**73**	64	72
minimum	-76,4	-94,3	-4,3	-5,6
dolní kvartil	-1,60	-2,30	-1,53	-2,23
medián	0,30	0,70	0,40	0,80
průměr	-0,58	-0,05	0,61	1,26
horní kvartil	3,00	4,90	3,08	4,93
maximum	6,2	7,8	6,2	7,8
směrodatná odchylka	9,99	11,83	2,94	3,87
variační koeficient (%)	-	-	-	-
šikmost	-6,9	-7,0	0,2	0,1
špičatost	49,8	53,9	-1,0	-1,3
Identifikace odlehlých pozorování (vnitřní hradby)				
dolní mez	-8,50	-13,10	-	
horní mez	9,90	15,70	-	



Obr 1 Pokles světelného toku (lm) po 30 sekundách od zapnutí při snížení okolní teploty z 22°C na 5°C u zářivek od výrobců Amber a Bright (krabicový graf, původní data)



Obr 2 Pokles světelného toku (lm) po 30 sekundách od zapnutí při snížení okolní teploty z 22°C na 5°C u zářivek od výrobců Amber a Bright (krabicový graf, histogramy, QQ-graf, data po odstranění odlehlých pozorování)

Analýza poklesu světelného toku zářivek výrobce Amber
(po 30 sekundách od zapnutí, při snížení okolní teploty z 22°C na 5°C)

Během testu byl zjišťován pokles světelného toku 65 kusů zářivek výrobce Amber. Zjištěný pokles se pohyboval v rozmezí -76,4 lm až 6,2 lm. Pokles světelného toku zářivky s identifikačním číslem 40 byl na základě metody vnitřních hradeb identifikován jako odlehlé pozorování, protože se nachází mimo interval (-8,50; 9,90) a nebude zahrnut do dalšího zpracování. Možná příčina vzniku odlehlého pozorování je: špatně provedené měření nebo vada zářivky, v závislosti na výši záporného poklesu neboli nárůstu svítivosti. Dále uvedené výsledky tedy pocházejí z analýzy poklesů světelného toku 64 kusů zářivek. Jejich průměrný pokles světelného toku byl 0,61 lm, směrodatná odchylka pak 2,94 lm. U poloviny testovaných zářivek pokles světelného toku nepřekročil 0,40 lm. V polovině případů se pokles světelného toku pohyboval v rozmezí -1,53 lm až 3,08 lm. Vzhledem k záporným hodnotám poklesu neboli nárůstu svítivosti, není vhodné využít variačního koeficientu, tedy nelze analyzovaný soubor považovat za homogenní.

Analýza poklesu světelného toku zářivek výrobce Bright
(po 30 sekundách od zapnutí, při snížení okolní teploty z 22°C na 5°C)

Během testu byl zjišťován pokles světelného toku 73 kusů zářivek výrobce Amber. Zjištěný pokles se pohyboval v rozmezí -94,3 lm až 7,8 lm. Pokles světelného toku zářivky s identifikačním číslem 95 byl na základě metody vnitřních hradeb identifikován jako odlehlé pozorování, protože se nachází mimo interval (-13,10; 15,70) a nebude zahrnut do dalšího zpracování. Možná příčina vzniku odlehlého pozorování je: špatně provedené měření nebo vada zářivky, v závislosti na výši záporného poklesu neboli nárůstu svítivosti. Dále uvedené výsledky tedy pocházejí z analýzy poklesů světelného toku 72 kusů zářivek. Jejich průměrný pokles světelného toku byl 1,26 lm, směrodatná odchylka pak 3,87 lm. U poloviny testovaných zářivek pokles světelného toku nepřekročil 0,80 lm. V polovině případů se pokles světelného toku pohyboval v rozmezí -2,23 lm až 4,93 lm. Vzhledem k záporným hodnotám poklesu neboli nárůstu svítivosti, není vhodné využít variačního koeficientu, tedy nelze analyzovaný soubor považovat za homogenní.

Ověření normality poklesu světelného toku zářivek výrobce Amber
(po 30 sekundách od zapnutí, při snížení okolní teploty z 22°C na 5°C)

Na základě grafického zobrazení (viz Obr 2) a výběrové šikmosti a špičatosti (výběrová šikmost i špičatost leží v intervalu $(-2; 2)$) dle QQ grafu a histogramu se můžeme domnívat, že pokles světelného toku zářivek výrobce Amber má normální rozdělení. Dle pravidla 3σ lze tedy očekávat, že přibližně 95 % zářivek bude mít pokles světelného toku v rozmezí -5,26 lm až 6,47lm.

Ověření normality poklesu světelného toku zářivek výrobce Bright
(po 30 sekundách od zapnutí, při snížení okolní teploty z 22°C na 5°C)

Na základě grafického zobrazení (viz Obr 2) a výběrové šikmosti a špičatosti (výběrová šikmost i špičatost leží v intervalu $(-2; 2)$), ale vzhledem ke QQ grafu a histogramu nelze předpokládat, že pokles světelného toku zářivek výrobce Bright má normální rozdělení. Dle pravidla Čebyševovy nerovnosti lze tedy očekávat, že více než 75 % zářivek bude mít pokles světelného toku v rozmezí -6,47 lm až 8,98 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).* ***Poznámka****: Byla-li grafická prezentace poklesů světelných toků v úkolů 1 bez připomínek, stačí do komentáře vložit odkaz na grafické výstupy z úkolu 1.*

Dle závěrů předešlé analýzy připomeneme, že u výrobce Amber i Bright bylo nalezeno jedno odlehlé pozorování (viz. Tab 1, Obr 1), které jsme se rozhodli odstranit z dalšího zpracování. Dle vizualizace srovnání poklesu svítivosti (viz. Obr 2) se můžeme domnívat, že u obou výrobců dochází k relativně podobným hodnotám poklesů po 30 sekundách od zapnutí při snížení okolní teploty z 22°C na 5°C.

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.*

Dle předchozí analýzy jsme usuzovali, že na základě QQ grafů (viz. Obr 2), šikmosti a stand. špičatosti (viz Tab 1) u výrobce Bright nelze data modelovat normálním rozdělením. U výrobce Amber bylo na základě QQ grafu velice těžké toto rozhodnutí posoudit.

Dle Shapirova-Wilkova nelze na hladině významnosti 0,05 pokles svítivosti zářivek Amber a Bright modelovat normálním rozdělením (viz. Tab 2), proto budou pro popis poklesu svítivosti použity neparametrické metody.

Rozdělení poklesu svítivosti lze u obou výrobců považovat za symetrické (viz Tab 2), proto lze pro intervalové odhady a test významnosti mediánu v případě obou výrobců použít Wilcoxonovu testovou statistiku.

Tab 2 Ověření normality a symetrie poklesů svítivosti u výrobců Amber a Bright

	Šikmost	Stand. špičatost	Shapirův-Wilkův test (p-hodnota)	Test symetrie (p-hodnota)
výrobce Amber	0,2	-1,0	0,041	0,475
výrobce Bright	0,1	-1,3	0,003	0,226

Vzhledem k tomu, že očekáváme kladné poklesy svítivosti (s menší okolní teplotou by měl pokles narůstat) volíme levostranné intervalové odhady / levostranné testy.

95% levostranný intervalový odhad mediánu poklesu svítivosti u výrobce Amber je (-0,10; +∞) lm. Na základě intervalového odhadu a p-hodnoty Wilcoxonova testu, se můžeme domnívat, že pokles svítivosti u výrobce Amber nelze považovat za statisticky významný na hladině významnosti 0,05. (viz. Tab 3)

95% levostranný intervalový odhad mediánu poklesu svítivosti u výrobce Bright je (0,49; +∞) lm. Intervalový odhad, stejně jako p-hodnota Wilcoxonova testu, ukazují, že medián poklesu svítivosti je statisticky významně větší než nula. Tj. na hladině významnosti 0,05 lze pokles svítivosti zářivek výrobce Bright považovat za statisticky významný. (viz. Tab 3)

Tab 3 Odhad mediánu poklesu svítivosti (lm) a test významnosti poklesu svítivosti u výrobců Amber, Bright

	Bodový odhad (lm)	**95% levostranný intervalový odhad (lm)**	**Wilcoxonův levostranný test (p-hodnota)**
výrobce Amber	0,40	(-0,10; +∞)	0,083
výrobce Bright	0,80	(0,49; +∞)	0,006

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.*

Vzhledem k zamítnutí předpokladu normality poklesu svítivosti zářivek výrobců Amber a Bright budeme i nadále pokračovat v aplikaci neparametrických metod.

Dle histogramů (viz Obr 2) lze předpokládat stejný tvar rozdělení. Pro odhad a test významnosti rozdílu mediánů proto použijeme metody založené na Mannově-Whitneyho statistice.

Tab 4 Srovnání mediánů poklesu svítivosti zářivek (lm) výrobce Amber ($x_{0,5}^{Amber}$) a Bright ($x_{0,5}^{Bright}$)

Bodový odhad $x_{0,5}^{Bright} - x_{0,5}^{Amber}$ (lm)	0,40
95% levostranný intervalový odhad $x_{0,5}^{Bright} - x_{0,5}^{Amber}$ (lm)	(-0,50, +∞)
Mannův-Whitneyho levostranný test (p-hodnota)	0,195

U zářivek výrobce Bright lze očekávat medián poklesu svítivosti zářivek o cca 0,40 lm vyšší než u výrobce Amber. Odpovídající 95% levostranný intervalový odhad tohoto rozdílu je (-0,50; +∞) lm. Dle intervalového odhadu a Mannův-Whitneyho testu (viz Tab 4) se můžeme domnívat, že na hladině významnosti 0,05 nelze považovat medián poklesu svítivosti zářivek výrobce Bright za statisticky významně vyšší než u výrobce Amber.

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 osmdesáti procent 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) *Srovnejte 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“ výrobci). 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“ výrobci). 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).*