

# **Domácí úkol 1S-4S**

**Zadání 68**

Jindřich Kvita (kvi0029)

Jméno studentky/studenta:	Jindřich Kvita
Osobní číslo:	kvi0029
Jméno cvičící/cvůčícího:	Tereza Kovářová

	Datum odevzdání	Hodnocení
Domácí úkol 1:		4/4?
Domácí úkol 2:		6,5/7
Domácí úkol 3:		
Domácí úkol 4:		
Celkem:	—	

**Ostrava, AR 2021/2022**

## Popis dtového 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_68.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 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.

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

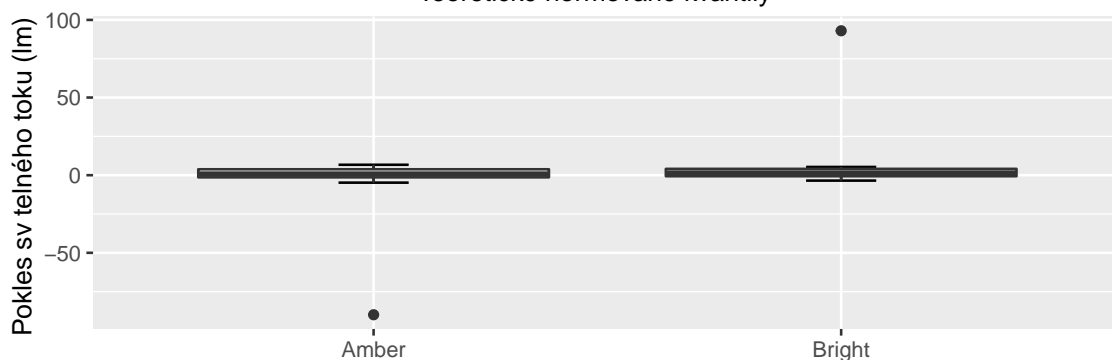
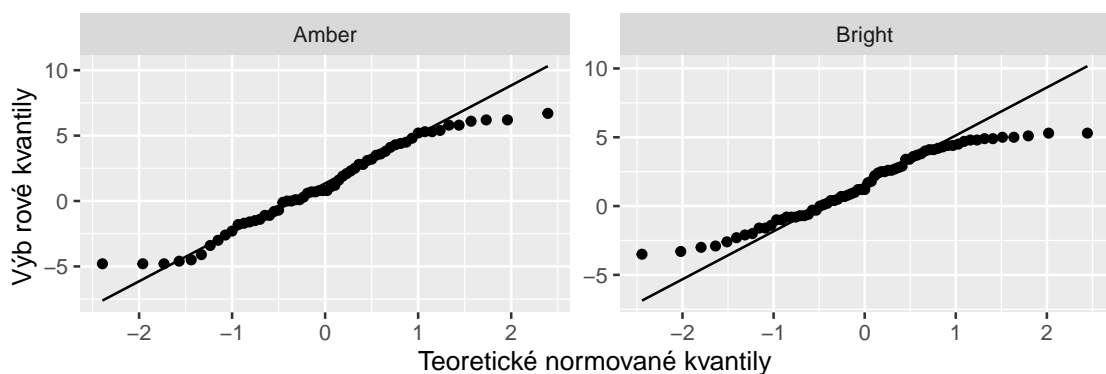
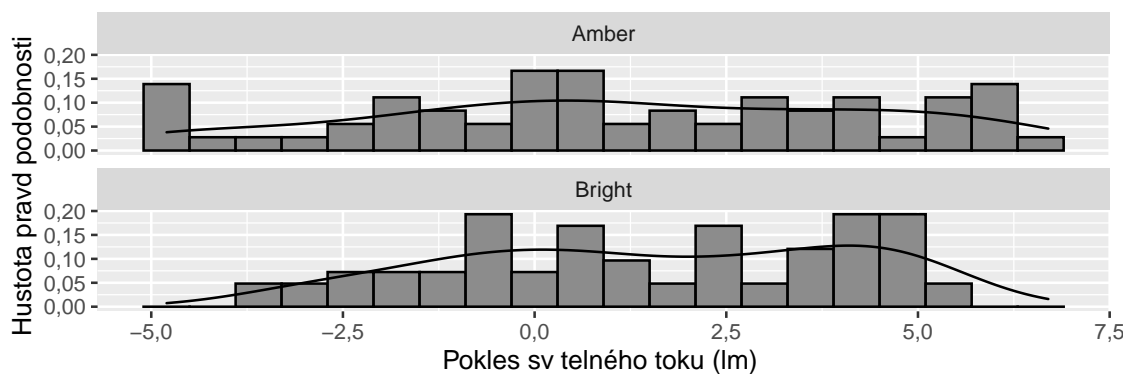
Tab. 1: Pokles světelného toku po 30 sekundách od zapnutí při snížení okolní teploty z 22°C na 5°C u zářivek od výrobců Amber a Bright (souhrnné statistiky)

	Původní data		Data bez odlehlých pozorování	
	Amber	Bright	Amber	Bright
rozsah souboru	61	70	60	69
minimum	-89,9	-3,5	-4,8	-3,5
dolní kvartil	-1,4	-0,7	-1,18	-0,70
medián	0,8	1,4	0,80	1,20
průměr	-0,3	2,8	1,16	1,51
horní kvartil	3,8	4,1	3,87	4,00
maximum	6,7	93	6,7	5,3
směrodatná odchylka	12,1	11,2	3,28	2,57
variační koeficient (%)	3673,6	398,1	282,0	169,6
šikmost	-6,8	7,5	-0,1	-0,2
špičatost	47,8	58,2	-0,9	-1,2
Identifikace odlehlých pozorování (vnitřní hradby)				
dolní mez	-9,20	-7,80		
horní mez	11,60	11,20		

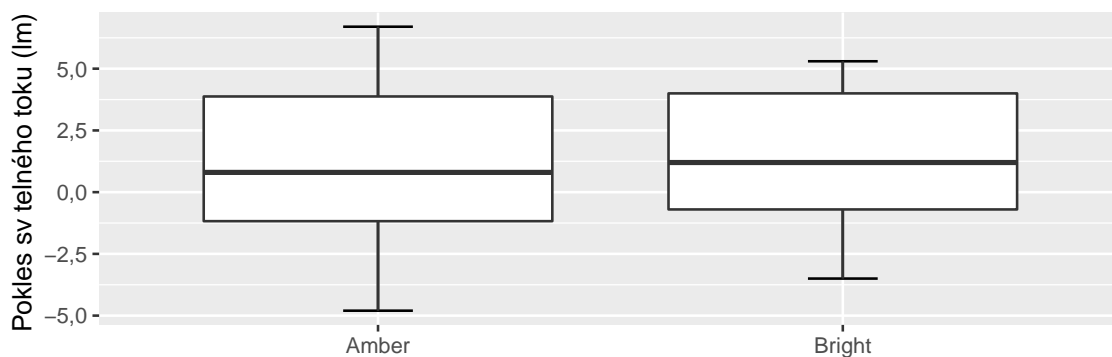
Tabulka 1: Pokles světelného toku po 30 sekundách od zapnutí při snížení okolní teploty z 22°C na 5°C u zářivek od výrobců Amber a Bright (souhrnné statistiky)



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



(data z p vodního datasetu)



(data bez odlehlých pozorování)

**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 70 kusů zářivek výrobce Bright. Zjištěný pokles se pohyboval v rozmezí -3,5 lm až 93 lm. Pokles světelného toku zářivek č. byl na základě metody vnitřních hradeb identifikován jako odlehlé pozorování a nebude zahrnut do dalšího zpracování. Možné příčiny vzniku odlehlých pozorování jsou: chyba měření, výjimečná vada zářivky, chyba zpracování dat. Dále uvedené výsledky tedy pocházejí z analýzy poklesů světelného toku 69 kusů zářivek. Jejich průměrný pokles světelného toku byl 1,51 lm, směrodatná odchylka pak 2,57 lm. U poloviny testovaných zářivek pokles světelného toku nepřekročil 1,20 lm. V polovině případů se pokles světelného toku pohyboval v rozmezí -0,70 lm až 4,00 lm. Vzhledem k příliš vysoké hodnotě variačního koeficientu (169,6 %) není variační koeficient, v tomto případě, vhodný ukazatel homogenity souboru.

**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 Obrázek 1) a výběrové šikmosti a špičatosti (výběrová šikmost i špičatost leží v intervalu ) lze předpokládat, že pokles světelného toku zářivek výrobce Amber má normální rozdělení. Dle pČebyševovy nerovnosti lze tedy očekávat, že přibližně více než 75 % zářivek bude mít pokles světelného toku v rozmezí -5,40 lm až 7,73 lm.

**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 Obrázek 1) a výběrové šikmosti a špičatosti (výběrová šikmost i špičatost leží v intervalu ) nelze předpokládat, že pokles světelného toku zářivek výrobce Bright má normální rozdělení. Dle Čebyševovy nerovnosti lze tedy očekávat, že přibližně Více než 75 % zářivek bude mít pokles světelného toku v rozmezí -3,62 lm až 6,64 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 úkolu 1 bez připomínek, stačí do komentáře vložit odkaz na grafické výstupy z úkolu 1.

Při porovnání dat po odstranění odlehlých pozorování (v obou případech se jednalo o odstranění jedné hodnoty) se jeví data u výrobce Amber, že mají normální rozdělení, avšak data u výrobce Bright, dle histogramu u (obrázek 1) a vychýlení na qq grafu (obrázek 2), že se nebude jednat o normální rozdělení. Dle krabicového grafu lze usuzovat, že rozptyl u výrobce Bright bude menší, tudíž budou konzistentnější. Takto bylo usouzeno dle faktu, že je jak krabice, tak fousy u výrobce Bright menší)

- 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.

	šikmost	špičatost	Shapiro-Wilkův test (p-hodnota)	test symetrie (p-hodnota)
Amber	-0,1	-0,9	0,06	0,274
Bright	-0,2	-1,2	0,005	0,217

Tabulka 1: Ověření normality a symetrie poklesu světelného toku zářivek dle výrobce

Dle Shapiro-Wilkova testu, u výrobce Amber, na hladině významnosti 5% normalitu nezamítáme. U výrobce Bright ji na stejné hladině významnosti zamítáme. Z tohoto důvodu (a aby bylo možné dále výrobce mezi sebou porovnávat), bude u obou výrobců využito neparametrických testů.



Dále byl použit levostranný intervalový odhad a test významnosti mediánů pro oba výrobce. Test významnosti byl proveden pomocí Wilcoxonova testu.

	bodový odhad (lm)	95% levostranný IO (lm)	Wilcoxonův levostranný test (p-hodnota)
Amber	0,80	$(0,45; +\infty)$	0,006
Bright	1,20	$(0,99; +\infty)$	< 0,001

Tabulka 2: Odhad mediánu poklesu světelného toku (lm) dle výrobce a test významnosti poklesu toku

U výrobce Amber je pokles svitu u poloviny zářivek větší, než 0,80 lm. 95% levostranný intervalový odhad mediánu poklesu je  $(0,45; +\infty)$  lm. Na hladině významnosti 0,05 jsou poklesy svitu statisticky významné, protože se bodový odhad nachází v intervalu  $(0,45; +\infty)$ . Obdobná situace je i u výrobce Bright, jen s rozdílem, že polovina poklesů svitu je větší, než 1,20 lm a 95% levostranný intervalový odhad mediánu poklesu je  $(0,99; +\infty)$

- 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.

Tvar rozdělení poklesů toků u Amber je normální avšak u Bright ne. Kvůli tomu je nutno použít neparametrické testy pro oba výrobce, budou tedy použity metody založené na Mannově-Whitneyho statistice.

	bodový odhad (lm)	95% oboustranný IO (lm)	Mannův-Whitneyho oboustranný test (p-hodnota)
Amber - Bright	-0,40	$(-1,50; 0,80)$	0,628

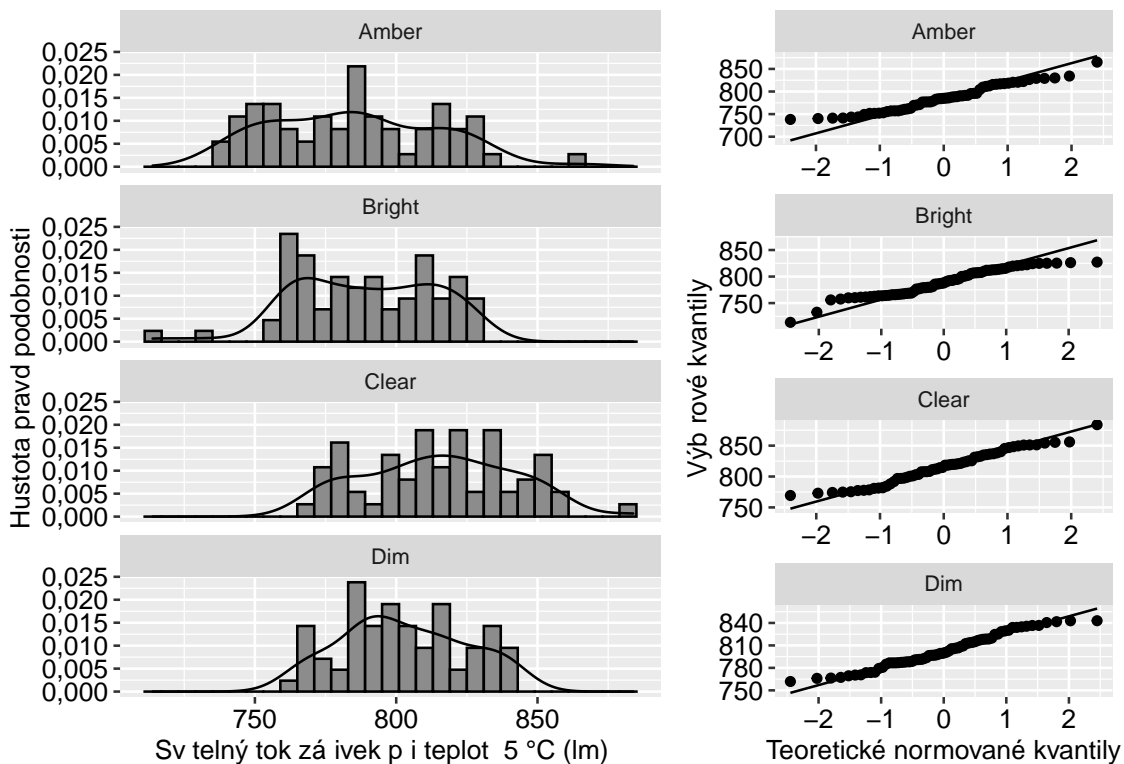
Tabulka 3: Odhad rozdílu mediánu poklesu světelného toku (lm) dle výrobce a test významnosti rozdílu mediánu

Dle bodového odhadu mediánu poklesu světelného toku očekávám pokles toku u zářivek výrobce Amber o cca -0,40 lm než u zářivek Bright. Tomu odpovídá i 95% oboustranný intervalový odhad tohoto rozdílu, dle kterého interval  $(-1,50; 0,80)$  pokrývá skutečný rozdíl mediánů s 95% pravděpodobností.

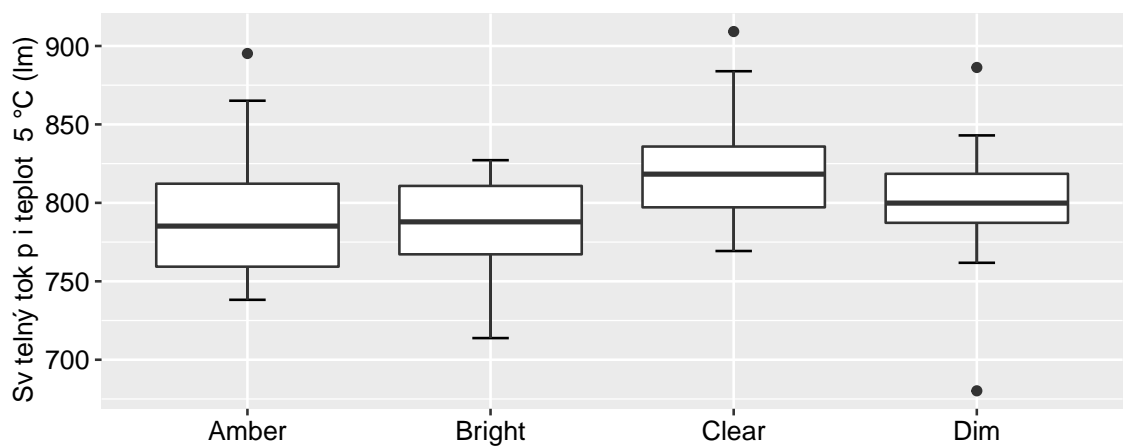
## Úkol 3

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.

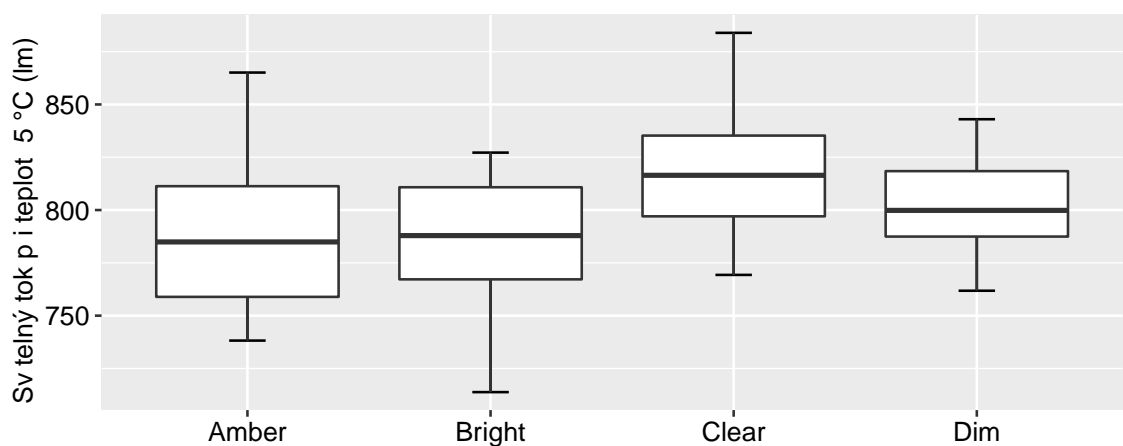
- a) Daný problém vhodným způsobem graficky prezentujte (vícenásobný krabicový graf, histogramy, q-q grafy). Srovnání okomentujte (včetně informace o případné manipulaci s datovým souborem).



Tyto grafy pochází z původního datasetu, tj. zde se nachází i odlehlá pozorování všech výrobců. Dle Vizuálního zhodnocení histogramů se zdá, že Bright a Clear nebudou mít normální rozdělení.



(data z p vodního datasetu)



(data bez odlehlých pozorování)

Obrázek 1: Krabicové grafy světelného toku zářivek (lm) při teplotě 5 °C dle výrobce

Tabulka 1: Ověření normality a symetrie světelného toku zářivek (lm) při teplotě 5 °C u všech výrobců

	šikmost	špičatost	Shapirův- Wilkův test (p-hodnota)	Test symetrie (p-hodnota)
<b>Amber</b>	0,3	-0,6	0,065	0,902
<b>Bright</b>	-0,3	-0,3	0,007	0,704
<b>Clear</b>	0,1	-0,7	0,163	0,640
<b>Dim</b>	0,1	-0,9	0,067	0,198

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

Dle Shapiro-Wilkova testu, na hladině významnosti 5% zamítáme nulovou hypotézu o normálním rozdělení u výrobce Bright. Dle šikmosti a špičatosti lze určit, že se však jedná o symetrické rozdělení.

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

Dle krabicových grafů a poměru největšího a nejmenšího rozptylu světelného toku (1,8\$) nezamítáme shodu rozptylů. Na základě Levenova testu, kde p-hodnota = 0,160 nelamítáme předpoklad shodě rozptylů

- 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) 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).