

Pravděpodobnost a statistika
Domácí úkoly 1S-4S
Zadání 123

Martin Pustka

7. května 2021

Jméno studenta: Martin Pustka
Osobní číslo: PUS0065
Jméno cvičícího: Ing. Michal Bérés

	Datum odevzdání	Hodnocení
Domácí úkol 1:	9. duben 2021	5/5
Domácí úkol 2:	30. duben 2021	2.5/5
Domácí úkol 3:	7. května 2021	
Domácí úkol 4:		
Celkem:		

Obsah

1	Úkol 1S	4
1.a	Zadání	4
1.b	Tabulkové řešení	4
1.c	Grafické řešení	5
1.d	Textové řešení	8
2	Úkol 2S	9
2.a	9
2.b	11
2.c	12
3	Úkol 3S	13
3.a	13
3.b	15
3.c	16
3.d	17
3.e	18

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.

1 Úkol 1S

1.a Zadání

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.

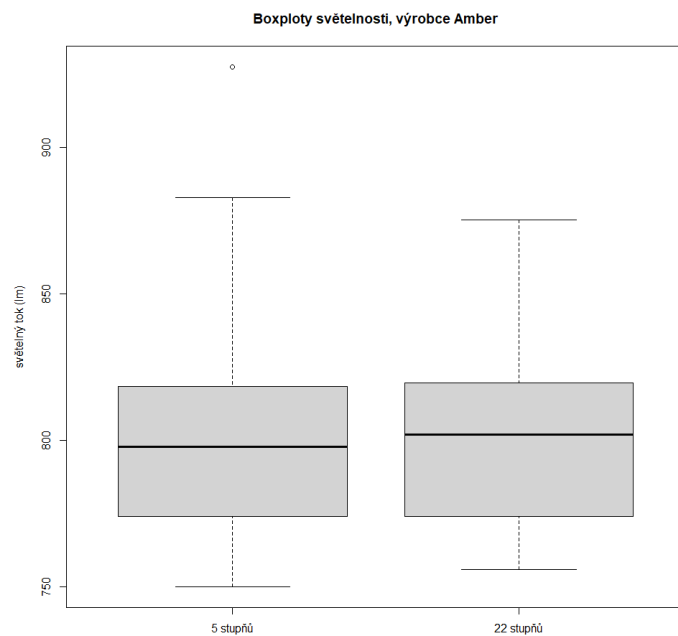
Výsledky popisné statistiky lze vidět v tabulce 1 a na obrazcích 1,2,3,4 a 5. Boxplot obsahuje odlehlá pozorování, ostatní grafy jsou bez odlehlých pozorování.

1.b Tabulkové řešení

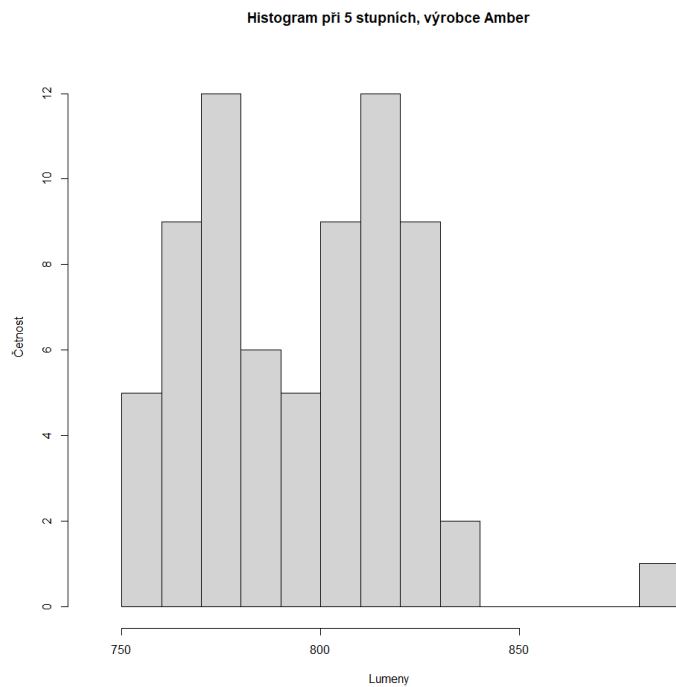
Světelný tok zářivek Amber (lm)			Po odstranění odlehlých pozorování	
	5°C	22°C	5°C	22°C
rozsah souboru	71	71	70	71
minimum	750,0	755,8	750,0	755,8
dolní kvartil	774,0	774,0	773,4	774,0
medián	797,8	801,9	797,4	801,9
průměr	797,5	797,6	795,6	797,6
horní kvartil	818,3	819,6	817,9	819,6
maximum	927,4	875,1	882,9	875,1
směrodatná odchylka	30,5	25,8	26,4	25,8
variační koeficient (%)	3,8	3,2	3,3	3,2
šikmost	1,2	0,1	0,3	0,1
špičatost	3,4	-0,5	0,0	-0,5
Identifikace odlehlých pozorování – vnitřní hradby				
dolní mez	707,6	705,6	706,6	705,6
horní mez	884,7	888,0	884,6	888,0

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

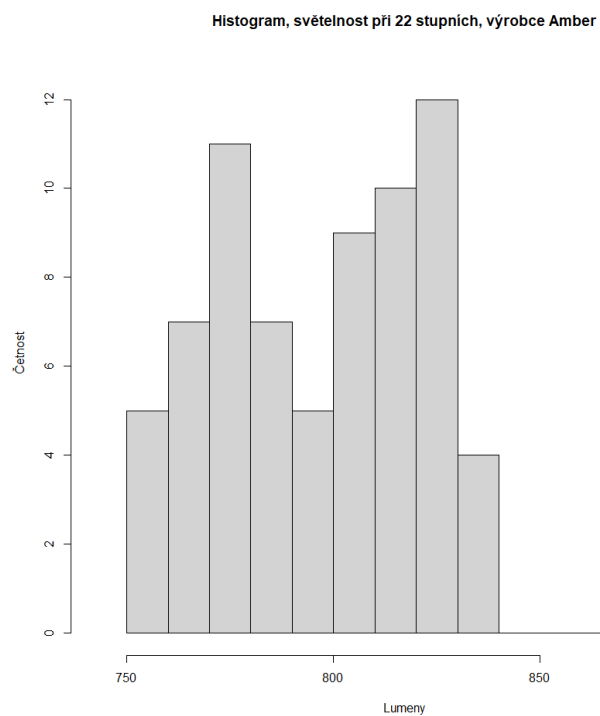
1.c Grafické řešení



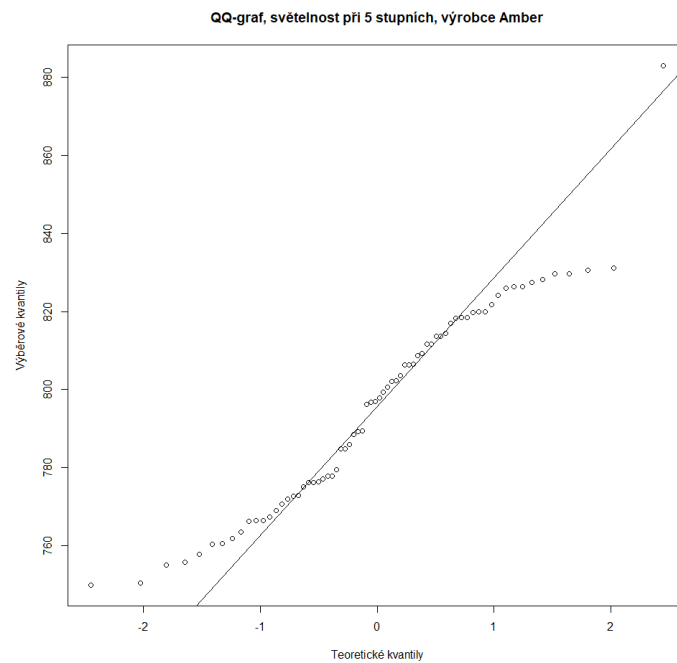
Obrázek 1: Boxploty světelnosti, Amber



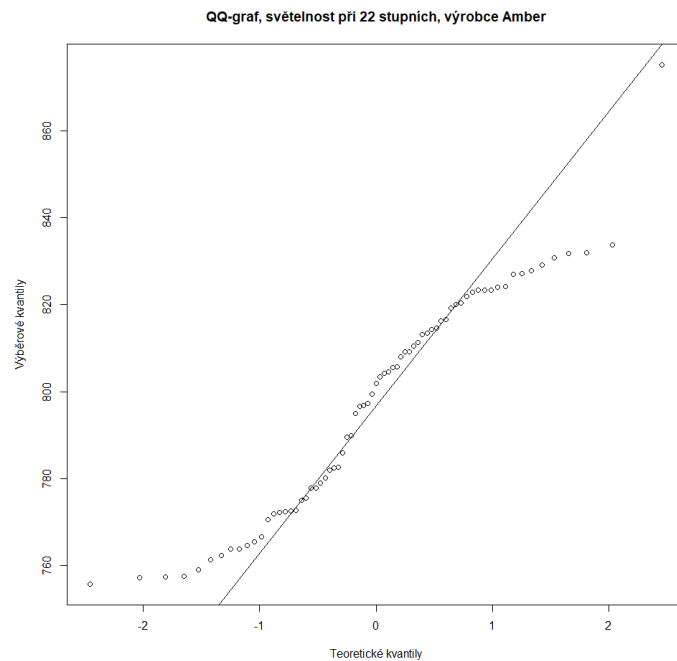
Obrázek 2: Histogram světelnosti při 5 stupních, výrobce Amber



Obrázek 3: Histogram světelnosti při 22 stupních, výrobce Amber



Obrázek 4: QQ graf světelnosti při 5 stupních, výrobce Amber



Obrázek 5: QQ graf světelnosti při 5 stupních, výrobce Amber

1.d Textové řešení

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

Během testu byl měřen světelný tok 71 kusů zářivek výrobce Amber. Naměřená světelný tok při teplotě 5°C se pohyboval v rozmezí 750,0 lm až 927,4 lm. Světelný tok zářivek č. 27 byl na základě metody vnitřních hradeb identifikován jako odlehlé pozorování a nebude zahrnut do dalšího zpracování. Možné příčiny vzniku odlehlých pozorování jsou: *chyby v měření nebo při výrobě*. Dále uvedené výsledky tedy pocházejí z analýzy světelný toku 70 kusů zářivek. Jejich průměrný světelný tok byl 795,6 lm, směrodatná odchylka pak 26,4 lm. U poloviny testovaných zářivek světelný tok nepřekročil 797,4 lm. V polovině měření se světelný tok pohyboval v rozmezí 773,4 lm až 817,9 lm. Vzhledem k hodnotě variačního koeficientu (3,3 %) lze 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 71 kusů zářivek výrobce Amber. Naměřená světelný tok při teplotě 22°C se pohyboval v rozmezí 755,8 lm až 875,1 lm. Žádné z měření nebylo identifikováno jako odlehlé pozorování. Dále uvedené výsledky tedy pocházejí z analýzy světelný toku 71 kusů zářivek. Jejich průměrný světelný tok byl 797,6 lm, směrodatná odchylka pak 25,8 lm. U poloviny testovaných zářivek světelný tok nepřekročil 801,9 lm. V polovině měření se světelný tok pohyboval v rozmezí 774,0 lm až 819,6 lm. Vzhledem k hodnotě variačního koeficientu (3,2 %) lze 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 obrázek 4) a výběrové šikmosti a špičatosti (výběrová šikmost i špičatost leží v intervalu (-2;2)) lze předpokládat, že světelný tok zářivek výrobce Amber při teplotě 5°C má normální rozdělení. Dle pravidla 3σ lze tedy očekávat, že přibližně 95 % zářivek bude mít světelný tok v rozmezí 742,9 lm až 848,3 lm.

Ověření normality světelného toku zářivek výrobce Amber po 30 sekundách od zapnutí při teplotě 22°C na základě explorační analýzy

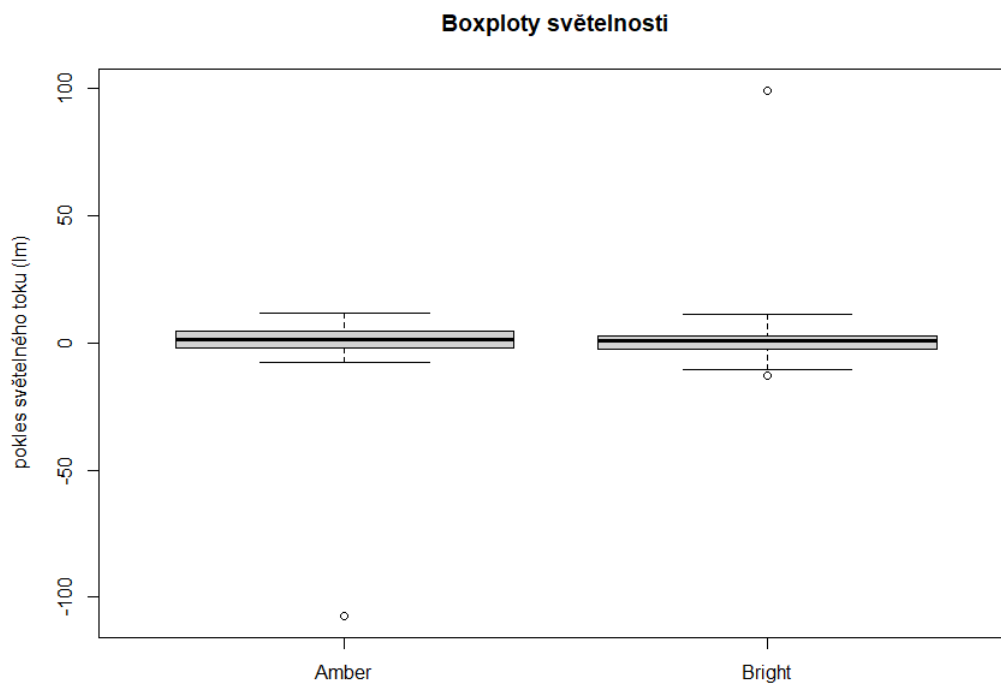
Na základě grafického zobrazení (viz obrázek 5) a výběrové šikmosti a špičatosti (výběrová šikmost i špičatost leží v intervalu (-2;2)) lze předpokládat, že světelný tok zářivek výrobce Amber při teplotě 22°C má normální rozdělení. Dle pravidla 3σ lze tedy očekávat, že přibližně 95 % zářivek bude mít světelný tok v rozmezí 746,0 lm až 849,2 lm.

2 Úkol 2S

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.

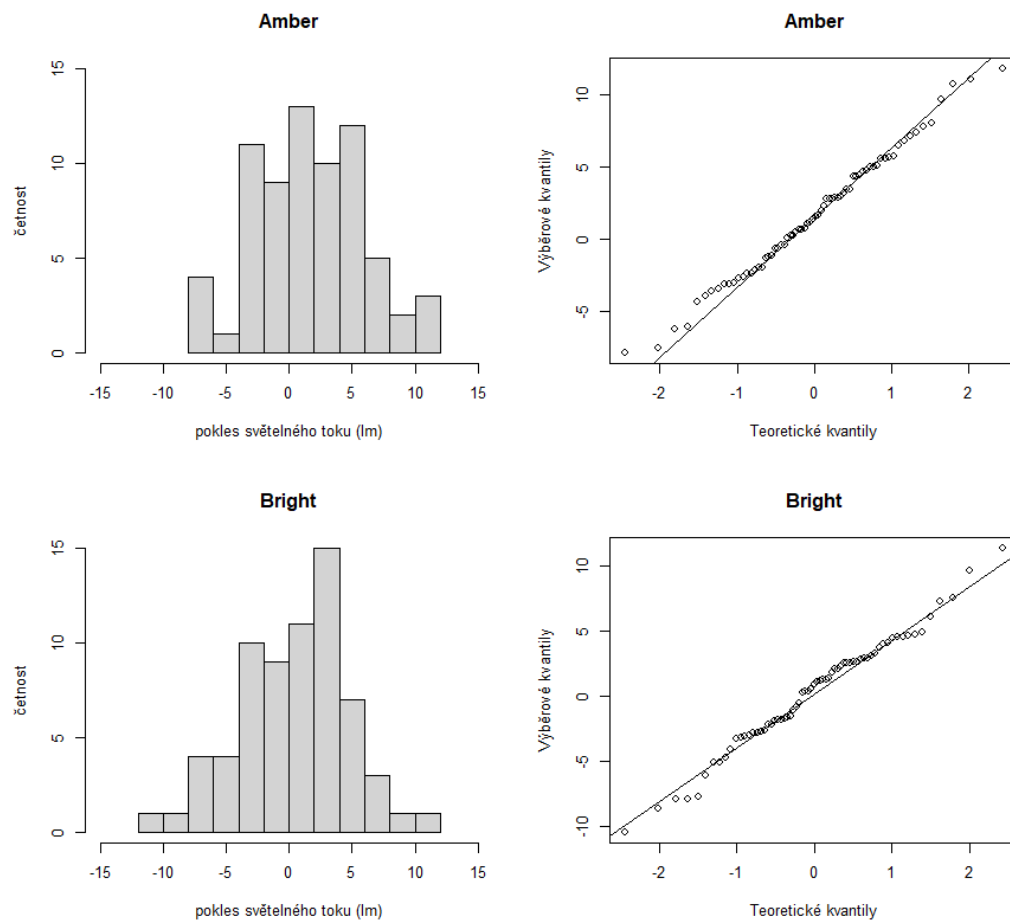
2.a

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).



Obrázek 6: Boxploty světelnosti, Amber a Bright

U obou výrobců byly pozorovány 3 odlehlé pozorování (viz Obr. 6), tyto jsme se rozhodli z dalšího zpracování vypustit. Podle vizualizace srovnání poklesů světelného toku zářivek výrobců Amber a Bright (viz Obr. 6), Obr. 7)) lze usoudit, že k výraznému poklesu světelného toku při snížení okolní teploty nedošlo.



Obrázek 7: Histogramy a Q-Q grafy světelnosti, Amber a Bright, odstraněná pozorování

2.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 prezentovaných Q-Q grafů (viz Obr. 7) lze usuzovat, že poklesy světelného toku zářivek u výrobce Amber i výrobce Bright lze modelovat normálním rozdělením. Šikmost i špičatost (viz Tab. 2) odpovídají normálnímu rozdělení.

	šikmost	stand. špičatost	Shapiro-Wilkův test (p-hodnota)	Test symetrie (p-hodnota)
výrobce Amber	0,1	-0,4	0,872	0,681
výrobce Bright	-0,1	0,1	0,646	0,110

Tabulka 2: Ověření normality poklesu světelného toku zářivek

Dle Shapiro-Wilkova testu lze na hladině významnosti 0,05 pokles světelného toku zářivek u výrobce Amber i u výrobce Bright modelovat normálním rozdělením (viz Tab. 2).

Rozdělení poklesů světelných toků zářivek lze u výrobce Amber i u výrobce Bright považovat za symetrické. (viz šikmosti, výsledky testu symetrie (Tab. 2) a tvar histogramů (Obr. 7)), proto lze pro intervalové odhady a test významnosti stř. hodnot v případě obou výrobců použít Studentův t-test.

Vzhledem k tomu, že očekáváme poklesy světelnosti (s klesající teplotou se světelný tok zářivek snižuje), volíme jednostranné intervalové odhady / jednostranné testy.

U zářivek výrobce Amber lze očekávat, že polovina bude vykazovat pokles světelného toku menší než cca 1.67 lm. 95% jednostranný intervalový odhad stř. hodnot poklesu světelného toku zářivek výrobce Amber je $(0,78; \infty)$ lm. Intervalový odhad, stejně jako p-hodnota, ukazují, že stř. hodnota poklesu světelnosti je statisticky významně větší než nula. Tj. na hladině významnosti 0,05 lze pokles světelného toku žárovek výrobce Amber považovat za statisticky významný (viz Tab. 3).

U zářivek výrobce Bright lze očekávat, že polovina bude vykazovat pokles světelného toku menší než cca 0.27 lm. 95% jednostranný intervalový odhad stř. hodnot poklesu světelného toku zářivek výrobce Bright je $(-0,62; \infty)$ lm. Intervalový odhad, stejně jako p-hodnota, ukazují, že stř. hodnota poklesu světelnosti je statisticky významně větší než nula. Tj. na hladině významnosti 0,05 nelze pokles světelného toku žárovek výrobce Bright považovat za statisticky významný (viz Tab. 3).

	bodový odhad (lm)	95% jednostranný intervalový odhad (lm)	Studentův jednostranný t-test (p-hodnota)
výrobce Amber	1,67	$(0,78; \infty)$	0,001
výrobce Bright	0,27	$(-0,62; \infty)$	0,308

Tabulka 3: Odhad stř. hodnoty a test významnosti poklesu světelných toků zářivek (lm) dle výrobce

2.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 potvrzení normality u výrobce Amber i Bright je třeba zjistit zda mezi rozptyly výrobců panuje shoda.

Na základě výsledku F-testu (viz Tab. 4) lze rozhodnout o shodě rozptylů (homoskedasticitě) a použít Dvouvýběrový Studentův t-test pro test stř. hodnot.

U zářivek výrobce Amber lze očekávat pokles ve světelném toku o cca 1,4 lm větší než u výrobce Bright. Odpovídající 95% jednostranný intervalový odhad tohoto rozdílu je (0,63;1,68) lm. Intervalový odhad, stejně jako Dvouvýběrový Studentův t-test (viz Tab. 4) ukazují, že stř. hodnota poklesu světelného toku zářivek výrobce Amber není statisticky významně vyšší než u výrobce Bright.

Bodový odhad $\mu^A - \mu^B$ (lm)	1,40
F-test (p-hodnota)	0,889
95% oboustranný intervalový odhad $\mu^A - \mu^B$ (lm)	(-0,08;2,88)
Dvouvýběrový Studentův t-test (p-hodnota)	1,000

Tabulka 4: Srovnání středních hodnot poklesu světelných toků zářivek (lm) výrobce Amber a Bright

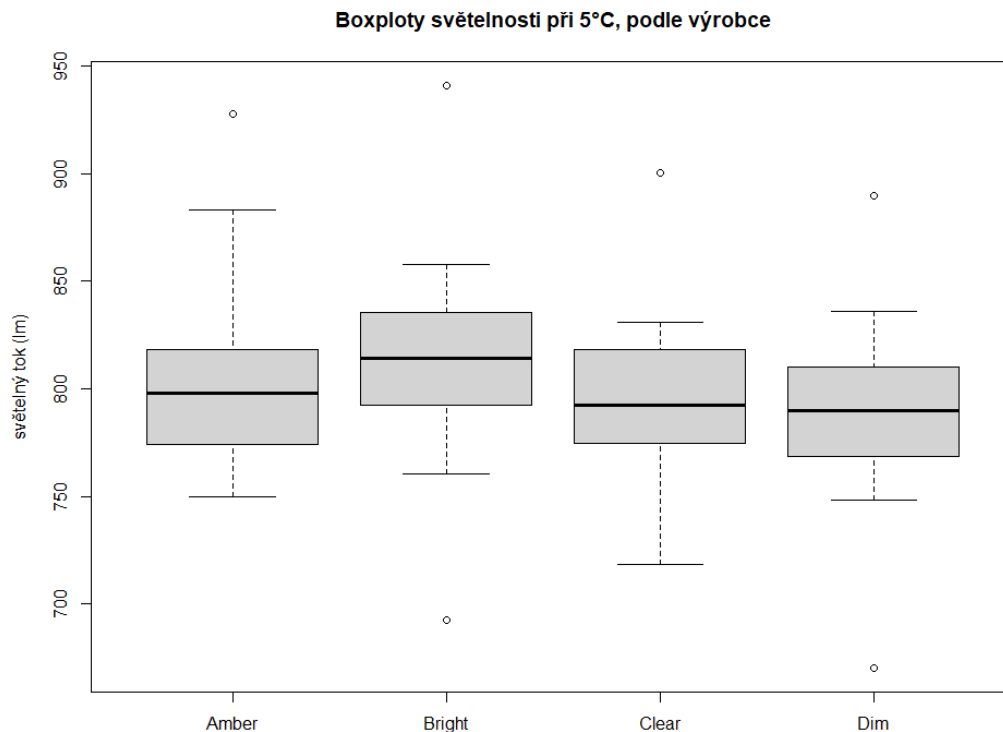
3 Úkol 3S

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.

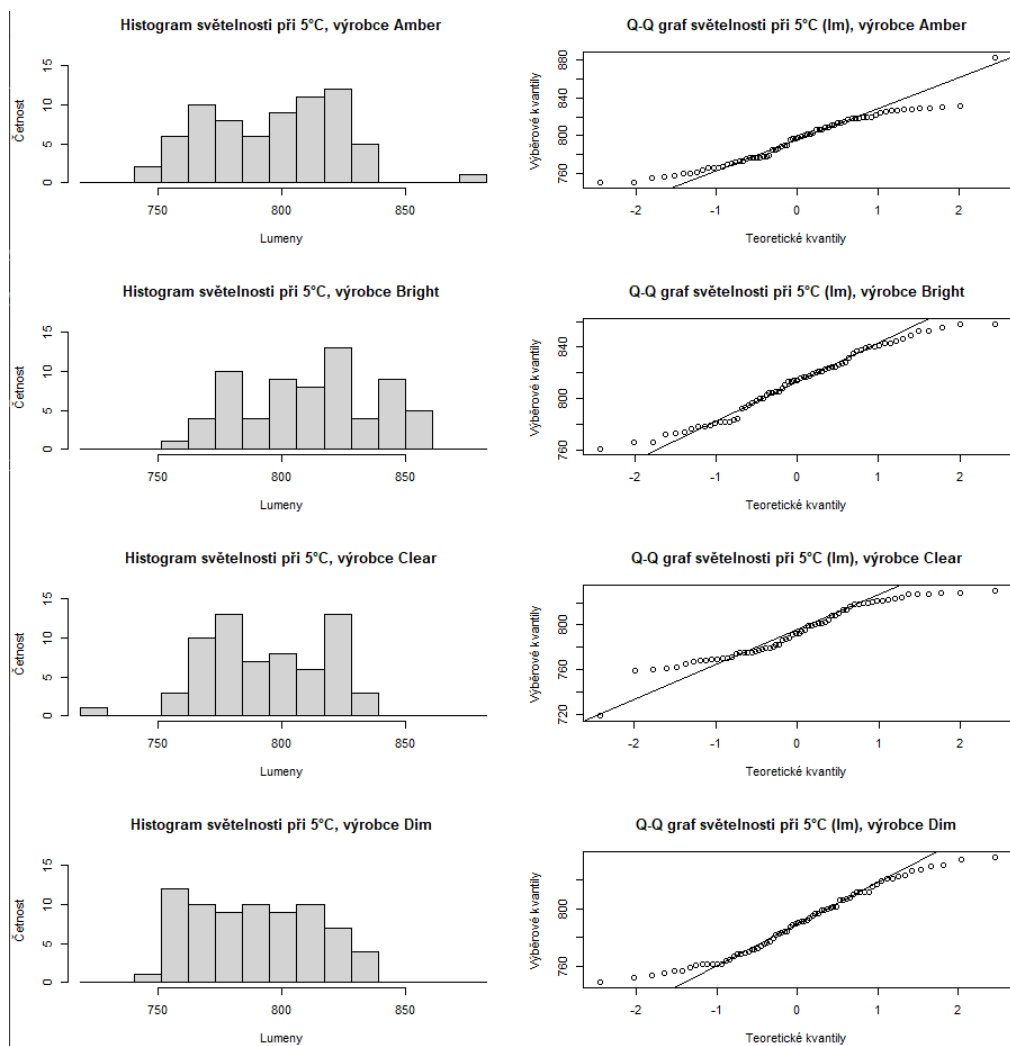
3.a

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

U všech výrobců byla pozorována odlehlá pozorování, celkem 6 (viz Obr. 8), která byla z dalšího zpracování odstraněna. Z histogramů a Q-Q grafů (viz Obr. 9) nelze, mezi výrobci, pozorovat významné rozdíly mezi světelnostmi jejich zářivek.



Obrázek 8: Boxploty světelnosti při 5°C. Výrobce Amber, Bright, Clear a Dim



Obrázek 9: Histogramy a Q-Q grafy světelnosti při 5°C. Výrobci Amber, Bright, Clear a Dim, odstraněná pozorování

3.b

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

Podle prezentovaných grafů (viz Obr. 9) nelze usuzovat, že světelný tok při 5 stupních je normálním rozdělením.

Šikmost i špičatost (viz Tab. 5) normálnímu rozdělení odpovídá, ale Shapirův-Wilkův test normalitu zamítl u všech výrobců s výjimkou Bright.

Podle tvaru histogramů (viz Obr. 9), výsledků šikmostí a výsledků testu symetrie (viz Tab. 5) lze u všech výrobců světelnost zářivek při 5°C považovat za symetrickou.

Výrobce	Šikmost	Špičatost	Shapirův-Wilkův test (p-hodnota)	Test symetrie (p-hodnota)
Amber	0,3	0,0	0,009	0,477
Bright	-0,1	-1,0	0,079	0,448
Clear	-0,3	-0,3	0,005	0,638
Dim	0,1	-1,1	0,031	0,823

Tabulka 5: Srovnání normality a symetrie poklesu světelného toku zářivek, podle výrobce

3.c

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

Podle Obr. 9 a poměru největšího a nejmenšího rozptylu světelností (viz Tab. 6), $S_A^2/S_D^2 = 1,23$, se zdá, že srovnávané rozptyly lze považovat za srovnatelné.

Na základě Leveneho testu (p -hodnota = 0,781) nezamítáme předpoklad o shodě rozptylů.

Výrobce	Rozptyl
Amber	694,9
Bright	684,3
Clear	582,5
Dim	564,8

Tabulka 6: Srovnání rozptylů světelnosti zářivek při 5°C, podle výrobce

3.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.)

Předpoklad o symetrii rozdělení nelze zamítnout (viz Tab. 5). Vzhledem k zamítnutí předpokladu o normálním rozdělení je pro srovnání mediánu použit Kruskalův-Wallisův test.

Pro výpočet bodového odhadu jsou použity mediány a pro výpočet 95% intervalového odhadu Wilcoxonův test.

Výrobce	Bodový odhad (lm)	95% intervalový odhad (lm)
Amber	797,4	(789,1 ; 801,5)
Bright	814,2	(805,5 ; 819,5)
Clear	792,2	(787,6 ; 799,6)
Dim	790,0	(783,3 ; 795,3)

Tabulka 7: Odhad mediánů světelnosti zářivek při 5°C, podle výrobce

Medián světelnosti zářivek výrobce Amber očekáváme při teplotě 5 °C na cca 797,4 lm. 95% intervalový odhad tohoto mediánu je (789,1 ; 801,5) lm. Výsledky výrobců Bright, Clear a Dim lze interpretovat obdobně. Výrobce Bright poskytuje při 5°C vyšší světelnost o přibližně 20 lumenů než ostatní výrobci. Výrobci Amber, Clear a Dim se v tomto parametru liší v řádu jednotek lumenů.

3.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.)

Na hladině významnosti 0,05 zamítám hypotézu o shodě mediánů (Kruskalův-Wallisův test, p-hodnota $<<0,001$), tj. mediány světelnosti při teplotě 5°C se významně liší mezi výrobci.

Výrobce	Amber	Bright	Clear	Dim
Amber	x	0,002	1,000	1,000
Bright		x	0,001	$<<0,001$
Clear			x	1,000
Dim				x

Tabulka 8: Dunnová post hoc analýza

Dle Dunnové (varianta bonferroni) post hoc analýzy (viz Tab. 8) můžeme konstatovat, že světelnost při 5°C se statisticky významně liší mezi podskupinami a(Bright) a b(Amber, Clear, Dim).

Tab. 9 znázorňuje setřídění efektů podle výrobců a jejich rozdělení do podskupin. V podskupině b(Amber, Clear, Dim) není statisticky významný rozdíl ve světelnosti zářivek při 5°C.

Výrobce	Efekt				
Bright	15.6	a			
Amber	-1.2		b		
Clear	-6.4		b	c	
Dim	-8.6		b	c	d

Tabulka 9: Efekt, setřídění a rozdělení do podskupin