

**VŠB – Technická univerzita Ostrava**  
**Fakulta elektrotechniky a informatiky**

**PRAVDĚPODOBNOST A STATISTIKA**

**Zadání 4**

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 2017/2018**

**Popis datového souboru**

Pro srovnání kvality LED žárovek byl změřen světelný tok 100 LED žárovek s příkonem 15 W čtyř výrobců (V1, V2, V3, V4). Světelný tok byl změřen nejprve ihned po zakoupení žárovky a následně po 10 000 h provozu.

V souboru *ukol\_4.csv* jsou pro každou LED žárovku uvedeny následující údaje: výrobce (V1, V2, V3, V4), světelný tok LED žárovky změřený po zakoupení a po 10 000 h provozu.

**Poznámka**

Všude dále v tomto textu se žárovkou myslí 15 W LED žárovka. Světelný tok (někdy chybně uváděn jako svítivost) žárovky je udáván v lumenech (lm). Jedná se o údaj, který bývá uveden na obalech žárovek.

Pro pochopení dané problematiky je možno použít např. odkaz:

<http://v-tac.cz/content/8-svitivost-svetelny-tok>

**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**

a) Popište strukturu datového souboru, tj. počty testovaných žárovek dle jednotlivých výrobců. Použijte tabulku četností a výsledky vhodným způsobem vizualizujte.

b) Pomocí nástrojů explorační analýzy srovnajte poklesy světelných toků LED žárovek výrobců V1, V2 po 10 000 h provozu (vzhledem k světelnému toku žárovek po zakoupení). Data vhodně graficky prezentujte (krabicový graf, histogram, q-q graf) a doplňte následující tabulky a text.

*Tab. 1: Výběrové charakteristiky poklesu světelných toků LED žárovek výrobců V1, V2 po 10 000 h provozu (vzhledem k světelnému toku žárovek po zakoupení)*

Pokles světelných toků LED žárovek po 10 000 h provozu (lm)			Po odstranění odlehlých pozorování	
	Výrobce V1	Výrobce V2	Výrobce V1	Výrobce V2
rozsah souboru				
<b>Míry polohy</b>				
minimum				
dolní kvartil				
medián				
průměr				
horní kvartil				
maximum				
<b>Míry variability</b>				
směrodatná odchylka				
variační koeficient (%)				
<b>Míry šikmosti a špičatosti</b>				
šikmost				
špičatost				
<b>Identifikace odlehlých pozorování – vnitřní hradby</b>				
dolní mez				
horní mez				

Jméno:

Číslo zadání: 1

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

**Analýza poklesu světelného toku žárovek výrobce V1 po 10 000 h provozu**

Během testu byl změřen pokles světelného toku ..... kusů žárovek výrobce V1 po 10 000 h provozu. Změřený pokles světelného toku se pohyboval v rozmezí ..... až ..... lm. Hodnoty poklesu světelného toku ležící mimo interval ..... až ..... lm (vnitřní hranby) 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 poklesů světelných toků ..... kusů žárovek. Jejich průměrný pokles světelného toku byl ..... lm, směrodatná odchylka pak ..... lm. U poloviny testovaných žárovek pokles světelného toku nepřekročil ..... lm. V polovině měření se pokles světelného toku pohyboval v rozmezí ..... až ..... lm. Vzhledem k hodnotě variačního koeficientu (.....%) lze / nelze analyzovaný soubor považovat za homogenní.

**Analýza poklesu světelného toku žárovek výrobce V2 po 10 000 h provozu**

Během testu byl změřen pokles světelného toku ..... kusů žárovek výrobce V2 po 10 000 h provozu. Změřený pokles světelného toku se pohyboval v rozmezí ..... až ..... lm. Hodnoty poklesu světelného toku ležící mimo interval ..... až ..... lm (vnitřní hranby) 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 poklesů světelných toků ..... kusů žárovek. Jejich průměrný pokles světelného toku byl ..... lm, směrodatná odchylka pak ..... lm. U poloviny testovaných žárovek pokles světelného toku nepřekročil ..... lm. V polovině měření se pokles světelného toku pohyboval v rozmezí ..... až ..... lm. Vzhledem k hodnotě variačního koeficientu (.....%) lze / nelze analyzovaný soubor považovat za homogenní.

**Ověření normality poklesu světelného toku žárovek výrobce V1 po 10 000 h provozu (na základě explorační analýzy)**

Na základě grafického zobrazení (viz ..... ) a výběrové šikmosti a špičatosti (viz Tab. 1, výběrová šikmost i špičatost leží / neleží v intervalu  $(-2; 2)$ ) lze / nelze předpokládat, že pokles světelného toku LED žárovek výrobce V1 po 10 000 h provozu má normální rozdělení. Dle pravidla  $3\sigma$  / Čebyševovy nerovnosti lze tedy očekávat, že přibližně 95 % / více než 75 % žárovek bude mít pokles světelného toku po 10 000 h provozu v rozmezí ..... až ..... lm.

**Ověření normality poklesu světelného toku žárovek výrobce V2 po 10 000 h provozu (na základě explorační analýzy)**

Na základě grafického zobrazení (viz ..... ) a výběrové šikmosti a špičatosti (viz Tab. 1, výběrová šikmost i špičatost leží / neleží v intervalu  $(-2; 2)$ ) lze / nelze předpokládat, že pokles světelného toku LED žárovek výrobce V2 po 10 000 h provozu má normální rozdělení. Dle pravidla  $3\sigma$  / Čebyševovy nerovnosti lze tedy očekávat, že přibližně 95 % / více než 75 % žárovek bude mít pokles světelného toku po 10 000 h provozu v rozmezí ..... až ..... lm.

**Úkol 2**

Porovnejte poklesy světelných toků LED žárovek výrobců V1, V2 po 10 000 h provozu (vzhledem k světelnému toku žárovek po zakoupení). 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) Vraťte se ke grafické prezentaci z úkolu 1 a vytvořte si úsudek o srovnání poklesů světelných toků žárovek po 10 000 h provozu výrobců V1 a V2.

b) Určete bodové a 95% intervalové odhady pro střední hodnoty (resp. mediány) poklesů světelných toků žárovek po 10 000 hodin provozu pro výrobce V1 a V2.

c) Na hladině významnosti 5 % rozhodněte, zda jsou poklesy světelných toků žárovek výrobců V1 a V2 statisticky významné.

Jméno:

Číslo zadání: 1

d) Určete bodový a 95% intervalový odhad rozdílu středních hodnot (resp. mediánů) poklesů světelných toků žárovek po 10 000 h provozu výrobců V1 a V2.

e) Na hladině významnosti 5 % rozhodněte, zda je rozdíl středních hodnot (resp. mediánů) poklesů světelných toků žárovek (po 10 000 h provozu) výrobců V1 a V2 statisticky významný.

**Úkol 3**

Na hladině významnosti 5 % rozhodněte, zda se hodnoty světelného toku LED žárovek po 10 000 h provozu liší v závislosti na tom, od kterého výrobce pocházejí. (Srovnejte všechny výrobce, tj. V1, V2, V3 a V4.) Posouzení proveďte nejprve na základě explorační analýzy a následně pomocí vhodného statistického testu včetně ověření potřebných předpokladů. V případě, že se kapacity žárovek jednotlivých výrobců statisticky významně liší, určete, které LED žárovky (od kterého výrobce) se statisticky významně odlišují od ostatních, tj. určete homogenní podskupiny výrobců.

a) Daný problém vhodným způsobem graficky prezentujte (vícenásobný krabicový graf, histogramy, q-q grafy).

b) Ověřte normalitu světelných toků po 10 000 h provozu všech čtyř výrobců (empiricky i exaktně).

c) Ověřte homoskedasticitu (shodu rozptylů) světelných toků výrobců V1 až V4 po 10 000 h provozu (empiricky i exaktně).



- d) Určete bodové a 95% intervalové odhady středních hodnot (resp. mediánů) světelných toků žárovek po 10 000 h provozu 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) světelných toků žárovek statisticky významný na hladině významnosti 5 %. Pokud ano, určete pořadí výrobců podle naměřených hodnot světelných toků testovaných žárovek po 10 000 h provozu. (Nezapomeňte na ověření předpokladů pro použití zvoleného testu.)

**Úkol 4**

U každé z žárovek zjistěte, zda světelný tok po 10 000 h provozu poklesl o více než 10 % z původní kapacity (po zakoupení). (Definujte si novou dichotomickou proměnnou (pokles o více než 10 %), která bude nabývat hodnot {ANO, NE}.)

a) Analyzujte skladbu žárovek dle toho, zda vykazují nebo nevykazují pokles světelného toku o více než 10 % (pro výrobce V1, V2, V3, V4). Výsledky prezentujte pomocí kontingenční tabulky, vhodného grafu a vhodné míry kontingence.

b) Určete bodový i 95% intervalový odhad rizika (pravděpodobnosti) poklesu hodnoty světelného toku žárovek o více než 10 % pro výrobce V3. (Nezapomeňte na ověření předpokladu pro použití příslušného intervalového odhadu.)

c) Určete bodový i 95% intervalový odhad relativního rizika poklesu hodnoty světelného toku o více než 10 % pro žárovky od „nejhoršího“ výrobce (vzhledem k žárovkám od „nejlepšího“ výrobce). Výsledky slovně interpretujte.

- d) Určete bodový i 95% intervalový odhad poměru šancí poklesu hodnoty světelného toku o více než 10 % pro žárovky od „nejhoršího“ výrobce (vzhledem k žárovkám od „nejlepšího“ výrobce). Výsledky slovně interpretujte.
- e) Pomocí chí-kvadrát testu nezávislosti rozhodněte, jestli pravděpodobnost poklesu světelného toku o více než 10 % závisí statisticky významně na tom, od kterého výrobce žárovka pochází. (Nezapomeňte na ověření předpokladů testu.)

## 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).