

VŠB - Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky

PRAVDĚPODOBNOST A STATISTIKA

Ostrava, AR 2018/2019

Zadání 20

JMÉNO

STUDENTKY/STUDENTA:

OSOBNÍ ČÍSLO:

JMÉNO CVIČÍCÍ/CVIČÍCÍHO:

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

Popis datového souboru

Výrobce zubní pasty se rozhoduje jaký typ abrazivní složky (A1, A2, A3, A4) do pasty přidat. U abrazivní složky sledujeme velikost zrn (měřeno v nm) a výsledek testu chemické netečnosti (prošel/neprošel). Velikosti zrn abrazivní složky jsou pro každý vzorek zjišťovány automaticky a manuálně pomocí mikroskopu.

V souboru [ukol_20.xlsx](#) jsou pro každý vzorek abraziva uvedeny následující údaje: typ (A1, A2, A3, A4), výsledky měření automaticky, výsledky manuálního měření a výsledek testu chemické netečnosti.

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 vzorků dle jednotlivých typů abraziva. Použijte **tabulku četností** a výsledky **vhodným způsobem vizualizujte**.
- b) Pomocí nástrojů explorační analýzy srovnajte **velikosti zrn typu A1 a typu A2 měřených automaticky**. Data vhodně **graficky prezentujte** (vícenásobný krabicový graf, histogram, q-q graf) a doplňte následující tabulky a text.

Tab. 1: Výběrové charakteristiky velikost zrn abrazivní složky (nm) typu A1 a typu A2

Velikost zrn abrazivní složky (nm)		Po odstranění odlehlých pozorování		
	Typ A1	Typ A2	Typ A1	Typ A2
rozsah souboru				
Míry polohy				
minimum				
dolní kvartil				
medián				
průměr				
horní kvartil				
maximum				
Míry variability				
směrodatná odchylka				
variační koeficient (%)				
Míry šikmosti a špičatosti				
šikmost				
špičatost				
Identifikace odlehlých pozorování - vnitřní hranice				
dolní mez				
horní mez				

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

Dále následuje analýza velikosti zrn měřená automatickou metodou.

Analýza velikosti zrn abrazivní složky typu A1

Během testu byla změřena velikost zrn abrazivní složky u vzorků typu A1. Změřená velikost zrn se pohybovala v rozmezí až nm. Hodnoty velikosti zrn ležící mimo interval až nm (vnitřní hradby) 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 velikosti zrn abrazivní složky u vzorků typu A1. Jejich průměrná velikost byla nm, směrodatná odchylka velikosti pak nm. U poloviny testovaných vzorků velikost zrn nepřekročila nm. V polovině měření se velikost zrn pohybovala v rozmezí až nm. Vzhledem k hodnotě variačního koeficientu (.....%) lze / nelze analyzovaný soubor považovat za homogenní.

Analýza velikosti zrn abrazivní složky typu A2

Během testu byla změřena velikost zrn abrazivní složky u vzorků typu A2. Změřená velikost zrn se pohybovala v rozmezí až nm. Hodnoty velikosti zrn ležící mimo interval až nm (vnitřní hradby) 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 velikosti zrn abrazivní složky u vzorků typu A2. Jejich průměrná velikost byla nm, směrodatná odchylka velikosti pak nm. U poloviny testovaných vzorků velikost zrn nepřekročila nm. V polovině měření se velikost zrn pohybovala v rozmezí až nm. Vzhledem k hodnotě variačního koeficientu (.....%) lze / nelze analyzovaný soubor považovat za homogenní.

Ověření normality velikosti zrn abrazivní složky typu A1 (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 velikost zrn abrazivní složky typu A1 má normální rozdělení. Dle pravidla 3σ / Čebyševovy nerovnosti lze tedy očekávat, že přibližně 95 % / více než 75 % vzorků bude mít velikost zrn v rozmezí až nm.

Ověření normality velikosti zrn abrazivní složky typu A2 (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 velikost zrn abrazivní složky typu A2 má normální rozdělení. Dle pravidla 3σ / Čebyševovy nerovnosti lze tedy očekávat, že přibližně 95 % / více než 75 % vzorků bude mít velikost zrn v rozmezí až nm.

Úkol 2.

Porovnejte velikosti zrn abrazivní složky **typu A1 a typu A2** měřené **automatickou metodou**. Dále ověřte, zda se dají použité metody měření (manuální a automatická) považovat za srovnatelné. 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 velikosti zrn abrazivní složky typu A1 a typu A2 měřené automatickou metodou.
- b) Určete bodové a 95% intervalové odhady pro střední hodnoty (resp. mediány) velikosti zrn abrazivní složky typu A1 a typu A2 měřené automatickou metodou.
- c) Určete bodový a 95% intervalový odhad rozdílu středních hodnot (resp. mediánů) velikosti zrn abrazivní složky typu A1 a typu A2 měřené automatickou metodou.
- d) Na hladině významnosti 5 % rozhodněte, zda je rozdíl středních hodnot (resp. mediánů) velikosti zrn abrazivní složky typu A1 a typu A2 měřené automatickou metodou statisticky významný.
- e) Určete bodový a 95% intervalový odhad střední hodnoty (resp. mediánu) rozdílů velikosti zrn zjištěných manuální a automatickou metodou. srovnatelné. (použijte veškerá data, nezávisle na typu abraziva)
- f) Na hladině významnosti 5 % rozhodněte, zda jsou manuální a automatická metoda srovnatelné. Tedy, zda je rozdíl mezi oběma měřeními statisticky významný. (použijte veškerá data, nezávisle na typu abraziva)

Úkol 3.

Na hladině významnosti 5 % rozhodněte, zda se velikosti zrn abrazivní složky měřené **automatickou metodou** liší v závislosti na typu abraziva. Srovnajte všechny typy abraziva, tj. **A1, A2, A3, A4**. Posouzení proveďte nejprve na základě explorační analýzy a následně pomocí vhodného statistického testu, včetně ověření potřebných předpokladů. V případě, že se velikosti zrn měřené automatickou metodou pro jednotlivé typy abrazivní složky statisticky významně liší, určete, které typy abrazivní složky se statisticky významně odlišují od ostatních, tj. určete homogenní podskupiny typů abrazivních složek (setřídte sestupně dle velikosti zrn).

- a) Daný problém vhodným způsobem graficky prezentujte (vícenásobný krabicový graf).
- b) Ověřte normalitu velikosti zrn abrazivní složky měřené automatickou metodou všech čtyř typů (**empiricky i exaktně**).
- c) Ověřte homoskedasticitu (shodu rozptylů) velikostí zrn abrazivních složek měřených automatickou metodou pro všechny typy abraziv (**empiricky i exaktně**).
- d) Určete bodové a 95% intervalové odhady středních hodnot (resp. mediánů) velikostí zrn abrazivních složek měřených automatickou metodou pro všechny typy abraziv. (Nezapomeňte na ověření předpokladů pro použití příslušných intervalových odhadů.) **Použijte takové odhady, aby korespondovaly s vícevýběrovým testem použitým v bodě e**.
- e) Čistým testem významnosti ověřte, zda je pozorovaný rozdíl středních hodnot (resp. mediánů) velikostí zrn abrazivních složek měřených automatickou metodou statisticky významný na hladině významnosti 5 %. Pokud ano, určete pořadí typů abraziv dle naměřených velikostí zrn vzestupně. (Nezapomeňte na ověření předpokladů pro použití zvoleného testu.)

Úkol 4.

Zjistěte (empiricky i exaktně), zda existuje statisticky významná závislost mezi typem abraziva a výsledkem chemického testu. **Použijte veškerá dostupná data.**

- a) Analyzujte skladbu abraziv dle toho, zda prošla, nebo neprošla chemickým testem (pro výrobce A1, A2, A3, A4). Výsledky prezentujte pomocí kontingenční tabulky, vhodného grafu a vhodné míry kontingence.

- b) Určete bodový i 95% intervalový odhad rizika (pravděpodobnosti) negativního výsledku chemického testu pro typ A3. (Nezapomeňte na ověření předpokladu pro použití příslušného intervalového odhadu.)

- c) Určete bodový i 95% intervalový odhad relativního rizika negativního výsledku chemického testu abraziva typu A3 vzhledem k ostatním typům dohromady (zde skupina A3 vs. skupina A1+A2+A4). Výsledky slovně interpretujte.

- d) Určete bodový i 95% intervalový odhad poměru šancí negativního výsledku chemického testu abraziva typu A3 vzhledem k ostatním typům dohromady (zde skupina A3 vs. skupina A1+A2+A4). Výsledky slovně interpretujte.

- e) Pomocí chí-kvadrát testu nezávislosti rozhodněte, jestli pravděpodobnost negativního výsledku chemického testu závisí statisticky významně na typu testovaného abraziva. (Nezapomeňte na ověření předpokladů testu.)

Jak identifikovat, zda jsou v datech odlehlá pozorování?

Emiprické posouzení:

- použití vnitřních (vnějších) hradeb, resp. z – souřadnice, resp. mediánová souřadnice,
- 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?

Emiprické 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)?

Emiprické 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).