**VŠB – Technická univerzita Ostrava**

**Fakulta elektrotechniky a informatiky**

**Biostatistika**

**ZADÁNÍ 7**

|  |  |
| --- | --- |
| 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 2018/2019**

**Popis datového souboru**

V datovém souboru *diabetes\_2.xlsx* najdete údaje o 3 skupinách pacientů. O pacientech, u nichž byl diagnostikován diabetes mellitus I. typu, o pacientech, u nichž byl diagnostikován diabetes mellitus II. typu a pacientech, u nichž diabetes nebyl diagnostikován. V souboru jsou uvedeny hodnoty glykemie měřeny nalačno při odběru, na jehož základě bylo diagnostikováno onemocnění diabetem a informace o tom, zda byl diabetes diagnostikován alespoň u jednoho z rodičů pacienta. U pacientů, u nichž byl diagnostikován diabetes, jsou navíc uvedeny hodnoty glykemie zjištěné nalačno při kontrolním odběru uskutečněném 6 měsíců po zahájení léčby.

**Obecné pokyny:**

* Domácí úkoly odevzdávejte vždy v termínu, který určil váš cvičící.
* Portfolio domácích úkolů budete odevzdávat postupně. Tj. nejdříve odevzdáte titulní stránku + úkol 1, následně doplníte úkol 2, atd.
* Domácí úkoly zpracujte dle obecně známých typografických pravidel.
* **Všechny** tabulky i obrázky musí být opatřeny titulkem.
* Do domácích 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**

1. Prezentujte strukturu datového souboru, tj. strukturu pacientů dle jejich zdravotního stavu (diabetes mellitus I. typu, diabetes mellitus II. typu, diabetes nediagnostikován). Použijte tabulku četností a výsledky vhodným způsobem vizualizujte.
2. Srovnejte hodnoty poklesu glykemie během šesti měsíců u pacientů s diabetem I. a II. typu na základě explorační analýzy, data graficky prezentujte (histogram, vícenásobný krabicový graf) a doplňte následující tabulky a text.

Tab. 1: Výběrové charakteristiky poklesu glykemie během šesti měsíců u pacientů s diabetem I. a II. typu

**o**

|  |  |  |
| --- | --- | --- |
| **Pokles glykemie (mmol/l)** |  | **Po odstranění odlehlých pozorování** |
|   | **Diabetes Mellitus I. typu** | **Diabetes Mellitus II. typu** | **Diabetes Mellitus I. typu** | **Diabetes Mellitus II. typu** |
| **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 hradby** |  |  |  |  |
| **dolní mez** |  |  |  |  |  |  |
| **horní mez** |  |  |  |  |  |  |

**Grafická prezentace (histogramy, vícenásobný krabicový graf):**

**Pacienti s diabetem I. typu**

Byly analyzovány záznamy o poklesu glykemie …………. pacientů, u nichž byl diagnostikován diabetes mellitus I. typu (viz Tab. 1). Hodnoty poklesů glykemie se pohybovaly v rozmezí …………… až …………………. mmol/l. Hodnoty poklesů glykemie ležící mimo interval ……………… až ………………….. mmol/l byly identifikovány jako odlehlá pozorování a příslušní pacienti byli z dalšího zpracování vyřazeni. / Žádné z měření nebylo identifikováno jako odlehlé pozorování. Níže uvedené výsledky pocházejí z analýzy datového souboru o rozsahu ………………. pacientů.

Průměrný pokles glykemie byl …………….. mmol/l, směrodatná odchylka ………………….. mmol/l. U poloviny pacientů byl pozorován pokles glykemie o méně než …………….. mmol/l. (Podrobněji: U čtvrtiny pacientů s diabetem I. typu byl pozorován pokles glykemie o méně než ………………… mmol/l, u čtvrtiny pacientů byl pozorován pokles glykemie o více než ………………… mmol/l.) Vzhledem k hodnotě variačního koeficientu (……….%) lze / nelze analyzovaný soubor považovat za homogenní.

Obdobně lze popsat výsledky analýzy poklesu glykemie pacientů, u nichž byl diagnostikován diabetes mellitus II. typu.

**Ověření normality glykemie pacientů s diabetem I. typu 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 $\left〈-2;2\right〉$) lze / nelze předpokládat, že glykemie pacientů s diabetem I. typu má normální rozdělení. Dle pravidla 3$σ$ / Čebyševovy nerovnosti lze tedy očekávat, že 95% / více než 75% pacientů s diabetem bude mít hodnotu glykemie v rozmezí …………..……………. až ……………..……… mmol/l.

**Úkol 2**

Porovnejte pokles glykemie u pacientů s diabetem mellitus I. a II. typu. 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.

1. Vraťte se ke grafické prezentaci z úkolu 1 a vytvořte si úsudek o srovnání poklesů glykemie u pacientů s diabetem mellitus I. a II. typu.
2. Určete bodové a 95% intervalové odhady středního poklesu glykemie (resp. mediánu poklesu glykemie) pro obě skupiny pacientů.
3. Čistým testem významnosti ověřte, zda je pozorovaný pokles glykemie statisticky významný na hladině významnosti 5%. Test proveďte pro obě skupiny pacientů.
4. Určete bodový a 95% intervalový odhad rozdílu středních hodnot (resp. mediánů) poklesů glykemie srovnávaných skupin pacientů.
5. Na hladině významnosti 5 % rozhodněte, zda je rozdíl středních hodnot (resp. mediánů) poklesů glykemie u srovnávaných skupin pacientů statisticky významný.

**Úkol 3**

Na hladině významnosti 5 % rozhodněte, zda se statisticky významně liší hodnoty glykemie měřeny nalačno při odběru, na jehož základě bylo diagnostikováno onemocnění diabetem (uvažujte nejen pacienty s diabetem mellitus I. a II. typu, ale i pacienty, u nichž diabetes mellitus diagnostikován nebyl). Posouzení proveďte nejprve na základě explorační analýzy a následně pomocí vhodného statistického testu, včetně ověření potřebných předpokladů.

1. Daný problém vhodným způsobem graficky prezentujte (vícenásobný krabicový graf, histogramy, q-q grafy).
2. Ověřte normalitu glykemie všech tří skupin pacientů empiricky i exaktně.
3. Ověřte homoskedasticitu (shodu rozptylů) glykemie srovnávaných skupin pacientů (empiricky i exaktně).
4. Určete bodové i 95% intervalové odhady středních hodnot (resp. mediánů) glykemie u srovnávaných skupin pacientů. (Nezapomeňte na ověření předpokladů pro použití příslušných intervalových odhadů.)
5. Na hladině významnosti 5 % ověřte, zda je pozorovaný rozdíl středních hodnot (resp. mediánů) glykemie u srovnávaných skupin pacientů statisticky významný. Pokud ano, zjistěte, zda lze některé skupiny pacientů označit (z hlediska jejich glykemie) za homogenní. (Nezapomeňte na ověření předpokladů pro použití zvoleného testu.)

**Úkol 4**

Analyzujte vliv výskytu diabetu v rodinné anamnéze na výskyt diabetu II. typu. Analýzu proveďte na základě údajů o pacientech s diabetem II. typu a pacientech s nediagnostikovaným diabetem.

1. Pomocí vhodné asociační tabulky prezentujte vliv výskytu diabetu II. typu v rodinné anamnéze na výskyt diabetu II. typu. Tabulku doplňte o řádkové relativní četnosti.
2. Analyzovanou závislost prezentujte pomocí vhodného grafu.
3. Určete alespoň jednu míru kontingence. Na základě posouzení vizualizace analyzované závislosti a nalezené míry kontingence posuďte míru analyzované závislosti.
4. Určete šanci na onemocnění diabetem II. typu u obou skupin pacientů, tj. u pacientů, v jejichž rodinné anamnéze se diabetes vyskytoval i u těch, v jejichž rodinné anamnéze se diabetes nevyskytoval. (Komentujte.)
5. Určete relativní šanci na onemocnění diabetem II. typu u rizikové skupiny pacientů (včetně 95% intervalového odhadu). Na základě svého zjištění určete, zda lze vliv výskytu diabetu v rodinné anamnéze považovat za statisticky významný na hladině významnosti 5%.
6. Určete riziko onemocnění diabetem II. typu (včetně 95% intervalového odhadu) u obou skupin pacientů, tj. u pacientů, v jejichž rodinné anamnéze se diabetes vyskytoval i u těch, v jejichž rodinné anamnéze se diabetes nevyskytoval. Nezapomeňte na ověření předpokladů pro použití příslušných intervalových odhadů.
7. Určete relativní riziko onemocnění diabetem II. typu u rizikové skupiny pacientů (včetně 95% intervalového odhadu). Na základě svého zjištění určete, zda lze vliv výskytu diabetu v rodinné anamnéze považovat za statisticky významný na hladině významnosti 5%.
8. Ověřte, zda existuje závislost mezi výskytem diabetu v rodinné anamnéze a výskytem diabetu II. typu u pacienta pomocí Chí-kvadrát testu nezávislosti. Nezapomeňte ověřit předpoklady pro použití testu.

## Jak identifikovat, zda jsou v datech odlehlá pozorování?

*Empirické 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?

*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).*