

Pravděpodobnost a statistika - verze 240603B

(0 – 60) bodů, požadované minimum: 30 bodů

Praktická část - (0 – 50) bodů

1. Student při odpovídání na otázku s více možnými odpověďmi (4 odpovědi, z toho je jedna správná) s pravděpodobností 70 % zná správnou odpověď, s pravděpodobností 30 % správnou odpověď nezná a “hádá”.

- a) S jakou pravděpodobností student na danou otázku odpoví správně? **(5b)**
b) Student na danou otázku odpověděl správně. S jakou pravděpodobností správnou odpověď pouze “uhádl”? **(5b)**

2. Hmotnost kapra (v kilogramech) pocházejícího z Brčálníku modelujeme spojitou náhodnou veličinou X , jejíž distribuční funkce je dána následujícím předpisem:

$$F_X(x) = \begin{cases} 0 & \text{pro } x \in (-\infty, 0) \\ c(x^2 - 20x) & \text{pro } x \in (0, 10) \\ d & \text{pro } x \in (10, \infty) \end{cases}$$

- a) Určete hodnoty konstant c a d . **(2b)**
b) Určete předpis hustoty pravděpodobnosti $f_X(x)$ a načrtněte její graf. **(3b)**
c) Určete medián hmotností kaprů z Brčálníku. **(3b)**
d) Určete pravděpodobnost, že kapr z Brčálníku bude mít hmotnost větší než 1kg. **(2b)**
3. Počet metrů čtverečních povrchu, jež jsme schopni obrousit jedním smirkovým papírem označme X . Předpokládejme, že $X \sim N(\mu, \sigma^2)$, kde $\mu = 1 \text{ m}^2$ a $\sigma = 0,1 \text{ m}^2$.
- a) Určete pravděpodobnost, že jedním smirkovým papírem obrousíme $0,8 \text{ m}^2$ až $1,0 \text{ m}^2$ povrchu. **(2b)**
b) Načrtněte graf hustoty pravděpodobnosti náhodné veličiny X a vyznačte v něm plochu, jejíž velikost je rovna pravděpodobnosti určené v (a). **(2b)**
c) S 30% pravděpodobností nejsme schopni jedním smirkovým papírem obrousit plochu větší než $a \text{ m}^2$. Určete hodnotu a . **(2b)**
d) Určete pravděpodobnost, že pomocí čtyřiceti smirkových papírů zvládneme obrousit 40 m^2 až 41 m^2 povrchu. **(4b)**

Popis datového souboru potřebného pro úlohy 4 a 5:

Byla provedena studie hodnotící výnosy čtyř odrůd pšenice (A, B, C, D). Každá odrůda byla pěstována v letech 2022 a 2023 na 100 polích. V roce 2023 byly z hlediska počasí výrazně lepší podmínky, což pravděpodobně přispělo k vyšším výnosům.

V datovém souboru jsou pro každé pole uvedeny následující údaje:

- *id* ... identifikační číslo pole,
- *odruda* ... název odrůdy, která byla na tomto poli pěstována (A/B/C/D),
- *vynos_2022* ... výnos pole v roce 2022 v tunách na hektar (t/ha),
- *vynos_2023* ... výnos pole v roce 2023 v tunách na hektar (t/ha),
- *typ_hnojiva* ... typ hnojiva použitý na poli (syntetické/přirodní).

V následujících úkolech bude zapotřebí analyzovat meziroční zvýšení výnosu, tedy

$$\text{zvyseni} = \text{vynos_2023} - \text{vynos_2022} \text{ (t/ha)}.$$

4. Na základě datového souboru analyzujte meziroční **zvýšení výnosu pšenice odrůdy A**. Nezapomeňte ověřit předpoklady pro použití metod statistické indukce.
 - a) Identifikujte odlehlá pozorování a z následující analýzy je vyřaďte. Uveďte *id* polí, která nebyla do další analýzy zařazena. **(1b)**
 - b) Určete bodový a 95% jednostranný intervalový odhad střední hodnoty (popřípadě mediánu) meziročního **zvýšení výnosu odrůdy A**. **(3b)**
 - c) Čistým testem významnosti na hladině významnosti 5 % ověřte, zda je meziroční **zvýšení výnosu odrůdy A** statisticky významné. Okomentujte výsledek v návaznosti na intervalový odhad z bodu (b). **(3b)**
 - d) Odhadněte pravděpodobnost (bodově a pomocí 95% oboustranného intervalového odhadu), že na náhodně vybraném poli, na kterém se pěstuje **odrůda A** byl v roce 2023 **nižší** výnos než v roce 2022 (tedy, že meziroční zvýšení výnosu pro dané pole bylo menší než 0). **(3b)**
5. Dodavatel **odrůdy A** se rozhodl prozkoumat, zda má **typ hnojiva** vliv na meziroční **změnu výnosu** (pouze pro pole, na kterých se pěstuje **odrůda A**). Konkrétně ho zajímá zda může přírodní hnojivo beztrátově nahradit syntetickým, t.j. zda existuje závislost mezi použitým **typem hnojiva** (syntetického nebo přírodního) a meziroční **změnou výnosu** (snížením nebo zvýšením) (t/ha).
 - a) Uveďte asociační tabulku vhodnou pro analýzu závislosti mezi **typem hnojiva** a **změnou výnosu**. Uspořádejte tabulku s ohledem na bod (c), tedy exponovaná pole jsou ta, na nichž je použito **syntetické** hnojivo a sledovanou událostí je **snížení výnosu**. **(2b)**
 - b) Načrtněte graf pro vizualizaci závislosti mezi **typem hnojiva** a **změnou výnosu** (korepondující s tabulkou uvedenou v bodě (a)) a na základě asociační tabulky, daného grafu a vhodné míry kontingence interpretujte závěry, k nimž jste ohledně sledované závislosti došli. **(2b)**
 - c) Určete bodový a 95% intervalový odhad **relativního rizika snížení výnosu** pro pole, na nichž bylo použito **syntetické** hnojivo. Výsledek interpretujte. **(3b)**
 - d) Pomocí intervalového odhadu z bodu (c), rozhodněte na hladině významnosti 5 %, zda je pozorovaná závislost mezi typem hnojiva a meziroční změnou výnosu statisticky významná. Pokud ano, uveďte který typ hnojiva vykazuje statisticky významnější **snížení výnosu**. **(3b)**