

Statistická analýza dat

Statistické minimum – příklady zkouškových otázek.

1. Definujte pojem *střední hodnota* náhodné veličiny a *aritmetický průměr*. Vysvětlete rozdíl mezi nimi.
2. Uvažujte náhodný vektor \mathbf{X} . Definujte *kovarianční* a *korelační matici*. Jaké mají tyto matice vlastnosti? K čemu se dají použít?
3. Co je to *distribuční funkce*? Co je to *kvantil* a *kvantilová funkce*? Co je to *funkce pravděpodobnostní hustoty*? Definujte pojmy formálně a uveďte jaký je mezi nimi vztah.
4. Vysvětlete vlastními slovy význam *p-hodnoty* statistického testu. Jaký je výklad *hladiny významnosti* α ? Předpokládejte, že *p-hodnota* nějakého testu je 0.045? Jaká je pravděpodobnost, že nulová hypotéza tohoto testu neplatí? Jak to souvisí s α ?
5. Vysvětlete pojem *interval spolehlivosti* (confidence interval). Co nejpřesněji popište, jak byste spočítal/a 99% interval spolehlivosti odhadu střední hodnoty normálního rozdělení z malého vzorku velikosti m .
6. Vysvětlete význam pojmů *chyba prvního druhu* a *chyba druhého druhu* používaných k popsání konkrétních chyb v procesu testování statistických hypotéz. Vysvětlete jak spolu tyto chyby souvisí. Popište, jak se dá výskyt těchto chyb ovlivnit.
7. Formulujte *centrální limitní větu*. Kde se dá využít?
8. Motivujte zavedení *Studentova t-rozdělení*. Ke kterému rozdělení se *t-rozdělení* asymptoticky blíží s rostoucím počtem stupňů volnosti? Vysvětlete za jakých okolností a jak se od tohoto rozdělení odlišuje. K čemu se používá?
9. Definujte *věrohodnostní funkci* (likelihood). K čemu se používá metoda maximální věrohodnosti? Pojmenujte alespoň dvě metody, které se k maximalizaci věrohodnosti používají. Vysvětlete, proč se často používá logaritmus věrohodnosti. jak se dá věrohodnosti využít při testování statistických hypotéz?
10. Formálně definujte *multivariátní normální rozdělení*. Kolik parametrů budeme obecně potřebovat pro popis tohoto rozdělení v d dimenzích? Lze počet těchto parametrů nějak omezit? Jaké důsledky toto omezení může mít?
11. Definujte alespoň dvě často používaná rozdělení diskretní náhodné veličiny. Pojmenujte jejich parametry. Na příkladech naznačte, kdy se tato rozdělení dají využít.
12. Vysvětlete rozdíl mezi parametrickým a neparametrickým statistickým testem. Pojmenujte základní výhody a nevýhody obou přístupů. Jmenujte alespoň jeden parametrický a jeden neparametrický test.
13. K následující statistické úloze přiřaďte vhodný test. Bylo testováno 11 automobilů určité značky. Ověřte, zda se jejich pravé a levé přední pneumatiky ojíždějí srovnatelně. (Předpokládejte, že ojetí pneumatik [mm] má normální rozdělení. Z následujících testů vyberte ten nejlepší, svou volbu vysvětlete (dvouvýběrový t-test, Friedmanův test, jednovýběrový t-test, jednovýběrový Wilcoxonův test, test o parametru alternativního rozdělení, test o rozptylu normálního rozdělení, párový t-test). Co by se stalo pokud byste použili druhý nejlepší test z daného seznamu?