

**Ukázkový test k přijímacímu řízení
Statistika a data science**

- 1) Při použití chí-kvadrát testu nezávislosti v kontingenční tabulce o 5 řádcích a 3 sloupcích má testová statistika:
 - a) $3 \times 5 = 15$ stupňů volnosti
 - b) $15 - 1 = 14$ stupňů volnosti
 - c) $(5 - 1) \times (3 - 1) = 8$ stupňů volnosti**
 - d) $15 - 3 - 5 = 7$ stupňů volnosti

- 2) Jsou porovnávány platy zaměstnanců ve dvou obdobích. V druhém období se plat každého zaměstnance zvýšil o 200 Kč. Pro charakteristiky platů ve druhém období platí:
 - a) průměr se zvýšil o 200 Kč, rozptyl se také zvýšil o 200 Kč
 - b) průměr se zvýšil o 200 Kč, rozptyl se nezměnil**
 - c) nezměnil se ani průměr, ani rozptyl
 - d) průměr se nezměnil, směrodatná odchylka se zvýšila o 200 Kč

- 3) Trend roční časové řady určitého ukazatele (v tis. Kč) lze popsat rovnicí $T_t = 9,836 - 0,142 \cdot t$. Na základě odhadnutých parametrů lze tvrdit, že hodnota ukazatele se každý rok (oproti roku předcházejícímu) v průměru:
 - a) sníží o 142 Kč**
 - b) sníží o 13,2 %
 - c) sníží o 984 Kč
 - d) sníží o 86,76 %

- 4) Všechny teploty, naměřené v lednu v horské meteorologické stanici, se pohybovaly pod bodem mrazu od -22 v noci do -2 ve dne. Rozptyl těchto teplot
 - a) bude v intervalu $(-22, -2)$
 - b) bude nezáporný**
 - c) bude záporný
 - d) není pro záporné hodnoty definován

- 5) Pro indikátorové proměnné typu dummy v regresní analýze platí:
- používají se pro kvantitativní vysvětlující proměnné
 - je jich právě tolik, jaký je počet kategorií vysvětlující proměnné, ke které se vztahují
 - je jich vždy o jednu méně, než jaký je počet kategorií vysvětlující proměnné, ke které se vztahují**
 - mohou nabývat hodnot 0, 1 nebo -1
- 6) Které z následujících tvrzení nejlépe vysvětluje, proč u některých dat platí: modus = medián = aritmetický průměr?
- medián se vždy rovná průměru u diskrétních dat
 - nejčtetnější hodnota je často zároveň prostřední hodnotou
 - průměr vždy leží uprostřed seřazených dat
 - datový soubor je symetrický**
- 7) Pravděpodobnost, že pojišťovací agent uzavře na jedné schůzce pojistnou smlouvu, je 0,1. Předpokládejme, že absolvuje 25 schůzek. Jak by se změnil rozptyl počtu uzavřených smluv, pokud by se pravděpodobnost uzavření smlouvy zvýšila z 0,1 na 0,5, přičemž počet schůzek zůstane stejný?
- rozptyl by se zmenšil
 - rozptyl by se zvětšil**
 - rozptyl by zůstal stejný
 - nelze rozhodnout
- 8) Které z následujících tvrzení správně popisuje použití statistických metod při zkoumání závislosti mezi různými typy proměnných?
- analýza rozptylu je vhodná pro určení, jak se liší rozptyly mezi jednotlivými hodnotami faktoru
 - korelační analýza je vhodná pro kategoriální proměnné
 - chí-kvadrát test nezávislosti se používá pro dvě kategoriální proměnné**
 - regresní analýza nevyžaduje kvantitativní závislou proměnnou
- 9) Které kritérium je konstruováno tak, aby při rozhodování o kvalitnějším regresním modelu penalizovalo zbytečně velký počet parametrů?
- upravený koeficient determinace**
 - koeficient determinace
 - reziduální rozptyl
 - reziduální součet čtverců

- 10) tvrzení nejlépe vystihuje důsledek centrální limitní věty?
- a) průměr z velkého výběru má stejnou variabilitu jako jednotlivé pozorování
 - b) rozdělení průměru má menší rozptyl než rozdělení jednotlivých hodnot**
 - c) centrální limitní věta platí pouze pro normální rozdělení
 - d) s rostoucím počtem pozorování roste variabilita průměru
- 11) Byla zjišťována závislost výdělku Y (Kč) v jistém typu zaměstnání na počtu odpracovaných let X . Jako vhodný model závislosti byla zvolena regresní přímka. Jak se změní směrnice této přímky, pokud výdělek budeme udávat v Eurech místo v korunách? Uvažujte převodní kurz 1 euro = 24 korun.
- a) hodnota směrnice se nezmění
 - b) hodnota směrnice se zvýší 24krát
 - c) hodnota směrnice se sníží 24krát**
 - d) hodnota směrnice se zvýší o 24 %
- 12) Průměr 18 hodnot ve výběru je 100 a rozptyl (definovaný jako průměrná kvadratická odchylka od průměru) je 400. Jak se změní průměr a rozptyl, přidáme-li do souboru hodnoty 90 a 110?
- a) průměr se nezmění, nový rozptyl je 370**
 - b) průměr se nezmění, pro výpočet rozptylu je málo informací
 - c) žádná ze statistik se nezmění
 - d) průměr je 100, nový rozptyl je 394
- 13) Firma prodávala svůj výrobek ve dvou prodejnách. V první prodejně prodala 100 ks za jednotkovou cenu 35 Kč. Ve druhé prodejně prodala výrobky za jednotkovou cenu 45 Kč. Celkové tržby firmy činily 6 740 Kč. Jaká byla průměrná cena prodaných výrobků?
- a) 35 Kč
 - b) 40 Kč
 - c) 39,2 Kč**
 - d) 38,0 Kč
- 14) Při testu hypotézy o střední hodnotě v normálním rozdělení byla určena hodnota testového kritéria 2,47 a kritický obor $W_{0,05} = (2,7; \infty)$. Potom
- a) p-hodnota pro tento test je určitě menší než 0,05
 - b) testové kritérium leží v kritickém oboru, a proto nulovou hypotézu zamítáme
 - c) nulovou hypotézu nezamítáme na hladině významnosti 0,05**
 - d) p-hodnota tohoto testu je rovna 0,05

- 15) Pro neznámou pravděpodobnost výhry ve hře byl sestroyen dvoustranný 95% interval spolehlivosti (0,32; 0,38). Potom při testu hypotézy $\pi = 0,4$ proti dvoustranné alternativě
- nezamítneme nulovou hypotézu na 5% hladině významnosti
 - zamítneme nulovou hypotézu ve prospěch alternativní hypotézy na 1% hladině významnosti
 - zamítneme nulovou hypotézu ve prospěch alternativní hypotézy na 5% hladině významnosti**
 - nelze rozhodnout, byly by třeba dva jednostranné 95% intervaly
- 16) U 100 výrobků bylo zjištěno, že 12 má prošlou záruční lhůtu a 18 má poškozený obal. Výrobků bez vady je 75. Jaká je pravděpodobnost, že náhodně vybraný výrobek má poškozený obal, víme-li, že má prošlou záruční lhůtu?
- 0,417**
 - 0,05
 - 0,3
 - 0,278
- 17) V souboru osob je 320 mužů a 480 žen. Z tohoto souboru opakovaně náhodně vybíráme osoby s vrácením. Která z následujících událostí je nejpravděpodobnější?
- prvních pět vybraných osob budou ženy
 - mezi prvními pěti vybranými osobami budou právě čtyři ženy**
 - prvních pět vybraných osob budou muži
 - mezi prvními pěti vybranými osobami budou právě čtyři muži
- 18) Máme k dispozici 8 čtvrtletních hodnot HDP pro Českou republiku za období 2023 až 2024: 786,5; 820,0; 807,1; 873,3; 841,3; 873,6; 864,4; 931,4. Jaká je první hodnota vhodného klouzavého průměru, který by tuto časovou řadu zbavil sezónnosti?
- 821,7
 - 825,6
 - 828,6**
 - 819,0

- 19) Ve firmě je 55 % mužů a 45 % žen. Podíl mužů, jejichž výška je více než 180 cm je 8 %, podíl žen je 2 %. Náhodný zaměstnanec firmy je vyšší než 180 cm. Jaká je pravděpodobnost, že je to žena?
- a) 0,02
 - b) 0,1
 - c) 0,25
 - d) 0,17**
- 20) Byl odhadnut klasický regresní model závislosti obratu firmy Y (v tis. Kč) na počtu zaměstnanců X : $Y = 740 + 75 \cdot X$. Model byl odhadnut na základě dat firem s počtem zaměstnanců v intervalu 5 až 20. Které z následujících tvrzení je na základě tohoto modelu správné?
- a) zvýšení počtu zaměstnanců z 8 na 9 zvýší obrat o 75 tis. Kč u každé firmy
 - b) parametr 75 vyjadřuje průměrný nárůst obratu o 75 tis. Kč na jednoho dodatečného zaměstnance**
 - c) parametr 740 je minimální možný obrat firmy
 - d) model nelze použít pro firmy s méně než 10 zaměstnanci