

Vypracujte nasledujúce úlohy pre Vaše dáta. Súbor následne vo formáte html alebo pdf odovzdajte na miesto na to určené v AIS-e, najneskôr deň pred termínom skúšky v decembri. Ak na skúšku v decembri nejdete, termín je 23.12. 23:59:59. Súbor nazvite zadanie_vášepriezvisko. Súbor by mal obsahovať všetky výpočty a slovné komentáre a zdôvodnenie vybraných metód na analýzy. Ak nepíšem inak, hladina významnosti testu je 0.05. Všade kde testujete niečo, treba uviesť aj nulové a alternatívne hypotézy testov. Prosím, pred riešením úlohy uveďte vždy aj jej zadanie (pre ľahšiu kontrolu).

V úlohách, kde filtrujete dáta, teda vytvárate nejaké podmnožiny, vždy vypíšte prvých 10 hodnôt podmnožiny, aby bolo vidieť, že pracujete s požadovanými hodnotami.

Vybrali ste si zaujímavé dáta, zďaleka som vymyslenými zadaniami nevyčerpala ich možnosti. Ale aj tak sa teším na zaujímavé čítanie výsledkov. Je to projekt, tak sa ho prosím snažte na zodpovedajúcej úrovni upraviť a odovzdať. Max. počet bodov je 20, čo predstavuje 20% Vášho celkového hodnotenia z predmetu. Projekt je povinný pre všetkých, ktorí chcú predmet úspešne absolvovať, dolná hranica bodov potrebná k skúške však nie je daná.

Pre porovnanie, že náročnosť je pre každého rovnaká je v dokumente zadanie každého z Vás. Pre tých, ktorí majú dáta odo mňa, boli dáta zaslané v správe v deň zverejnenia zadání.

V prípade akýchkoľvek otázok píšete mail na ballova@math.sk alebo konzultujte na cvičení.

Babich Nikita

1. Využitím popisnej štatistiky charakterizujte premennú RatingDiff Vašich protihráčov. Výsledky zobrazte graficky a popíšte aj použité grafy. Aká je najčastejšie sa vyskytujúca hodnota Vášho RatingDiff?
2. Vhodným testom overte na hladine významnosti 0.01, či je Váš RatingDiff v mesiaci september aspoň 6. Porovnajte Váš RatingDiff za mesiac október a september v tých hrách, ktoré majú Termination- normal. Je tento rozdiel štatisticky významný na hladine významnosti 0.1?
3. Existuje štatisticky významný rozdiel vo Vašom RatingDiff vzhľadom na Termination alebo vzhľadom na mesiac v ktorom prebehla hra? Overte vhodným testom a aj graficky. Existuje aj interakcia medzi premennými?
4. Je medzi RatingDiff víťazného hráča a RatingDiff hráča, ktorý prehral v hrách ktoré končili Time forfeit štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Bilyk Ivan

1. Využitím popisnej štatistiky charakterizujte premennú Vek. Výsledky zobrazte graficky a popíšte aj použité grafy. Akého pohlavia je väčšina zamestnancov?
2. Vhodným testom overte na hladine významnosti 0.1, či je Počet rokov v podniku aspoň 5. Porovnajte Počet rokov v podniku vzhľadom na pohlavie. Je tento rozdiel štatisticky významný na hladine významnosti 0.01?
3. Existuje štatisticky významný vplyv Pozície (zoberme len Grafický dizajnér, Informatik a Inžinier) alebo Pohlavia na vek zamestnanca? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými úrovňami? Existuje interakcia medzi týmito premennými?
4. Je medzi Počtom rokov a Vekom zamestnankýň štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Halušková Anna

1. Využitím popisnej štatistiky charakterizujte premennú Počet samovrážd na 100 tisíc obyvateľov za celé sledované obdobie pre celú Európu. Výsledky zobrazte graficky a popíšte aj použité grafy. Ktoré pohlavie je v súbore viac zastúpené?
2. Vhodným testom overte na hladine významnosti 0.01, či je Počet samovrážd na 100 tisíc obyvateľov na Slovensku za celé sledované obdobie najviac 10. Porovnajte Počet samovrážd na 100 tisíc obyvateľov na Slovensku za celé obdobie vzhľadom na pohlavie. Je tento rozdiel štatisticky významný na hladine významnosti 0.1?
3. Existuje štatisticky významný rozdiel v Počte samovrážd na 100 tisíc obyvateľov vzhľadom na rok (berme len roky 2012, 2013 a 2014) alebo vzhľadom na krajinu (zoberme Slovensko, Rakúsko a Česko)? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými? Existuje aj interakcia medzi premennými?
4. Je medzi Počtom samovrážd (Suicides_no) na Slovensku a premennou Population štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Harmaniak Adam

1. Využitím popisnej štatistiky charakterizujte premennú Náklady firmy. Výsledky zobrazte graficky a popíšte aj použité grafy. Z ktorej oblasti sa v súbore nachádza najviac firiem?
2. Vhodným testom overte na hladine významnosti 0.1, či je zisk firiem aspoň 230 tisíc. Porovnajte náklady firiem, ktoré pôsobia v oblasti služieb s tými, ktoré pôsobia v oblasti výroby. Je tento rozdiel štatisticky významný na hladine významnosti 0.01?
3. Existuje štatisticky významný rozdiel v očakávaných nákladoch vzhľadom na firmu alebo oblasť? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými? Existuje medzi týmito premennými interakcia?
4. Je medzi ziskom firiem, ktoré pôsobia v Európe a ich nákladmi štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Holík Marek

1. Využitím popisnej štatistiky charakterizujte premennú Total. Výsledky zobrazte graficky a popíšte aj použité grafy. Ktorý Field category je v súbore najviac zastúpený?
2. Vhodným testom overte na hladine významnosti 0.1, či je Analytic aptitude najviac 4. Porovnajte Analytic aptitude, vzhľadom na to, či ide o kontaktný alebo bezkontaktný šport. Je tento rozdiel štatisticky významný na hladine významnosti 0.01?
3. Existuje štatisticky významný rozdiel v premennej Total vzhľadom na Field category (berme do úvahy len team sports, winter sports a precision sports) alebo vzhľadom na to či je Durability pod 5 alebo 5 a viac? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými? Existuje aj interakcia medzi premennými?
4. Je medzi Endurance a Nerve bezkontaktných športov štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Hryshenko Kateryna

1. Využitím popisnej štatistiky charakterizujte premennú rating. Výsledky zobrazte graficky a popíšte aj použité grafy. Akého žánru je najviac medzi knihami?
2. Vhodným testom overte na hladine významnosti 0.01, či je rating týchto kníh aspoň 4,7. Porovnajte cenu kníh, ktoré majú rating najviac 4,6 a kníh, ktoré majú rating viac ako 4,6. Je tento rozdiel štatisticky významný na hladine významnosti 0.1?
3. Existuje štatisticky významný rozdiel v hodnotení kníh vzhľadom na rok (berme len roky 2015, 2016 a 2017) a vzhľadom na žánr? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými? Existuje interakcia medzi týmito premennými?
4. Je medzi ratingom a počtom reviewerov non Fiction kníh štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Kaššovic Tomáš

1. Využitím popisnej štatistiky charakterizujte premennú Čas. Výsledky zobrazte graficky a popíšte aj použité grafy. V ktorom roku sa narodil najmladší lyžiar?
2. Vhodným testom overte na hladine významnosti 0.01, či je váha lyžiarov nanajvýš 85 kg. Porovnajte váhu lyžiarov, ktorí debutovali pred a po 19 roku života. Je tento rozdiel štatisticky významný na hladine významnosti 0.05?
3. Existuje štatisticky významný rozdiel v čase vzhľadom na to či lyžiari majú váhu: menej ako 85 kg, od 86 do 92 kg a nad 92 kg alebo vzhľadom na to či majú nad 35 rokov, od 32 do 34 rokov alebo menej ako 32 rokov? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými? Existuje medzi týmito premennými interakcia?
4. Je medzi časom tých, ktorí sú starší ako 30 rokov a ich váhou štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Kretsul Nazarii

1. Využitím popisnej štatistiky charakterizujte premennú Critics' assessment . Výsledky zobrazte graficky a popíšte aj použité grafy. Aký je najčastejší budget pre daný súbor?
2. Vhodným testom overte na hladine významnosti 0.1, či je Critics' assessment najviac 87. Porovnajzte Critics' assessment RPG a FPS genre. Je tento rozdiel štatisticky významný na hladine významnosti 0.01?
3. Existuje štatisticky významný rozdiel v budget vzhľadom na genre (berme len RPG, FPS a Action-Adventure) alebo vzhľadom na úroveň premennej RTX on? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými? Existuje aj interakcia medzi týmito premennými?
4. Je medzi Critics' a players' assessment hier, ktoré majú budget pod 80 000 štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Kromková Stanislava

1. Využitím popisnej štatistiky charakterizujte premennú rating. Výsledky zobrazte graficky a popíšte aj použité grafy. Ktorý type je v súbore najviac zastúpený?
2. Vhodným testom overte na hladine významnosti 0.01, či je rating týchto anime aspoň na úrovni 8,2. Porovnajte počet members pre tie, ktoré majú rating od 7 do 8,2 a tých ktoré majú rating nad 8,2. Je tento rozdiel štatisticky významný na hladine významnosti 0.1?
3. Existuje štatisticky významný rozdiel v rating vzhľadom na type (TV, OVA, other) alebo vzhľadom na to, či ide o menej ako 50 epizód (rátajme aj filmy) alebo aspoň 50 epizód? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými?
4. Je medzi počtom epizód seriálov a ich ratingom štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Legeida Artem

1. Využitím popisnej štatistiky charakterizujte premennú Cena. Výsledky zobrazte graficky a popíšte aj použité grafy. Ktoré vydavateľstvo je v súbore najviac zastúpené?
2. Vhodným testom overte na hladine významnosti 0.01, či je počet strán týchto kníh aspoň 350. Porovnajte cenu kníh, vzhľadom na to, či boli preložené prekladateľom alebo prekladateľkou. Je tento rozdiel štatisticky významný na hladine významnosti 0.05?
3. Existuje štatisticky významný rozdiel v cene a v roku prvého vydania vzhľadom na tri vydavateľstvá: Slovart, Ikar a Odeon? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými?
4. Je medzi cenou kníh, ktoré napísal mužský autor a počtom strán týchto kníh štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Marek Alex

1. Využitím popisnej štatistiky charakterizujte premennú hmotnosť. Výsledky zobrazte graficky a popíšte aj použité grafy. Je v súbore viac mužov s výškou pod 178 cm alebo nad?
2. Vhodným testom overte na hladine významnosti 0.05, či je BMI mužov nanajvýš 28. Porovnajte BMI mužov, ktorí majú menej ako 35 rokov s tými, ktorí majú aspoň 35 rokov. Je tento rozdiel štatisticky významný na hladine významnosti 0.1?
3. Existuje štatisticky významný rozdiel v hmotnosti a výške mužov vzhľadom na to do akej triedy sa zaraďujú podľa hodnoty BMI? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými?
4. Je medzi vekom mužov, ktorí majú menej ako 180 cm a hodnotou ich BMI štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Modránszky Andrej

1. Využitím popisnej štatistiky charakterizujte premennú Cena. Výsledky zobrazte graficky a popíšte aj použité grafy. Z ktorej organizácie je v súbore najviac objektov?
2. Vhodným testom overte na hladine významnosti 0.01, či cena tých, ktoré majú mission_status success je viac ako 40. Porovnajte cenu vzhľadom na rocket_status. Je tento rozdiel štatisticky významný na hladine významnosti 0.1?
3. Existuje štatisticky významný rozdiel v cene vzhľadom na mission_status alebo vzhľadom na to či je organizácia- CASC, SPaceX alebo Rocket Lab? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými? Existuje interakcia medzi týmito premennými?
4. Druhé dáta, lebo neboli v súbore vhodné dáta na tieto analýzy. Je medzi hrúbkou telefónov, ktoré majú dotykový displej a ich kapacitou štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Ondrušová Júlia

1. Využitím popisnej štatistiky charakterizujte premennú Počet absolventov za sledované obdobie na celom Slovensku. Výsledky zobrazte graficky a popíšte aj použité grafy. Študovalo v sledovanom období viac študentov externe alebo dennou formou?
2. Vhodným testom overte na hladine významnosti 0.1, či je počet absolventov najviac 6000. Porovnajte počet absolventov v bratislavskom kraji, ktorí študovali externe s počtom tých, ktorí študovali dennou formou. Je tento rozdiel štatisticky významný na hladine významnosti 0.01?
3. Existuje štatisticky významný rozdiel v počte študentov z krajov západného Slovenska alebo vzhľadom na pohlavie? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými? Existuje interakcia medzi týmito premennými?
4. Je medzi počtom študentov a počtom absolventov, ktorí sú z východného Slovenska štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Palacka Boris

1. Využitím popisnej štatistiky charakterizujte premennú horsepower_bhp. Výsledky zobrazte graficky a popíšte aj použité grafy. Ktorá značka športových áut je najviac zastúpená v súbore?
2. Vhodným testom overte na hladine významnosti 0.01, či top_speed_mph je aspoň 220. Porovnajte top_speed_mph tých áut, ktoré boli vyrobené 2010 a neskôr s tými, ktoré boli vyrobené pred 2010. Je tento rozdiel štatisticky významný na hladine významnosti 0.1?
3. Existuje štatisticky významný rozdiel v engine_size_ci vzhľadom na značku (zoberme len autá značky BMW, Koenigsegg a Mercedes) alebo vzhľadom na to či premenná car_0_60_time_seconds má hodnoty menšie ako 3 alebo 3 a viac? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými? Existuje interakcia medzi týmito premennými?
4. Je medzi top_speed_mph a engine_size_ci áut, ktoré sú vyrobené v roku 2013 a 2014 štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.

Yakymets Taras

1. Využitím popisnej štatistiky charakterizujte premennú Výkon. Výsledky zobrazte graficky a popíšte aj použité grafy. Z ktorej krajiny sa v súbore nachádza najviac áut?
2. Vhodným testom overte na hladine významnosti 0.01, či je priemerný počet prejazdených kilometrov viac ako 90 000. Porovnajte objem motora európskych a ázijských automobilov. Je tento rozdiel štatisticky významný na hladine významnosti 0.1?
3. Existuje štatisticky významný rozdiel v cene vzhľadom na to, či je automobil francúzskej, japonskej alebo americkej výroby alebo vzhľadom na farbu (zoberme iba autá čierne, biele a sivé)? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými? Existuje interakcia medzi týmito premennými?
4. Je medzi výkonom a cenou áut, ktoré majú objem motora aspoň 2 litre, štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.