

**ŠTUDIJSKO LETO 2018/2019**

	<b>Naloga 1</b>	<b>Naloga 2</b>	<b>Naloga 3</b>
<b>Možne točke</b>	10	15	15

## **STATISTIČNA NALOGA**

### ***Navodila***

**Predmet: *Statistika 1*, 1. letnik (RST)**

**Nosilka predmeta: doc.dr. Nuša Erman**

## OSNOVNA NAVODILA

V nadaljevanju so predstavljena navodila za statistično nalogo. Podatki, ki jih boste potrebovali za izdelavo statistične naloge, so objavljeni na spletni strani predmeta Statistika 1 (Moodle FIŠ) v .csv formatu. Vir podatkov je spletna stran <https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps>, na kateri boste našli vse potrebne informacije za predstavitev podatkov.

Naloge rešujte s pomočjo programskega okolja R, s katerim pripravite vse potrebne izračune ter grafične predstavitve. Kodo, s katero boste rešili naloge, posebej shranite in ustrezno poimenujte (*VpisnaŠtevilka.R*).

S pomočjo izpisov in izrisanih grafikonov pripravite poročilo v vnaprej pripravljeni predlogi (*Predloga.docx*), ki je dostopna na spletni strani predmeta Statistika 1 (Moodle FIŠ), dokument preimenujte (za ime uporabite svojo vpisno številko: *VpisnaŠtevilka.docx*). Dokument (poročilo) in .R datoteko z uporabljenimi kodo oddajte preko spletne strani predmeta na Moodle-u, v za to pripravljenem razdelku (Oddaja statistične naloge).

Rok za oddajo statistične naloge je **en dan pred izpitnim rokom**, na katerem boste opravljali pisni izpit (na katerega ste prijavljeni).

## NAVODILA NALOG

### 1. Opisovanje in urejanje podatkov

- a) Na osnovi opisa podatkov (<https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps/home>) natančno opredelite populacijo (stvarna, krajevna, časovna opredelitev), statistično enoto in statistične spremenljivke. Posamezno statistično spremenljivko natančno opredelite (vrsta spremenljivke glede na tip merjenja, vrsta spremenljivke glede na mersko lestvico).
- b) Podatke (AppleStoreData-sample.csv) uvozite v R in jih prikažite (vstavite sliko podatkov).
- c) Pripravite podatke za kasnejšo uporabo na naslednji način:
  - v podatkih obdržite dve številski spremenljivki, in sicer: **size\_bytes** in **price**, ter dve številski spremenljivki, in sicer: **cont\_rating** in **prime\_genre**;
  - smiselno preimenujte spremenljivke tako, da bo ime posamezne spremenljivke sestavljeno iz le enega imena (npr. size\_bytes preimenujemo v Velikost);
  - tako pripravljene podatke ponovno prikažite (vstavite sliko podatkov).

## 2. Analiza številskih spremenljivk

- a) Izračunajte glavne opisne statistike (srednje vrednosti in mere variabilnosti) za obe številski spremenljivki in jih prikažite. Rezultate tudi vsebinsko interpretirajte.
- b) Številski spremenljivki predstavite tudi grafično s pomočjo histograma in okvirja z ročaji. Na osnovi grafičnih prikazov opišite značilnosti posamezne spremenljivke.
- c) S pomočjo številskih spremenljivk narišite razsevni grafikon. Ustrezno preimenujte naslova osi, grafikonu dodajte tudi glavni naslov. Ali razsevni grafikon nakazuje na povezanost obravnavanih spremenljivk? Svoj odgovor pojasnite!
- d) Med obravnavanima spremenljivkama preverite njuno povezanost s pomočjo Pearsonovega korelacijskega koeficienta. Na osnovi njegove vrednosti opredelite smer in jakost povezanosti med obravnavanima spremenljivkama!
- e) Glede na vsebino spremenljivk določite odvisno in neodvisno spremenljivko. Izračunajte ustrezno regresijsko premico, jo zapišite ter vsebinsko interpretirajte. Regresijsko premico vrišite tudi v razsevni grafikon iz točke c) te naloge.
- f) Izračunajte determinacijski koeficient in standardno napako za izračunano regresijsko premico! Na osnovi njunih vrednosti opredelite kakovost dobljenega regresijskega modela!

## 3. Analiza opisnih spremenljivk

- a) Obe opisni spremenljivki najprej predstavite s pomočjo izpisa, ki prikazuje frekvence in odstotke posameznih vrednosti spremenljivk.
- b) Opisni spremenljivki predstavite tudi grafično, izbiro grafikona prilagodite številu vrednosti posamezne spremenljivke. Na osnovi grafičnih prikazov opišite značilnosti posamezne spremenljivke.
- c) Za obravnavani opisni spremenljivki izračunajte kontingenčno tabelo ter jo predstavite.
- d) Na osnovi podatkov v kontingenčni tabeli predpostavite, kateri strukturni odstotki so primernejši za vsebinsko interpretacijo (po vrsticah ali po stolpcih)? Zakaj?
- e) Vsebinsko primernejše strukturne odstotke nato tudi izračunajte in jih prikažite grafično s strukturnimi stolpci! Skrbno uredite tudi legendo, da bodo strukturni stolpci razumljivi.
- f) Na osnovi kontingenčne tabele izračunajte tabelo pričakovanih frekvenc in jo izpišite.
- g) S pomočjo hi-kvadrat testa preverite, ali sta obravnavani spremenljivki med seboj povezani. Izračunajte tudi kontingenčne koeficiente. Na osnovi vrednosti hi-kvadrat testa opredelite, ali med

spremenljivkama obstaja povezava, s pomočjo vrednosti kontingenčnih koeficientov pa opredelite jakost povezanosti med obravnavanima spremenljivkama.

- h) Ugotovitve iz točke g) te naloge smiselno povežite in interpretirajte na osnovi grafičnega prikaza dvorazsežne strukture iz točke e) te naloge.