

Viri podatkov

**1. srečanje: Uvodne teme
(predstavitev načina dela in obveznosti
študentov pri predmetu; razdelitev tem;
osnovni pojmi)**

Mojca Bavdaž (mojca.bavdaz@ef.uni-lj.si)



Petek, 21.2.2020

Izvajalci

Mojca Bavdaž

E-mail: mojca.bavdaz@ef.uni-lj.si

Skype: mojcab

Tel.: 01 5892-630

Lokacija: EF, R-412

Govorilne ure: ponedeljek ob 14h oz. po dogovoru

Janez Štebe

E-mail: janez.stebe@fdv.uni-lj.si

Tel.: 01 5805-292

Lokacija: FDV, C 233

Govorilne ure: torek 10:00 – 12:00

IŠ 1. torek v mesecu 15:00 – 16:00

Gostje iz prakse



Izvedba predmeta

- Kontaktne ure kot kombinacija:
 - predavanj z aktivno udeležbo študentov (diskusije, predstavitve)
 - seminarjev z gosti
 - individualnega in skupinskega reševanja nalog/problemov
- Sprotno delo študentov:
 - Priprave na kontaktne ure
 - Recenzije
 - Predstavitve
 - Seminarska naloga
- Izpit



Struktura ocene

Izpit	50 točk
-------	---------

Sprotno delo s seminarsko nalogo	50 točk
----------------------------------	---------

Iz vsakega dela je potrebno doseči vsaj 25 točk

Ocena:

10	nad 90 točk
----	-------------

9	nad 80 točk
---	-------------

8	nad 70 točk
---	-------------

7	nad 60 točk
---	-------------

6	nad 50 točk
---	-------------

Bonus 1

Petek, 6.3., 9:00: Metka Zaletel (*Nacionalni inštitut za javno zdravje, Zdravstveno podatkovni center*)

Naloga (rok oddaje pred predavanjem na e-mail mojca.bavdaz@ef.uni-lj.si):

- Od treh oseb zbrati podatke za vprašalnik o zdravstveni pismenosti.
- Pripraviti poročilo (max 2 strani) o izkušnji z zbiranjem podatkov (izbor oseb, razmerje do osebe, zadržki glede podajanja podatkov, trajanje, občutki anketarja in anketirancev po izvedbi, ocena kakovosti - popolnosti in resničnosti – podatkov) ter razmislekom, kako na osnovi tovrstnih podatkov oblikovati ukrepe.
- Udeležba na predavanju gosta in sodelovanje v diskusiji.
- Do 10 izpitnih točk (šteje v kvoto 25 minimalno zahtevanih točk).

Bonus 2

Petek, 10.4., 9:00: Urban Brulc (*samostojni svetovalec
Informacijske pooblaščenke RS*)

Naloga (rok oddaje pred predavanjem na e-mail
mojca.bavdaz@ef.uni-lj.si):

- Pripraviti opis petih (po možnosti čimbolj različnih) situacij, v katerih ste se osebno srečali z varstvom osebnih podatkov, skupaj z vašim mnenjem, ali je bilo za (vase) osebne podatke ustrezno poskrbljeno ali ne.
- Priprava treh vprašanj za diskusijo.
- Udeležba na predavanju gosta in sodelovanje v diskusiji.
- Do 7 izpitnih točk (šteje v kvoto 25 minimalno zahtevanih točk).

Sprotno delo

Max 50 točk.

- A. Priprava na uro s področja uradne statistike (5)
- B. Priprava na uro s področja družboslovnih podatkov (5)
- C. Predstavitev primera »big data« (10)
- D. Seminarska naloga (25)
 - Osnutek grafične predstavitve (5)
 - Končna predstavitev z zagovorom (20)
- E. Končni trije aha momenti (5)

A. Priprava na uro s področja uradne statistike (1)

Viri:

Where do you get credible statistics from? (8. julij 2016) Dostopno na <https://www.quora.com/Where-do-you-get-credible-statistics-from>

How reliable are official government statistics in the US? (2015 in 2016). Dostopno na <https://www.quora.com/How-reliable-are-official-government-statistics-in-the-US>

Evropski statistični sistem (2017). Kodeks ravnanja evropske statistike. Brošura dostopna na <https://www.stat.si/statweb/FundamentalPrinciples/CodeOfPractice> ALI <http://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice>

The Economist (26. maj 2018). Plunging response rates to household surveys worry policymakers / Don't even ask! Dostopno na <https://www.economist.com/international/2018/05/24/plunging-response-rates-to-household-surveys-worry-policymakers>

A. Priprava na uro s področja uradne statistike (2)

Naloga:

Napišite eno stran vsebinskega poročila, povzetka (ne copy-paste), ki bo zajel bistvo vseh štirih virov.

Navedite tri »aha momente«.

Navedite eno vprašanje, ki bi ga želeli prediskutirati.

Oddaja:

Format: pdf

Oddajte na spletni strani predmeta v poglavju Sprotno delo

Rok: **sreda, 26.2.**

B. Priprava na uro s področja družboslovnih podatkov (1)

Vira:

Digital Curation Centre. How-to guides & Checklists.

<http://www.dcc.ac.uk/resources/how-guides>

Research Data Alliance: All recommendations & Outputs

<https://www.rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs>

B. Priprava na uro s področja družboslovnih podatkov (2)

Naloga:

Vsak študent povzame en vodič/priporočilo na eni strani
(po predhodni odobritvi, da ni ponavljanja).

Navedite tri »aha momente«.

Navedite eno vprašanje, ki bi ga želeli prediskutirati.

Oddaja:

Format: pdf

Oddajte na spletni strani predmeta v poglavju Sprotno delo

Rok: **sreda, 25.3.**

C. Predstavitev primerov »big data«

1. Cene
2. Facebook in družba
3. Google Flu Trends
4. Netflix
5. Pametne stavbe
6. Pristranskost zaradi kvantifikacije
7. Promet
8. Target



Predstavitev: petek, 20.3.

D. Seminarska naloga

Vir podatkov: SLORA: Slovenija in rak. Epidemiologija in register raka. Onkološki inštitut Ljubljana. www.slora.si

Namen: Razvoj veščin za grafično in vsebinsko predstavitev podatkov

Cilj: Priprava grafične predstavitve za splošno in strokovno javnost vključno s kratkimi vsebinskimi pojasnili in z metodološkimi opombami glede značilnosti in omejitvami podatkov.

Koraki priprave:

1. Oblikovanje skupin.
2. Izbira vsebinskega vidika in določitev obsega naloge ter pridobitev potrditve na mojca.bavdaz@ef.uni-lj.si.
3. Diskusije na predavanjih in individualne konzultacije.
4. Predstavitev.

E. Končni aha momenti

Kaj se vam bo najbolj vtisnilo v spomin pri tem predmetu?

Navedba treh a-ha momentov in argumentacija.

Pregled pomembnih datumov

Datum dogodka ali roka oddaje	Oblika dela	Točke
Rok oddaje: sre, 26.2.	A. Priprava na uro s področja uradne statistike	5
Rok oddaje do predavanj pet, 6.3.	Bonus 1	10 (izpitne)
Predstavitev: pet, 20.3.	C. Predstavitev primera big data	10
Rok oddaje: sre, 25.3.	B. Priprava na uro s področja družboslovnih podatkov	5
Rok oddaje do predavanj pet, 10.4.	Bonus 2	7 (izpitne)
Konzultacije: pet, 10.4.	D. Osnutek grafične predstavitve	5
Predstavitev in rok oddaje: pet, 24.4.	D. Končna predstavitev z zagovorom	20
Rok oddaje: pon, 4.5.	E. Končni aha moment	5

Vrste virov podatkov



I. Način pridobivanja

- Primarni podatki
zbrani za znan namen
field research: (neomejeno) ustvarjaš vprašanja
- Sekundarni podatki
iz obstoječih virov
ponovno uporabljeni
desk research: izbiraš med razpoložljivimi
vprašanji/opazovanji

Ponudniki sekundarnih podatkov

ponujajo primarno zbrane podatke

ponujajo podatke iz sekundarnih virov



Prednosti sekundarnih podatkov

Ker so že zbrani in (če) jih zberejo kompetentni ponudniki:

- prihranki
- ne obremenjujejo poročevalskih enot
- namen raziskave ne vpliva na odgovore poročevalskih enot, torej potencialno boljše razkritje informacij
- večji doseženi vzorci
- dodatne spremenljivke/informacije za triangulacijo
- pogosto edini vir za longitudinalno in mednarodno analizo
- stalnost in periodičnost zbiranja podatkov, predvidljivost objave

Slabosti sekundarnih podatkov

Izvirni »greh«: nezmožnost vplivanja na metodologijo in odsotnost nadzora nad kakovostjo:

- neskladnost s potrebami tokratnega raziskovanja
- neustreznost osnovnih opredelitev koncepta, enote, ciljne populacije itd.
- nezadostna razčlenjenost podatkov
- prepozne objave
- problemi časovnih serij
- neprimerljivost merskih enot

II. Namen uporabe

Kvantitativni podatki (izraženi številsko, numerično)

Kvalitativni podatki (izraženi opisno, nenumerično)

Narava podatka načeloma izhaja iz obravnave in namena uporabe:

Ali želimo pojav opisati ali ugotoviti pogostost?

Vprašanje:

Ali kvantitativni podatki temeljijo na kvalitativni presoji?

Ali lahko kvalitativne podatke spremenimo v kvantitativne?

III. Obseg podatkov

Small data

Big data = 3V

(Gartner. IT Glossary: Big data. Najdeno na <http://www.gartner.com/it-glossary/big-data/>)

high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making and process automation.

+ Veracity etc.



Formati

pdf

surovi podatki

sistemske datoteke raznih programov oz. prenosne datoteke

Ali je format še relevanten na današnji ravni razvoja IT?



Skupinsko delo (1)

Cilj: povečati ozaveščenost o (različnih) virih podatkov

Izvedba: Študenti se razdelijo v dve heterogeni skupini

Naloga 1:

Viharjenje možganov: identificirajte čim več virov podatkov.

Ko boste imeli vsaj 25 virov podatkov, jih poskušajte razvrstiti v skupine ter skupine poimenovati.

Skupinsko delo (2)

Naloga 2:

Vsaka skupina prejme list s tipologijami virov podatkov. Preverite, ali vaša razvrstitev sovпада s katero od podanih tipologij.

Izberite dva čimbolj različna vira podatkov in ju razvrstite v podane tipologije virov podatkov.

Naloga 3:

Identificirajte po eno spremenljivko v vsakem od izbranih dveh virov podatkov in jo opišite z vidika merskih lastnosti in kakovosti.



Primerjajte vaše opise z opisi, podanimi na dodatnem listu.

Citiranje podatkovnih nizov (APA stil)

Skupno ne glede na vrsto vira:

Author/Rightsholder. (Year). *Title of data set* (Version number) [Description of form].

Dodatek glede na vrsto vira:

Fizični vir: Location: Name of producer.

Elektronski vir – naslov vira: Retrieved from <http://>

Elektronski vir – naslov strani: Available from <http://>

Za neobjavljene podatke:

Author, F. N. (Year). [Description of study topic].

Unpublished raw data.

Citiranje podatkovnih nizov (APA stil)

Primeri:

United States Department of Housing and Urban Development. (2008). *Indiana income limits* [Data file]. Retrieved from http://www.huduser.org/Datasets/IL/IL08/in_fy2008.pdf

Pew Hispanic Center. (2008). *2007 Hispanic Healthcare Survey* [Data file and code book]. Available from <http://pewhispanic.org/datasets/>

Smith, J.A. (2006). [Personnel survey]. Unpublished raw data.

U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies. (2013). *Treatment episode data set -- discharges (TEDS-D) -- concatenated, 2006 to 2009* [Data set]. doi:10.3886/ICPSR30122.v2

Citiranje grafičnih podatkov (APA stil)

Primeri:

Centers for Disease Control and Prevention. (2005). [Interactive map showing percentage of respondents reporting "no" to, During the past month, did you participate in any physical activities?]. Behavioral Risk Factor Surveillance System. Retrieved from <http://apps.nccd.cdc.gov/gisbrfss/default.aspx>

Solar Radiation and Climate Experiment. (2007). [Graph illustration the SORCE Spectral Plot May 8, 2008]. Solar Spectral Data Access from the SIM, SOLSTICE, and XPS Instruments. Retrieved from http://lasp.colorado.edu/cgi-bin/ion-p?page=input_data_for_spectra.ion