

STATISTIČNA OBDELAVA PODATKOV

Statistika je znanost zbiranja, organiziranja in interpretiranja numeričnih dejstev, ki jih imenujemo podatki. V vsakodnevno smo s podatki takorekoč bombardirani. Večina ljudi povezuje “statistiko” z biti podatkov, ki izhajajo v dnevnem časopisju, novicah, reportažah: povprečna temperatura na današni dan, procenti pri košarkaških prostih metih, procent tujih vlaganj na našem trgu, in anketa popularnosti predsednika in premierja. Reklame pogosto trdijo, da podatki kažejo na superiornost njihovega produkta. Vse strani v javnih debatah o ekonomiji, izobraževanju in socialni politiki izhajajo iz podatkov. Kljub temu pa uporabnost statistike presega te vsakodnevne primere.

Podatki so pomembni pri delu mnogih, zato je izobraževanje na področju statistike izredno pomembno pri številnih poklicih. Ekonomisti, finančni svetovalci, vodstveni kader, vlade in gospodarstveniki preučujejo najnovejše podatke o nezaposlenosti in inflaciji. Zdravniki morajo razumeti izhodišča in verodostojnost podatkov, ki bodo objavljeni v medicinskih revijah. Industrijske odločitve so običajno zasnovane na raziskavah tržišč, ki govorijo o preferencah kupcev. Večina akademskih raziskav uporablja številke in tako hočeš nočes izkorišča statistične metode.

Nič bolj nismo uspešni pri izogibanju podatkov kot če bežimo od uporabe besed. Tako kot so besede na papirju brez pomena za nepismenega ali slabo izobraženega človeka, tako so lahko tudi podatki privlačni, zavajajoči ali enostavno irelevantni. Statistična pismenost, tj. sposobnost za sledenje in razumevanje argumentov, ki izhajajo iz podatkov, je pomembna za sleherno osebo.

Učbenik. For all practical purposes, *Mathematical Literacy in Today’s World*, Project director: Solomon Garfunkel, Consortium for Mathematics and its Applications, 5. izdaja, Freeman, NewYork, 2000.

Part I: MANAGEMENT SCIENCE

*(J. Malkevitch, York College, CUNY,
R. Mayer, Nassau Community College,
W. Mayer, Adelphi University)*

1. Street Networks,
2. Visiting vertices,
3. Planning and Scheduling,
4. Linear Programming.

Part II: STATISTICS: THE SCIENCE OF DATA

(D.S. Moore, Purdue University)

5. Producing data,
6. Exploring data,
7. Probability: The mathematics of Chance,
8. Statistical Inference.

Part III: CODING INFORMATION

(J. Gallian, University of Minnesota-Daluth)

9. Identification numbers,
10. Transmitting information.

Part IV: SOCIAL CHOICE AND DECISION MAKING

*(S.J. Brams, New York University,
B.P. Conrad, Temple University,
A.D. Taylor, Union College)*

11. Social Choice: The Impossible Dream,
12. Weighted Voting System,
13. Fair Division,
14. Apportionment,
15. Game theory:
The mathematics of Competition.

Part V: ON SIZE AND SHAPE

(P.J. Campbell, Beloit College)

16. Growth and Form,
17. Symmetry and Patterns,
18. Tilings.

Part VI: MODELING IN MATHEMATICS

*(P.J. Campbell, Beloit College,
J. Malkevitch, York College, CUNY,
A.D. Taylor, Union College)*

19. Logic and Modeling,
20. Consumer Finance Models.

STATISTIKA: ZNANOST O PODATKIH

I. poglavje: Zbiranje podatkov

1. Vzorčenje
2. Slabe navade vzorčenja
3. Enostaven naključni vzorec
4. Statistično ocenjevanje
5. Eksperimentiranje
6. Primerjalni eksperimenti z naključno metodo

POD ŽAROMETOM: Sir Ronald A. Fisher (1890-1962)

7. Statistična evidenca (statistična stopnja značilnosti)

8. Statistika v praksi

POD ŽAROMETOM: Charles Hennekens

Slovarček:

eksperiment

enostaven naključni vzorec

histogram

kontrolna skupina

marginna napake (meja napake)

populacija

priročni vzorec

pristranost (angl. bias)

prostovoljni vzorec (na prostovoljni osnovi)

statistična signifikantnost (stopnja značilnosti)

tabela naključnih števil

vzorec

zavajujoč?? (angl. confounding)

Predlagana literatura: G.W. Cobb, Design and analysis of Experiments, 1998.
G. Kalton, Introduction to Sampling, 1983
D.S. Moore, Statistics: Concepts and Contraversies, 1997.

D.S. Moore, The Basic Practice of Statistics: 1999.

J.M. Tanur, Samples and surveys, 1992.

Predlagani viri na internetu:

Kateri so evropski biroji za ankete,

US - the Bureau of the Statistics

(<http://stats.bls.gov>)

<http://www.census.gov>

<http://www.gallup.com>

Journal of the American Medical Association

(www.ama-assn.org/public.journals/jama/)
ali New England Journal of Medicine
(www.nejm.org).

Preverjanje znanja: 7; **Projekti:** 5;
Naloge: 3+4+5+7+7+9+2+3 (dod. 17).

Tehnološki kotiček:

1. generiranje naključnih števil, 2. uporaba histogramov za predstavljanje podatkov, 3. Raziskovanje.

Cilji poglavja:

- Identificiraj populacijo v dani vzorčni ali eksperimentalni situaciji.
- Identificiraj vzorec v dani vzorčni ali eksperimentalni situaciji.
- Pojasni razliko med populacijo in vzorcem.
- Izračunaj delež vzorca za dano velikost vzorca in število pozitivnih odgovorov.
- Določi iz deleža vzorcev in velikosti vzorca število pozitivnih odgovorov.
- Analiziraj vzorečni primer in odkrij izvoro pristranosti.
- Identificiraj številne primere vzorčenja, ki se pojavljajo v naši družbi.
- Izberi shemo za označevanje populacije, iz katere bo izbran naključni vzorec v tabeli z naključnimi števili.
- Uporabi tabelo naključnih števil za izbiro naključnega vzorca iz majhne populacije.
- Pojasni razliko med eksperimentalno skupino in med kontrolno skupino pri eksperimentu.
- Opiši placebo efekt.
- Obravnavaj zakaj je dvojna zaslepljenost zelena pri eksperimentu.
- Naštej dve osnovni lastnosti statistično pripravljenih eksperimentov.
- Prepoznavaj confounding efektov dveh spremenljivk pri eksperimentu.
- Konstruiraj Latinski kvadrat za poenostavitev načrta za za eksperiment.
- Za dan delež populacije in ustrezno mejo za dopustno napako navedi območje za 95% interval zaupanja.

II. poglavje: Raziskovanje podatkov

1. Predstavljanje porazdelitve: histogrami
 2. Interpretiranje histogramov
- POD ŽAROMETOM: W. Edwards Deming
3. Opisovanje centra: povprečje in mediana
 4. Opisovanje razpotegnjenosti: kvartili
 5. Povzetek v obliki peterice in škatle z brki
 6. Opisovanje razpotegnjenosti: stand. odklon
 7. Opisovanje zveze med 2-ma spremenljivkama
 8. Regresijska premica
 9. Korelacija
- POD ŽAROMETOM: Florence Nightngale (1820-1910)
10. Regresija in metoda najmanjših kvadratov
 11. Moderna analiza podatkov

Slovarček:

distribucija/porazdelitev
 histogram
 izstopajoči podatek (lažnjivec)
 korelacija
 kvartil
 mediana
 posameznik
 poševna porazdelitev
 povprečje
 povzetek v obliki peterice: \min , Q_1 , M , Q_3 , \max
 raziskovalna analiza podatkov
 razpršenost
 regresijska premica
 regresijska premica po metodi najmanjših kvadratov
 simetrična porazdelitev
 spremenljivka
 standardni odklon
 škatla z brki
 varianca

Predlagana literatura: W.S. Cleveland, The Elements of Graphing Data, 1985
 D.S. Moore, The Basic Practice of Statistics: 1999.
 A.J. Rossman, Workshop Statistics: Discovery with Data, 1996.
 E.R. Tufte, The visual Display of Quantitative information, 1983.

Predlagani viri na internetu:

The Datas and Story Library:
lib.stat.cmu.edu/DASL/,
www.dartmouth.edu/~chance,
www.amstat.org

Preverjanje znanja: 7; **Projekti:** 4;
Naloge: 6+12+2+4+12 (dod. 15).

Tehnološki kotichek:

1. Računanje povprečja in standardnega odklona, 2. Računanje premice najmanjših kvadratov, 3. Raziskovanje.

Cilji poglavja:

- Izračunaj povprečje za množico podatkov.
- Po sortiranju množice podatkov določi mediano.
- Določi spodnji in zgornji kvartil za dano množico podatkov.
- Izračunaj območje za dano množico podatkov, potem ko najdeš največjega in najmanjšega med njimi.
- Navedi pet-elementni povzetek za dano množico podatkov.
- Konstruiraj diagram škatel z brki za dan pet-elementni povzetek.
- Konstruiraj histogram za majhno množico podatkov.
- Identificiraj iz histograma možne izstopajoče elemente za dano množico podatkov.
- Navedi in opiši dva tipa porazdelitev za histogram.
- Nariši scatterplot? za majhno množico podatkov, ki se sestojijo iz parov podatkov.
- Iz scatterplot? nariši ocenjeno premico (From a scatterplot, draw an estimated line fit).
- Razloži kako je koncept razdalje uporabljen za določanje regresijske premice s metodo najmanjših kvadratov.
- Uporabi dano enačbi regresijske premice za napovedovanje (Y) vrednosti za dane causal? (X) vrednosti.
- Izračunaj korelacijo med dvema kvantitativnima spremenljivkama.
- Razumeti statistično značilnost korelacije med dvema spremenljivkama in jo oceniti iz scatterplot.

III. poglavje: Verjetnost: matematika šans

“It is a remarkable fact that the aggregate result of many thousands of chance outcomes can be known with near certainty.”

1. Kaj je verjetnost
2. Modeli verjetnosti in pravila
3. Enako verjetni dogodki
4. Povprečje verjetnostnega modela
5. Porazdelitev vzorcev
6. Normalna porazdelitev
7. Centralni limitni izrek

Slovarček:

centralni limitni izrek
dogodek
enako verjetni izidi
fenomen naključja
kombinatorika
nezdružljivi dogodki
normalna porazdelitev
pravilo za vsoto za nezdružljive dogotke
prostor vzorcev
sistem enostavnih dogodkov
standardni odklon normalne krivulje
standardni odklon verjetnostnega modela
statistika
verjetnost
verjetnostni histogram
verjetnostni model
vzorčna porazdelitev
zakon velikih števil

Predlagana literatura:

F. Mosteller, R.E.K. Rouke in G.B.Thomas, Probability with Statistical Applications, 1970.

I. Olkin, L.J. Gleser and S. Derman, Probability Models and Applications, 1994.

L.J. Snell, Introduction to Probability, 1998.

Predlagani viri na internetu:

www.ruf.rice.edu/~lane/hyperstat/index

Preverjanje znanja: 7; **Projekti:** 3;

Naloge: 6+7+6+6+2+7+16 (dod. 16).

Tehnološki kotiček:

1. Designing a Game of Chance,

2. Simulating the Game with a Spreadsheet,
3. Raziskovanje

Cilji poglavja:

- Opiši vzorčni prostor za dan naključni fenomen.
- Pojasni na kaj mislimo z izrazom verjetnost dogodka.
- Naštej dva zakona, ki jih zadovoljuje verjetnost.
- Uporabi zakone verjetnosti za določitev veljavnosti verjetnostnega modela.
- Identificiraj kateri verjetnostni zakon ni zadovoljen za nelegitimen verjetnostni model.
- Skonstruiraj drevesni diagram za analizo enostavnega binomskega verjetnostnega primera.
- Za dan drevesni diagram naštej vzorčni prostor in priredi verjetnosti za dogodke, tako da bodo veljali verjetnostni zakoni.
- Izračunaj verjetnost dogodka za verjetnostni model danega eksperimenta.
- Opiši normalno krivuljo.
- Določi povprečje in standardni odklon iz grafa normalne krivulje.
- Pojasni pravilo 68-95-99,7
- Uporabi pravilo 68-95-99,7 za računanje normalnih verjetnosti.
- Given the mean and standard deviation of a normally distributed data set, compute the percent of the population that falls within a given interval.
- Compute the expected value of an outcome when the associated probability model is defined.
- Pojasni pomen centralnega limitnega izreka.
- Uporabi pravilo za vsoto dveh dogodkov za računanje verjetnosti kombinacije večih disjunktih dogotkov.
- Nariši verjetnostni histogram verjetnostnega modela in ga nato uporabi za določanje verjetnosti slučajnih dogodkov.

IV. poglavje: Statistična inferenčnost

1. Ocenjevanje deleža populacije
2. Intervali zaupanja
3. Ocenjevanje populacijskega povprečja
POD ŽAROMETOM: How the pool was taken
4. Statistična kontrola procesa
5. Perils of analize podatkov

Slovarček:

delež vzorca
interval zaupanja
kontingenčna tabela
kontrolna razporednica
parameter
povprečje vzorca
statistika

Predlagana literatura: David S. Moore, The Basic Practice of Statistics: 1999.
Lincoln E. Moses, The reasoning statistical interference, 1962.

Predlagani viri na internetu: the Gallup Organization (<http://www.gallup.com>)
(<http://www.louisharris.com>)
stats.bls.gov
www.ncpp.org

Preverjanje znanja: 7; **Projekti:** 3;
Naloge: 7+11+10+17+6 (dod. 15).

Tehnološki kotichek: testiranja generatorja psevdonaključnih števil v razporednici ocenjevanje μ z uporabo \bar{x} .

Cilji poglavja:

- Definiraj statistično inferenco.
- Opiši razliko med neformalno in formalno statistično inferenco.
- Pojasni razliko med parametrom in statistiko.
- Identificiraj parameter in statistiko pri enostavni inferenčni nastavitvi.
- Izračunaj vzorčni delež za dano velikost vzorca in dano število pozitivnih odgovorov.
- Uporabi ustrezno formulo za izračun standardnega odklona za dano statistiko.
- Izračunaj mejo dopustne napake za enostavno statistiko, če poznaš standardni odklon.
- Izračunaj razlike mej napak za povečane vzorčne velikosti.
- Obravnava učinek povečane velikosti vzorca za statistično mejo napake.
- Za dan standardni odklon enostavne statistike konstruiraj 95% interval zaupanja za delež populacije.
- Pojasni razliko med povprečjem populacije in med vzorčnim povprečjem.
- Obravnava uporabo procesne kontrolne charts v industrijskem okolju.
- Identificiraj centralno premico in kontrolne limite za dan procesni kontrolni chart.