Poglavje 4

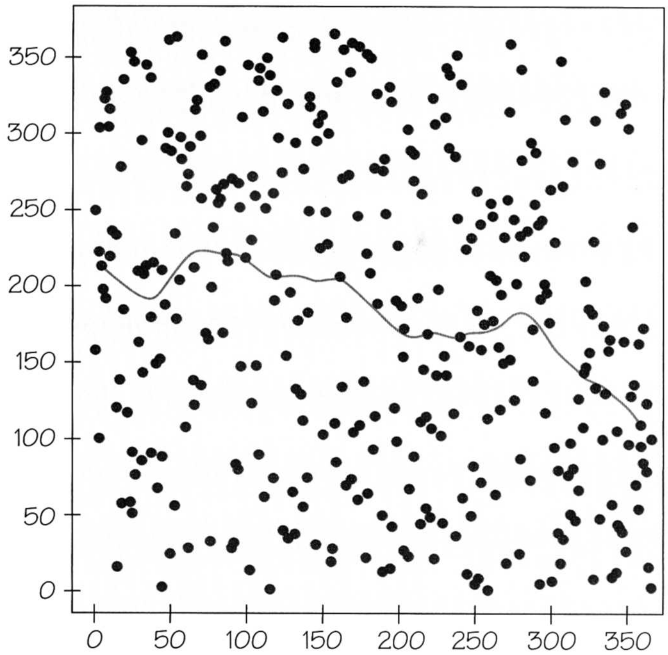
Statistično sklepanje

Sklepanje je proces, pri katerem pridemo do zaključkov na podlagi danih dokazov. Dokazi so lahko v mnogo različnih oblikah. V sojenju zaradi umora jih lahko predstavljajo izjave prič, posnetki telefonskih pogovorov, analize DNK iz vzorcev krvi in podobno. Pri statističnem sklepanju nam dokaze priskrbijo podatki. Neformalno statistično sklepanje velikokrat temelji na grafični predstavitvi podatkov. Formalno sklepanje, tema tega poglavja, uporablja verjetnost, da pove, do kakšne mere smo lahko prepričani, da so naši zaključki pravilni.

Primer. (Je bil nabor nepošten?) V času Vietnamske vojne so američani izbirali državljane za vojaško obveznost po sistemu loterije. Prvo so izvedli leta 1970. Vsi mediji so kmalu poročali, da je bila loterija pristranska proti tistim, ki so bili rojeni kasneje v letu. Na sliki 4.1 je razsevni diagram izbranih števil (manjša števila so bila izbrana prej) v odvisnosti od datuma rojstva, štetega od 1. januarja naprej. Krivuljo na diagramu smo dobili z izglajevalcem razsevnega diagrama. Vidimo, da kasneje v letu teži k nižjim številkam. Ta tendenca ni zelo močna, zato se lahko vprašamo, če je morda tak izid le naključen, ne pa rezultat sistematične pristranskosti v loteriji. Navsezadnje je pri vsaki loteriji nekaj slučajnega odstopanja od popolne poenotenosti.

Vendar pa izračun verjetnosti pokaže, da se tako močna odstopanja, kot je tisto na sliki 4.1, v zares slučajnih izborih pojavijo z verjetnostjo, manjšo od $\frac{1}{1000}$. Ta izračun je vse prepričal, da nabor ni bil naključen. Preiskava je pokazala, da so pri žrebanju slabo premešali kapsule, ki so vsebovale datume. ♦

V matematiki pridemo do zaključkov tako, da začnemo s hipotezo, nato pa z uporabo logičnega sklepanja pokažemo, da naši zaključki brezdvomno sledijo iz predpostavk.



Slika 4.1: Razsevni diagram števil, izbranih pri naboru (med 1 in 366) v odvisnosti od datuma rojstva (z začetkom pri 1. januarju). Splošni potek je opisan s pomočjo izglajevalca diagrama. Kaže, da je bila loterija pristranska v prid moškim, ki so bili rojeni v začetnem delu leta.

To je *deduktivno sklepanje* od hipoteze k posledicam. Statistični dokazi potekajo skoraj obratno. Če je bila loterija pristranska, pričakujemo, da bodo kasnejši datumi sistematično manj pogosto izbrani. V podatkih opazimo tak trend, zato dokazi govorijo v prid trditvi, da je bila loterija pristranska. To je *induktivno sklepanje* od posledic k hipotezi. Induktivno sklepanje ni dokaz. Manjša pogostost kasnejših datumov bi *lahko* bila naključna. *Statistično sklepanje uporabi verjetnost, da nam pove, kako močan je induktivni argument.* Težnja, ki jo opazimo v podatkih za leto 1970 se skoraj nikoli ne bi pojavila (verjetnost je manjša od 0,001) v slučajnem izboru. To močno govori v prid trditvi, da loterija ni bila zares naključna.

Primer. (Ali lahko zaupamo javnomnenjski raziskavi?) Kako lahko zaupamo rezultatom, ki jih dobimo s slučajnim vzorčenjem, če vemo, da bi z drugim vzorcem dobili drugačen rezultat? Pri Gallupovi raziskavi na 1493 ljudeh so ugotovili, da se 45% ljudi zaradi kriminala boji ponoči ven. V drugem slučajnem vzorcu bi izbrali nekih drugih 1493 ljudi in rezultat, ki bi ga dobili, bi bil različen od 45%. Kaj lahko kljub temu povemo o populaciji več kot 200 milijonov odraslih Američanov na osnovi Gallupovega vzorca?

Izračun verjetnosti nam pove, da pri 95% vseh Gallupovih vzorcev dobimo rezultat, ki se za manj kot tri odstotne točke razlikuje od resnične vrednosti za našo populacijo. Spoznamo torej, da je precej varno verjeti, da leži dejanska vrednost med 42% in 48%. Ker pa Gallupovi vzorci večkrat zgrešijo pravi rezultat za več kot 1%, ne moremo trditi, da leži resnični delež med 44% in 46%. ♦

V obeh primerih smo odgovor na vprašanje "Kaj bi se zgodilo, če bi to napravili večkrat?" dobili z izračunom verjetnosti. V velikem številu nepristranskih loterij bi le ena izmed tisoč pokazala tako močan trend, kot smo ga lahko opazili pri naboru. V velikem številu Gallupovih vzorcev bi jih 95% dalo rezultat, ki bi ležal $\pm 3\%$ od dejanske vrednosti. Takšne izjave so značilne za statistično sklepanje. Da bi ga razumeli, moramo najprej razumeti, kako pri tem uporabljamo verjetnost.

4.1 Ocenjevanje deleža populacije

Z uporabo poenostavljene verzije Gallupove raziskave o kriminalu bomo predstavili pomembno obliko statističnega sklepanja. Kot večina nacionalnih raziskav uporablja Gallupova anketa kompleksen večstopenjski načrt vzorca. Namesto tega si predstavljajmo, da izberemo enostavni slučajni vzorec 1500 odraslih in ugotovimo, da se jih med njimi 675 boji kriminala. **Delež vzorca**, ki ostane doma zaradi strahu pred kriminalom je enak

$$\hat{p} = \frac{675}{1500} = 0,45 = 45\%.$$

Delež vzorca bomo označili s $\hat{p}$ (kar preberemo "p streha"). Vedno bomo izražali delež vzorca v odstotkih.

Delež vzorca $\hat{p} = 45\%$ se nanaša na 1500 ljudi iz tega konkretnega vzorca. V resnici pa želimo poznati *delež populacije*, odstotek (imenujmo ga p) vseh odraslih Američanov, ki ponoči ostajajo doma zaradi strahu pred kriminalom. Da bi lahko razumno razpravljali o statističnem sklepanju, moramo jasno vedeti, katera števila opisujejo vzorce in katera populacije.

Število, ki opisuje populacijo, imenujemo **parameter**. Število, ki ga izračunamo iz vzorca, imenujemo **statistika**.

Ni si težko zapomniti, da **p**arameter pripada **p**opulaciji, statistika pa torej spada k vzorcu[1]. Pri statističnem sklepanju so parametri običajno neznani. Ne vemo, kolikšen delež p vseh odraslih zares ostaja doma zaradi strahu pred kriminalom. Za oceno neznanega parametra p uporabimo statistiko $\hat{p}$, ki smo jo dobili, ko smo dejansko anketirali ljudi iz vzorca. *Naš cilj ni zgolj oceniti p, temveč tudi povedati, kako natančna je ta ocena.* Da bi to izvedeli, se moramo vprašati "Kaj bi se zgodilo, če bi vzeli več vzorcev? Kako blizu neznanemu p bi običajno ležale ocene $\hat{p}$?"

Za odgovor na to vprašanje si natančneje ogledamo *porazdelitev vzorca $\hat{p}$.* To je porazdelitev vrednosti, ki jih zavzamejo deleži vzorcev, ko si ogledamo veliko število vzorcev iz iste populacije. V prejšnjih poglavjih smo že simulirali vzorčne porazdelitve. Zdaj želimo matematična dejstva. Tule so:

Izberimo enostavni slučajni vzorec velikosti n iz velike populacije, v kateri ima p% osebkov lastnost, ki nas zanima. Naj bo $\hat{p}$ delež osebkov v vzorcu, ki imajo to lastnost. Potem velja:

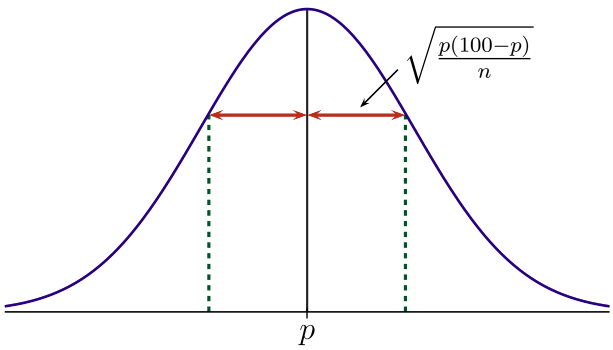
- Vzorčna porazdelitev $\hat{p}$ je *približno normalna* in je bližje normalni kadar je velikost vzorca n velika.
- *Srednja vrednost* vzorčne porazdelitve je natanko p.
- *Standardni odklon* vzorčne porazdelitve je
$$\sigma_{\hat{p}} = \sqrt{\frac{p(100-p)}{n}}.$$
S $\sigma_{\hat{p}}$ jo označimo zato, da ne pozabimo, da gre za standardni odklon, ki pripada porazdelitvi $\hat{p}$.

Na sliki 4.2 je prikazana ta vzorčna porazdelitev kot normalna krivulja. Središče (srednja vrednost) in razpon (standardni odklon) te krivulje vsebujeta pomembne statistične informacije.

Srednja vrednost krivulje je dejanski delež p vseh ljudi, ki se bojijo ponoči zapustiti domove zaradi kriminala. To nam pove, da $\hat{p}$ ni pristranska ali sistematično napačna cenilka za neznani p. Pri večkratnem vzorčenju bo dobljeni rezultat včasih višji in včasih nižji, povprečje velikega števila rezultatov, srednja vrednost vzorčne porazdelitve, pa bo pravilno. V praksi seveda ne poznamo vrednosti parametra p.

[1] V angleščini se ujemata tudi prvi črki izrazov za statistiko (statistic) in vzorec (sample). (Op. prev.)

Vemo pa, da se ne glede na to, kakšno vrednost ima p, opažene vrednosti statistike $\hat{p}$ kopičijo okoli p kot na sliki 4.2.



Slika 4.2: Vzorčna porazdelitev deleža vzorca $\hat{p}$. Je približno normalna s srednjo vrednostjo p in standardnim odklonom $\sqrt{\frac{p(100-p)}{n}}$.

Vendar pa nismo zadovoljni s tem, da imamo v povprečju prav. Dobra cenilka mora biti tudi ponovljiva, v smislu da daje skoraj enake odgovore pri večkratnem vzorčenju. Ponovljivost opišemo z razponom vzorčne porazdelitve, ki ga meri standardni odklon. Če bi vzorčenje ponovili velikokrat, tako da bi v vsaki ponovitvi poklicali naključno izbrane številke, bi vsakič dobili vrednost deleža vzorca $\hat{p}$ nekje vzdolž krivulje na sliki 4.2. Kako daleč od resničnega p ti rezultati ležijo, je odvisno od standardnega odklona $\sigma_{\hat{p}}$ te normalne krivulje. Standardni odklon se manjša, ko se velikost vzorca n povečuje. Tudi pri naših simulacijah iz prejšnjega poglavja (slika 3.6) je bilo tako. Zdaj natančno poznamo zvezo med n in standardnim odklonom. *Standardni odklon je odvisen od kvadratnega korena* $\sqrt{n}$. Če želimo razpon porazdelitve razpoloviti, moramo izbrati štirikrat večji vzorec.

Primer. (Vzorčna porazdelitev za raziskavo o kriminalu) Recimo, da se v resnici 40% odraslih boji ponoči zdoma zaradi kriminala. Se pravi, naj bo $p = 40\%$. Izberimo enostavni slučajni vzorec velikosti $n = 1500$. Pri večkratnem vzorčenju se bo delež vzorca $\hat{p}$ spreminjal v skladu z normalno porazdelitvijo s srednjo vrednostjo $p = 40\%$ in s standardnim odklonom

$$\sigma_{\hat{p}} = \sqrt{\frac{p(100-p)}{n}} = \sqrt{\frac{40 \cdot 60}{1500}} = \sqrt{1{,}6} = 1,265\%.$$

V praksi vrednosti parametra p ne poznamo. S kalkulatorjem izračunamo, da je standardni odklon porazdelitve $\hat{p}$ majhen, zato bo delež vzorca $\hat{p}$ ponavadi precej blizu p.

Zdaj pa predpostavimo, da je resnična vrednost za populacijo $p = 50\%$. Središče vzorčne porazdelitve se premakne na 50%. Standardni odklon se spremeni v

$$\sigma_{\hat{p}} = \sqrt{\frac{p(100-p)}{n}} = \sqrt{\frac{50 \cdot 50}{1500}} = \sqrt{1{,}67} = 1,29\%.$$

Standardni odklon $\sigma_{\hat{p}}$ se ne spremeni veliko, ko se spremeni p. Ko torej izbiramo vzorce istih velikosti iz različnih populacij, se središče vzorčne porazdelitve $\hat{p}$ premakne do parametra p, ki ustreza tej populaciji, razpon pa ostaja približno enak.

Velikost vzorca pomembno vpliva na razpon. Recimo, da bi izbrali vzorec velikosti $n = 375$ namesto 1500 iz populacije, za katero je $p = 40\%$. Srednja vrednost porazdelitve $\hat{p}$ je še vedno 40%, ker velikost vzorca ne spremeni središča vzorčne porazdelitve. Standardni odklon pa se poveča na

$$\sigma_{\hat{p}} = \sqrt{\frac{p(100-p)}{n}} = \sqrt{\frac{40 \cdot 60}{375}} = \sqrt{6{,}4} = 2,53\%.$$

Ker je nova velikost vzorca $n = 375$ le četrtina prvotnih 1500, je nov standardni odklon 2,53% dvakrat tolikšen kot prejšnji 1,265%. Tako na razpon vpliva $\sqrt{n}$. ♦

4.2 Intervali zaupanja

Pri našem anketiranju 1500 ljudi smo ugotovili, da je $\hat{p} = 45\%$. To je naša najboljša ocena za delež populacije p. Kako blizu resnične vrednosti p je najbrž naša ocena? Vemo, da se $\hat{p}$ spreminja v skladu z normalno porazdelitvijo. Pravilo 68–95–99,7 pravi, da $\hat{p}$ v 95% vseh vzorcev leži največ dva standardna odklona od p (srednje vrednosti vzorčne porazdelitve). Naša ocena, ki temelji na tem enem vzorcu, je torej najverjetneje od p oddaljena največ dva standardna odklona, se pravi kvečjemu

$$2\sigma_{\hat{p}} = 2\sqrt{\frac{p(100-p)}{1500}}.$$

Problem je v tem, da je ta standardni odklon odvisen od neznanega parametra p. Na srečo pa smo v zgornjem primeru videli, da se $\sigma_{\hat{p}}$ spreminja le zelo počasi, ko se spreminja p, če le ni p zelo blizu 0% ali pa 100%. Ker je $\hat{p}$ blizu p, enostavno zamenjamo v enačbi za standardni odklon neznani p s $\hat{p}$. Ker je tako dobljeni standardni odklon le ocena in ne poznamo natančne vrednosti, ga označimo s $s_{\hat{p}}$.

Primer. (Ocenjeni standardni odklon za raziskavo o kriminalu) Oceniti želimo standardni odklon deležev vzorca. Velikost vzorca je bila $n = 1500$ in za p

uporabimo oceno $\hat{p} = 45\%$, ki jo dobimo z anketo. Ocenjeni standardni odklon je

$$s_{\hat{p}} = \sqrt{\frac{45 \cdot 55}{1500}} = \sqrt{1{,}65} = 1,285\%.$$

♦

Zaključimo torej naslednje: v 95% vseh vzorcev bo delež vzorcev $\hat{p}$ kvečjemu za $2 \cdot 1,285\% = 2,6\%$ oddaljen od neznanega deleža populacije p. Izbrali smo en vzorec in dobili $\hat{p} = 45\%$. Od tod sklepamo, da p leži na intervalu

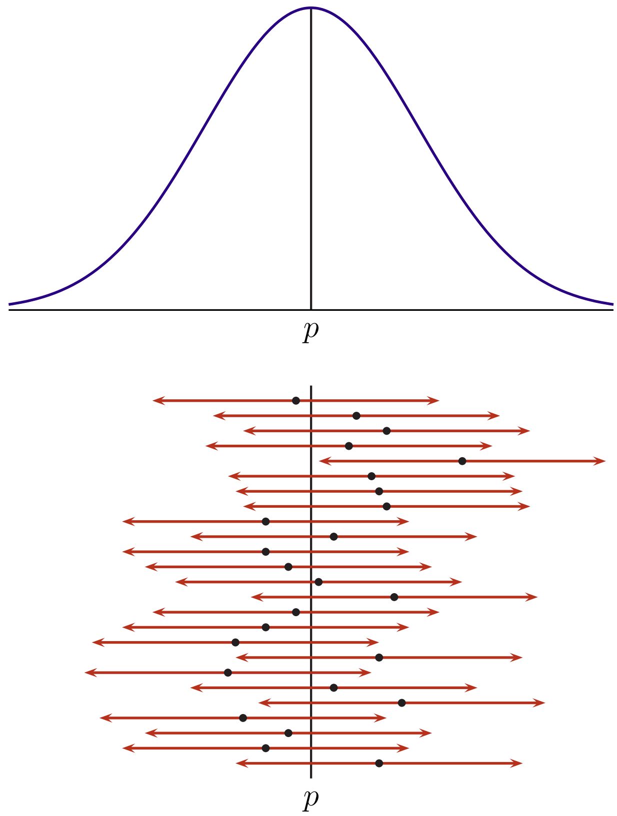
$$45\% \pm 2,6\%$$

oziroma med 42,4% in 47,6%. Pravimo, da je naše *zaupanje* v ta sklep 95%, ker smo dobili interval tako, da smo izračunali, kako blizu p ležijo deleži vzorcev v 95% vseh vzorcev. To je *95% interval zaupanja* za oceno neznanega deleža populacije.

Povedano po matematično, z verjetnostjo 0,95 je delež vzorca $\hat{p}$ oddaljen kvečjemu za ±2,6% od neznanega dejanskega deleža p vseh odraslih, ki jih je ponoči strah zapustiti domove zaradi kriminala. Slika 4.3 dodatno pojasnjuje to idejo. Normalna krivulja na zgornjem delu slike je vzorčna porazdelitev za $\hat{p}$. Ko izberemo več različnih vzorcev, se dejanske vrednosti $\hat{p}$ spreminjajo v skladu s to krivuljo. Na spodnjem delu slike so s pikami predstavljene vrednosti $\hat{p}$, ki smo jih dobili s 25 različnimi vzorci. Pri vsaki je narisan tudi interval zaupanja, ki sega 2,6% na vsako stran ustreznega $\hat{p}$. Dejanski delež populacije p je označen z navpično črto. Čeprav se intervali od vzorca do vzorca spreminjajo, vsi razen enega vsebujejo pravi p. Ko pravimo, da gre za 95% intervale zaupanja, želimo le povedati, da ti intervali vsebujejo p v 95% vseh vzorcev in ga zgrešijo le v 5% primerov. Pri tem je pomembno razumeti, da se teh 95% in 5% nanaša le na to, kaj bi se zgodilo, če bi izbirali take vzorce vedno znova in znova. Pri majhnem številu vzorcev je lahko število intervalov zaupanja, ki ne vsebujejo dejanskega p, nekoliko večje ali manjše od 5%. Na sliki 4.3 na primer le pri enem vzorcu od 25, torej v 4%, interval zaupanja ne vsebuje p.

Interval, ki ga dobimo iz podatkov o vzorcu z metodo, pri kateri za 95% vseh vzorcev dobimo interval, ki vsebuje dejanski parameter p dane populacije, se imenuje **95% interval zaupanja**.

Na sliki 4.3 lahko vidimo, da interval zaupanja vsakega vzorca vsebuje neznani parameter ali pa ne. Ne moremo vedeti, če je naš izbrani vzorec eden izmed 95% tistih, ki zadenejo parameter, ali pa med tistimi 5%, ki ga zgrešijo. Če za naš interval



Slika 4.3: Obnašanje 95% intervalov zaupanja pri večkratnem vzorčenju. Interval se spreminja od vzorca do vzorca, vendar pa na dolgi rok 95% vseh vzorcev porodi intervale, ki vsebujejo dejansko vrednost p.

45%±2,6% trdimo, da je 95% interval zaupanja, želimo reči "Ta interval smo dobili z metodo, ki v 95% primerov zadene pravi parameter."

Zdaj smo dosegli dvoje: ugotovili smo, kaj pomeni "95% zaupanje", poleg tega pa smo dejansko našli 95% interval zaupanja za ocenjevanje deleža populacije. Tule je recept:

Za enostavni slučajni vzorec velikosti n je **95% interval zaupanja za delež populacije** p enak

$$\hat{p} \pm 2s_{\hat{p}} = \hat{p} \pm 2\sqrt{\frac{\hat{p}(100-\hat{p})}{n}}.$$

Spomnimo se, da merimo tako p kot $\hat{p}$ v odstotkih. Ta recept je samo približno pravilen, vendar pa je precej natančen, ko je velikost vzorca n velika.

Primer. (Tvegano obnašanje v dobi AIDSa) Kako pogosto je obnašanje, pri katerem ljudje tvegajo okužbo? V nacionalni raziskavi so anketirali slučajni vzorec 2673 odraslih heteroseksualcev. Od teh jih je 170 v preteklem letu imelo več kot enega partnerja. Ustrezni delež vzorca je

$$\hat{p} = \frac{170}{2673} = 6,36\%.$$

S 95% stopnjo zaupanja je interval zaupanja za delež populacije p enak

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(100-\hat{p})}{n}} = 6,36 \pm 2\sqrt{\frac{6{,}36 \cdot 93{,}64}{2673}} = 6,36 \pm 0,94 = 5,42\% \text{ do } 7,30\%.$$

Kot ponavadi praktične težave lahko povzočijo dodatne napake, ki jih *ne* upoštevamo pri meji napake. Zelo verjetno je, da nekateri sodelujoči niso bili pripravljeni razkriti svojih resničnih navad. Rezultati raziskave torej najbrž podcenjujejo pogostost tveganega obnašanja, ne vemo pa, za koliko. ♦

Dolžina intervala zaupanja je odvisna od velikosti vzorca n: pri večjih vzorcih so intervali krajši. Niso pa intervali odvisni od velikosti populacije. To velja vse dokler je populacija veliko večja kot vzorec. Interval zaupanja iz zgornjega primera velja tako za vzorec 2673 ljudi iz mesta s 100 000 odraslimi prebivalci kot za vzorec iz države z 200 milijoni prebivalcev. Tisto, kar je pomembno, je, koliko ljudi anketiramo, ne pa kolikšen delež populacije kontaktiramo.

Vsak interval zaupanja ima dva bistvena podatka: interval sam in stopnjo zaupanja. Običajno ima obliko

$$\text{ocena } \pm \text{ meja napake}.$$

Ocena je vzorčna statistika, na primer $\hat{p}$, ki ocenjuje neznani parameter. Meja napake nam pove, kako natančna je ta ocena. V raziskavi o AIDSu je ocena 3,36%, meja napake pa je 0,94%.

Stopnja zaupanja nam pove, kako prepričani smo, da naš interval vsebuje dejansko vrednost parametra. Čeprav je pogosta stopnja zaupanja 95%, lahko zahtevamo večje zaupanje, na primer 99%, ali pa se zadovoljimo že samo z 90%. Naš 95% interval zaupanja je temeljil na srednjih 95% normalne porazdelitve. Za 99% stopnjo zaupanja zahtevamo srednjih 99% porazdelitve, zato je ustrezni interval zaupanja širši (ima večjo mejo napake). Podobno je 90% interval zaupanja krajši od 95% intervala zaupanja za iste podatke. Izbirati moramo med tem, kako natančno lahko določimo parameter (meja napake) in kako prepričani smo v ta rezultat.

Primer. (Razumevanje novic) V novicah se pogosto pojavljajo rezultati javnomnenjskih raziskav in drugih raziskav na vzorcih. Poročila velikokrat navajajo mejo napake, redko pa omenijo stopnjo zaupanja. (O izjemi lahko prebereš v okvirju.) Medijsko poročilo o naši raziskavi o kriminalu bi izgledalo takole: "Raziskava je pokazala, da se 45% Američanov zaradi kriminala boji ponoči zapustiti domove. Meja napake te raziskave je ± 2,6 odstotne točke."

Poznati pa moramo tako mejo napake kot stopnjo zaupanja, ker večja stopnja zaupanja zahteva večjo mejo napake. V medijih velja nenapisano pravilo: skoraj vse javnomnenjske raziskave podajo mejo napake pri 95% stopnji zaupanja. Če torej v novici o raziskavi podajo mejo napake brez omembe stopnje zaupanja, lahko ponavadi predpostavimo, da je ta enaka 95%. ♦

Pod žarometom

Kako je potekala raziskava?

New York Times je 15. junija 1998 objavil članek z naslovom "Raziskava kaže: večina naklonjena Microsoftu", ki sta ga napisala Steve Lohr in Marjories Connelly. V času ko je država napadala računalniškega giganta z antitrust zakoni, je raziskava javnega mnenja pokazala, da ima 55% odraslih o Microsoftu pozitivno mnenje. Časopis je ob strani objavil tudi nekaj podrobnosti o tem, kako je potekala raziskava:

> Najnovejša raziskava, ki sta jo izvedla *New York Times* in *CBS News*, je temeljila na telefonski anketi, ki je bila opravljena med 7. in 9. junijem in je zajela 1126 odraslih po vseh ZDA.
> Vzorec telefonskih central je bil izbran slučajno s pomočjo računalnika s popolnega seznama več kot 42 000 aktivnih central po celi državi. Znotraj vsake centrale so nato slučajno izbrali telefonske številke, tako da so bili v vzorec vključeni tudi tisti, ki sicer niso v telefonskem imeniku. V vsakem gospodinjstvu je nato na vprašanja odgovarjala slučajno izbrana odrasla oseba.
> ... Teoretično se v 19 primerih od 20 rezultati raziskav, ki temeljijo na tovrstnih vzorcih, razlikujejo od dejanske vrednosti za največ tri odstotne točke.
> ... Poleg napak, ki nastajajo pri vzorčenju, se lahko pojavijo še druge, ki so povezane s praktičnimi omejitvami pri izvajanju vseh javnomnenjskih raziskav. Odstopanja v formulaciji vprašanj ali v vrstnem redu lahko na primer pripeljejo do drugačnih rezultatov.

Urad za delavsko statistiko pa se je odločil, da bo mesečno stopnjo nezaposlenosti podajal z 90% stopnjo zaupanja. Zaključki so temeljili na raziskavi, ki je vključevala

50 000 gospodinjstev, Urad pa pravi, da je meja napake $\pm 0,2\%$. Medtem ko naslovi časopisov oznanjajo 5,9% stopnjo nezaposlenosti, želi Urad povedati, da je z 90% zaupanjem med 5,7% in 6,1% delovno aktivnega prebivalstva nezaposlenega.

Raziskave mnenja imajo velikokrat meje napake okoli $\pm 3\%$. Bistveno manjša meja napake pri raziskavi o nezaposlenosti je posledica veliko večjega vzorca. Pri večjih vzorcih je meja napake manjša pri isti stopnji zaupanja. Vendar pa kvadratni koren iz n, ki se pojavi v računih, pove, da moramo vzeti štirikrat večji vzorec, če želimo mejo napake razpoloviti. Da bi dobili zelo majhno mejo napake, se pri zgornji raziskavi potrudijo in anketirajo 50 000 ljudi, medtem ko jih Gallupova raziskava kontaktira le 1500. Pri Gallupovi raziskavi si lahko privoščijo 3% napako, stopnja nezaposlenosti pa mora biti bolj natančno določena, ker na njej temelji veliko ekonomskih in političnih odločitev.

4.3 Ocenjevanje srednje vrednosti populacije

Statistikova škatla z orodjem vsebuje veliko različnih intervalov zaupanja, ki ustrezajo velikemu številu različnih parametrov populacije, ki bi jih želeli oceniti. Spoznali smo že interval zaupanja za oceno deleža populacije p. Zdaj pa želimo oceniti srednjo vrednost populacije. Za opis središča množice podatkov smo uporabili **vzorčno povprečje** $\overline{x}$. Zdaj bomo uporabili vzorčno povprečje $\overline{x}$ za oceno neznane srednje vrednosti μ celotne populacije, iz katere smo izbrali vzorec. Srednjo vrednost populacije označimo z μ, ki označuje tudi srednjo vrednost verjetnostne porazdelitve, ker je srednja vrednost populacije tudi srednja vrednost porazdelitve rezultata, ki ga dobimo pri slučajni izbiri enega posameznika iz populacije. Vzorčno povprečje $\overline{x}$ je statistika, ki se bo med večkratnim vzorčenjem spreminjala, medtem ko je srednja vrednost populacije μ parameter in ostaja konstantna. Na srečo je interval zaupanja za ocenjevanje μ precej podoben že znanemu intervalu zaupanja za ocenjevanje p, ker oba temeljita na normalni vzorčni porazdelitvi.

Primer. (NAEP) Namen ameriške raziskave NAEP (*National Assessment of Educational Progress*) je ugotavljati, kako napreduje izobraženost. Raziskava vključuje tudi kratek test kvantitativnega razumevanja, ki preverja v glavnem osnovno računanje in sposobnosti uporabe le-tega v realističnih problemih. Rezultati testov ležijo med 0 in 500 točkami. Oseba, ki dobi 233 točk, lahko na primer sešteje dva zneska, ki se pojavita na bančnem izpisku, nekdo, ki je dobil 325 točk je sposoben iz menija razbrati, koliko bo stalo kosilo, oseba s 375 točkami lahko pretvori ceno iz centov na gram v dolarje na kilogram.

V zadnjem letu so v vzorec zajeli 840 moških med 21. in 25. letom starosti. Njihov povprečen rezultat je bil $\overline{x} = 272$. Teh 840 moških sestavlja enostavni slučajni vzorec populacije vseh mladih moških. Kaj lahko na osnovi tega vzorca povemo o povprečnem rezultatu μ populacije vseh 9,5 milijona moških med 21. in 25. letom starosti? ♦

Zakon velikih števil nam pove, da je vzorčno povprečje $\overline{x}$ pri velikih slučajnih vzorcih blizu neznani srednji vrednosti μ za dano populacijo. Ker je $\overline{x} = 272$, ugibamo, da je μ "nekje blizu 272". Da bi "nekje blizu 272" formulirali bolj natančno, se vprašamo: "Kako bi se spreminjalo vzorčno povprečje $\overline{x}$, če bi izbrali veliko vzorcev 840 oseb iz te populacije?" Na vprašanje nam odgovori *centralni limitni izrek* (str. 140).

Izberimo enostavni slučajni vzorec velikosti n iz velike populacije s srednjo vrednostjo μ in standardnim odklonom σ. Ko je velikost vzorca n velika, je **vzorčna porazdelitev vzorčnega povprečja** $\overline{x}$ približno normalna s srednjo vrednostjo μ in standardnim odklonom $\frac{\sigma}{\sqrt{n}}$.

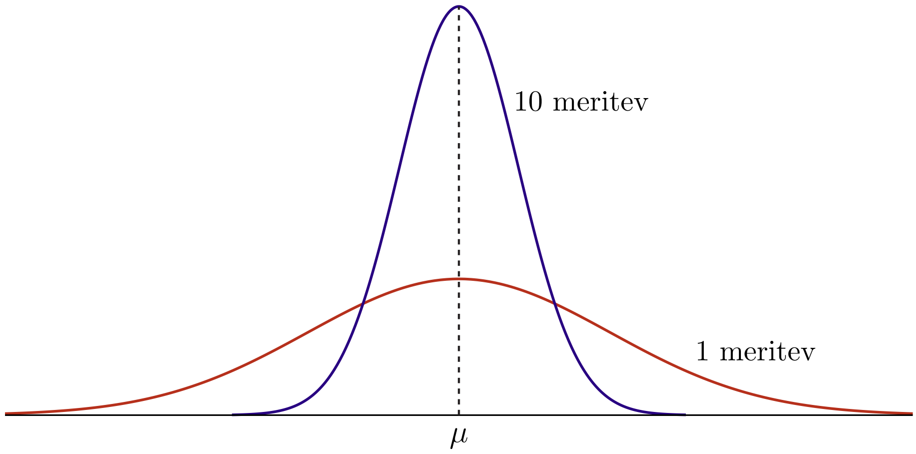
Tej že znani trditvi lahko dodamo še eno dejstvo. *Če je porazdelitev posameznikov v populaciji normalna, potem je vzorčna porazdelitev $\overline{x}$ popolnoma normalna.* To je res za vzorce poljubnih velikosti. Na sliki 4.4 je prikazana zveza med porazdelitvijo ene same vrednosti iz normalno porazdeljene populacije in porazdelitve povprečja več (v tem primeru 10) vrednosti. Povprečje večih vrednosti je manj spremenljivo kot posamezne vrednosti.

Če želimo uporabiti vzorčno porazdelitev $\overline{x}$, moramo za našo populacijo poznati standardni odklon σ. Iz preteklih izkušenj vemo, da je standardni odklon rezultatov NAEPa blizu $\sigma = 60$. Zdaj imamo informacije, ki jih potrebujemo za interval zaupanja srednje vrednosti.

Primer. (Ocena povprečnega NAEP rezultata) Normalna vzorčna porazdelitev za $\overline{x}$ ima srednjo vrednost, ki je enaka neznani srednji vrednosti populacije μ. Standardni odklon vzorčne porazdelitve je

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{840}} = 2{,}1.$$

Pravilo 68–95–99,7 nam pove, da bo $\overline{x}$ v 95% vseh vzorcev največ za dva standardna odklona oddaljen od μ. Ta razdalja je $2 \cdot 2{,}1 = 4{,}2$ odstotne točke. V našem vzorcu



Slika 4.4: Vzorčna porazdelitev vzorčnega povprečja $\overline{x}$ iz enostavnega slučajnega vzorca 10 meritev v primerjavi s porazdelitvijo ene same meritve.

smo dobili $\overline{x} = 272$, torej s stopnjo zaupanja 95% trdimo, da leži srednja vrednost populacije μ na intervalu

$$272 \pm 4{,}2$$

oziroma med 267,8 in 276,2. ♦

Spodaj je recept, ki povzame naša dognanja. Interval zaupanja ima spet obliko

$$\text{ocena} \pm \text{meja napake}.$$

Ocena je v tem primeru enaka vzorčnemu povprečju $\overline{x}$.

Naj bo μ neznana srednja vrednost populacije in σ znani standardni odklon. Iz te populacije izberemo enostavni slučajni vzorec velikosti n in izračunamo vzorčno povprečje $\overline{x}$. Potem je **95% interval zaupanja za srednjo vrednost populacije μ** enak

$$\overline{x} \pm 2\sigma_{\overline{x}} = \overline{x} \pm 2\frac{\sigma}{\sqrt{n}}.$$

Ta ocena velja v primeru, ko je populacija normalno porazdeljena, in je približno pravilna v drugih primerih, če so vzorci veliki.

Velikokrat v praksi ne poznamo vnaprej standardnega odklona σ za dano populacijo. V tem primeru moramo oceniti σ s standardnim odklonom s, ki ga izračunamo za naš vzorec. Če je vzorec velik, bo s blizu σ in zamenjava σ z s ne bo imela velikega vpliva na interval zaupanja.

Oglejmo si še en primer ocenjevanja srednje vrednosti populacije.

Primer. (Ocenjevanje količine prahu v rudnikih premoga) Povprečje več vrednosti je manj spremenljivo kot ena sama vrednost, zato je v primerih, ko je potrebna večja natančnost, dobro opazovati povprečje več meritev namesto ene same. Količino prahu v zraku v rudnikih premoga merimo tako, da izpostavimo filter zraku v rudniku, nato pa izmerimo količino prahu, ki se nanj nabere. Tehtanje ni povsem natančno. Večkratna tehtanja istega filtra se spreminjajo v skladu z normalno porazdelitvijo. Vrednosti, ki bi jih dobili s številnimi tehtanji, sestavljajo našo populacijo. Srednja vrednost μ te populacije je resnična teža (se pravi, da tehtanje ni pristransko). Standardni odklon populacije opisuje natančnost tehtanja. Vemo, da je $\sigma = 0{,}08$ mg. Vsak filter stehtamo trikrat in zapišemo povprečno težo.

Za enega od filtrov so izmerjene teže enake

$$123{,}1 \text{ mg} \qquad 122{,}5 \text{ mg} \qquad 123{,}7 \text{ mg}.$$

Kako bi poiskali 95% interval zaupanja za dejansko težo μ?

Najprej izračunamo vzorčno povprečje:

$$\overline{x} = \frac{123{,}1 + 122{,}5 + 123{,}7}{3} = \frac{369{,}3}{3} = 123{,}1 \text{ mg}.$$

Torej je 95% interval zaupanja enak

$$\overline{x} \pm 2\frac{\sigma}{\sqrt{n}} = 123{,}1 \pm 2\frac{0{,}08}{\sqrt{3}} = 123{,}1 \pm 0{,}09.$$

Smo torej 95% prepričani, da dejanska teža leži med 123,01 mg in 123,19 mg. ♦

4.4 Statistični nadzor procesov

Statistične metode uporabljamo pri zbiranju socialnih in ekonomskih informacij in pri raziskavah na številnih področjih. Večina primerov, ki smo si jih ogledali do sedaj, se je nanašala na ta dva načina uporabe statistike. Statistika pa prispeva tudi k izboljšanju kvalitete proizvedenih izdelkov. Skupaj z novo tehnologijo in novimi načini vodenja (na primer sodelovanjem z delavci in dobavitelji) so statistične ideje pomemben del bojev vsakega proizvajalca, ki želi tekmovati na globalnem tržišču. V tem razdelku si bomo ogledali preprosto a pomembno statistično orodje za kontroliranje in izboljšanje kakovosti, kontrolni diagram.

Primer. (Kontrola računalniških zaslonov) Proizvajalec računalniških zaslonov mora nadzorovati napetost na mreži tankih žic, ki ležijo pod površjem zaslona. Preveč napetosti bi poškodovalo mrežo, premalo bi povzročilo gubanje. Napetost merijo z električno napravo, ki izpisuje meritve v milivoltih (mV). Pravilna napetost je 275 mV. V proizvodnem procesu vedno prihaja do manjših odstopanj. Kadar proces poteka pravilno, je standardni odklon teh meritev $\sigma = 43mV$.

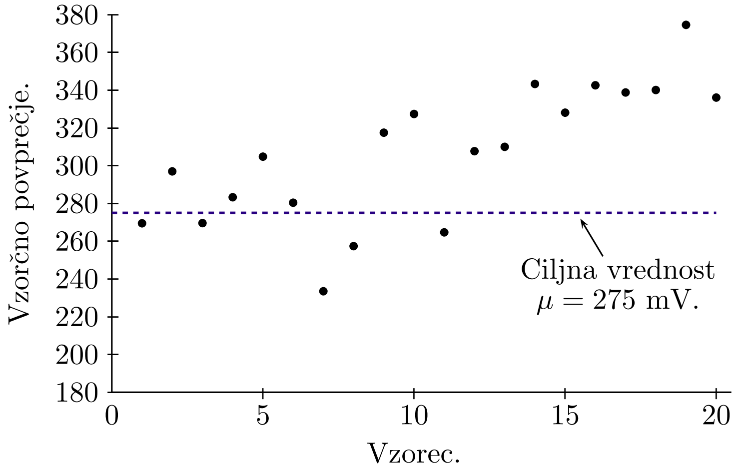
Tehnik vsako uro izmeri napetosti na vzorcu štirih zaslonov. Srednja vrednost $\overline{x}$ vsakega vzorca je ocena za povprečno napetost μ celotnega procesa v tistem trenutku. V tabeli 4.1 so prikazane vrednosti $\overline{x}$ za 20 zaporednih ur. Kako si lahko pomagamo s temi podatki, da bi obdržali proces stabilen? ♦

Vzorec	$\overline{\mathbf{x}}$	Vzorec	$\overline{\mathbf{x}}$
1	269,5	11	264,7
2	297,0	11	307,7
3	269,6	11	310,0
4	283,3	11	343,3
5	304,8	11	328,1
6	280,4	11	342,6
7	233,5	11	338,8
8	257,4	11	340,1
9	317,5	11	374,6
10	327,4	11	336,1

Tabela 4.1: Povprečja $\overline{x}$ za 20 vzorcev velikosti 4.

Graf podatkov v odvisnosti od časa nam bo pomagal videti, ali je bil proces moten. Na sliki 4.5 je diagram zaporednih povprečij meritev. Ker je ciljna vrednost procesa $\mu = 275$ mV, narišemo vzdolž diagrama *središčno črto* na tem nivoju. Povprečja iz kasnejših vzorcev vsa ležijo nad to črto in so dosledno višja od tistih iz prejšnjih vzorcev. Iz tega lahko sklepamo, da je srednja vrednost procesa μ dvignila stran od ciljne vrednosti 275 mV. Vendar pa morda ta dvig $\overline{x}$ predstavlja le naravno spremenljivost v procesu. Diagram moramo podpreti še z računom.

Pričakujemo, da bo $\overline{x}$ porazdeljen približno normalno. Ne samo da so meritve napetosti približno normalne, tudi centralni limitni izrek sugerira, da bodo vzorčna povprečja bližje normalni porazdelitvi kot posamezne meritve. Če ostane standardni



Slika 4.5: Graf vzorčnega povprečja meritev napetosti na računalniških zaslonih v odvisnosti od časa. Vodoravna črta je ciljna vrednost 275.

odklon pri $\sigma = 43$ mV, je standardni odklon povprečja $\overline{x}$ za štiri zaslone enak

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{43}{\sqrt{4}} = 21{,}5.$$

Dokler bo srednja vrednost enaka ciljni vrednosti $\mu = 275$ mV, nam bo pravilo 68–95–99,7 povedalo, da skoraj vse vrednosti $\overline{x}$ ležijo med

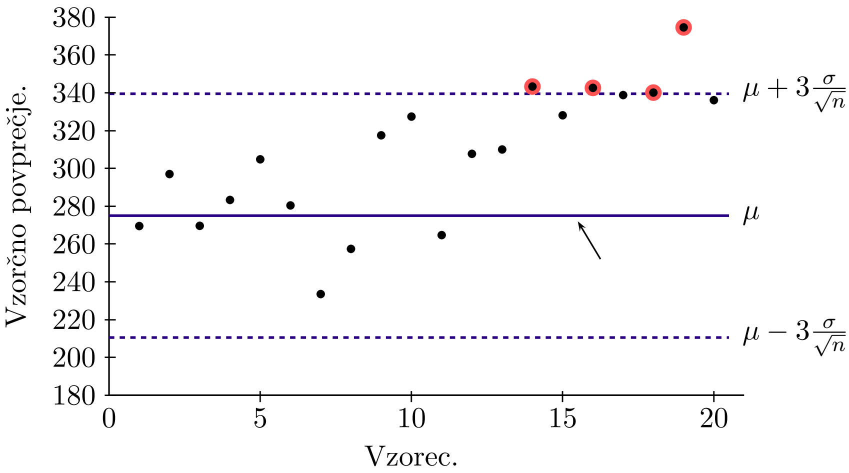
$$\mu - 3\sigma_{\overline{x}} = 275 - 3 \cdot 21{,}5 = 210{,}5 \text{ mV}$$

in

$$\mu + 3\sigma_{\overline{x}} = 275 + 3 \cdot 21{,}5 = 339{,}5 \text{ mV}.$$

Ti *kontrolni meji* smo vnesli v diagram kot črtkani črti.

Na sliki 4.6 je kontrolni diagram za meritve s slike 4.5. Štiri točke, ki so na diagramu obkrožene, ležijo nad zgornjo mejno vrednostjo. Ni zelo verjetno (verjetnost je manjša od 0,003), da bi kakšna od točk ležala nad to mejo, če bi μ in σ ostala nespremenjena. Te štiri točke so torej dober dokaz, da se je porazdelitev proizvodnega procesa spremenila. Zdi se, da se je srednja vrednost nekje pri 14. vzorcu premaknila navzgor. V praksi nadzorniki prežijo na takšne motnje v procesu takoj, ko opazijo prvo odstopanje, se pravi pri 14. vzorcu. Pomanjkanje nadzora bi lahko povzročili nov tehnik, nova serija mrež ali okvara na merilnem aparatu. Pojav meritev nad mejno vrednostjo nas opozori na spremembo, preden izdelamo veliko število pokvarjenih zaslonov. Spodaj je povzetek korakov za izdelavo kontrolnega diagrama, kakršen je na sliki 4.6.



Slika 4.6: Kontrolni diagram za $\overline{x}$ v primeru meritev napetosti na računalniških zaslonih. Črtkani kontrolni črti postavljata mejo pričakovanih odstopanj pri nemotenem procesu.

Za nadzorovanje stabilnosti procesa z danima standardoma μ in σ izdelamo **kontrolni diagram** $\overline{x}$ takole:

- Narišemo povprečja $\overline{x}$ rednih vzorcev velikosti n v odvisnosti od časa.
- Narišemo vodoravno *središčno črto* pri μ.
- Narišemo vodoravni kontrolni črti pri *mejnih vrednostih* $\mu \pm \frac{3\sigma}{\sqrt{n}}$.

Katerikoli $\overline{x}$, ki ne leži med obema kontrolnima črtama, je znak, da smo izgubili nadzor nad procesom.

Statistične ideje so temelji kontrolnih diagramov. Najprej se zavemo, da se v vseh procesih pojavljajo variacije. Naš cilj ni preprečiti te variacije, temveč le ločiti naravna odstopanja v procesu od tistih dodatnih, ki nas opozorijo na motnje. Poleg tega z uporabo normalne vzorčne porazdelitve za $\overline{x}$ in pravila 68–95–99,7 opišemo območje naravnih variacij. Nazadnje združimo to formalno sklepanje z diagramom podatkov, ki ga lahko uporabljajo tudi zaposleni v tovarnah, ki nimajo veliko statistične izobrazbe.

Zakaj vsako uro v vzorec izberemo le štiri zaslone? Namen statistične kontrole procesov ni preverjati funkcionalnosti zaslonov, te bodo podrobno pregledali, ko bodo dokončani. Cilj je spremljanje nameščanja mreže in pravočasna odprava morebitnih napak. Ni praktično preverjati vsakega zaslona na vsakem koraku proizvodnega procesa. Namesto tega nam statistične metode vzorčenja priskrbijo hiter in ekonomičen način, da poskrbimo za gladko delovanje procesa. Kontrolni diagrami, ki temeljijo na vzorcih, zmanjšajo stroške, ki bi lahko nastali, če napak ne bi odkrili pravočasno, ker dopuščajo, da napake sproti odpravljamo. Tako ni potrebno popravljati ali zavreči končanih izdelkov.

V praksi ne iščemo le posameznih točk, ki padejo preko mej, ampak tudi druge sumljive vzorce.

Primer. (Signal "niza devetih") Eden od pogostih signalov, ki nas opozarjajo na napake, je *niz* devetih zaporednih točk nad ali pod središčno črto. Tak niz je zelo malo verjeten, če ostaja srednja vrednost procesa pri ciljni vrednosti, s pomočjo katere narišemo središčno črto – enaka je verjetnosti, da pri devetih zaporednih metih kovanca dobimo same glave. Tak niz torej napeljuje na možnost, da se je srednja vrednost odmaknila s središčne črte.

Na diagramu $\overline{x}$ s slike 4.6 nas signal niza devetih opozori na napako šele pri vzorcu številka 20. Signal, pri katerem smo pozorni na posamezne točke, ki prekoračijo katero od mejnih črt, nas opozori na napako že pri vzorcu 14. V tem primeru je bil signal z nizom devetih počasen pri odkrivanju napake. Kadar pa se srednja vrednost procesa le počasi odmika od ciljne vrednosti, nas bo signal niza devetih opozoril na napako veliko prej, preden bo katerakoli posamezna meritev padla iz dovoljenega območja. Zaradi tega pogosto uporabljamo oba postopka za odkrivanje napak istočasno. ♦

4.5 Nevarnosti analize podatkov

Statistični načrti zbiranja podatkov lahko, kot v primeru študije na zdravnikih, vključujejo eksperimente. Lahko pa tudi uporabljajo posebne postopke vzorčenja kot so na primer tisti, ki jih v ZDA uporabljajo pri raziskavah prebivalstva ali pri kontroli procesov. V obeh primerih se zanašamo na slučajenje in matematično teorijo verjetnosti za izračun vzorčnih porazdelitev. Iz teh lahko potem dobimo rezultate z znanimi stopnjami zaupanja. Vendar pa je formalno statistično sklepanje, ki se odraža v stopnjah zaupanja, sekundarnega pomena ob dobro zasnovanem zbiranju podatkov in vpogledu v njihovo obnašanje. Sklepanje je neuporabno v primeru prostovoljnih vzorcev in ne more odpraviti napak, ki nastanejo zaradi neodziva pri raziskavah na vzorcih. Še več, učinki *skritih spremenljivk* lahko celo iz navidezno jasnega sklepa napravijo zavajujočega. Kot smo videli v prvem poglavju, lahko pri dobro osnovanem eksperimentu nadzorujemo pomešanje skritih spremenljivk z obrazložitvenimi. Kadar eksperiment ni možen, pa moramo včasih postati statistični detektivi. Oglejmo si primer. Čeprav je izmišljen, temelji na resnični situaciji.

Primer. (Katera bolnica je varnejša?) Z namenom pomagati potrošnikom, da bi lažje sprejemali odločitve, povezane z zdravstvom, vlada objavi podatke o usodah pacientov v posameznih bolnicah. Primerjati želiš bolnico A in bolnico B, ki sta v tvoji bližini. Tabela 4.2 predstavlja podatke o preživetju pacientov, ki so bili operirani v teh dveh bolnicah. Vključeni so vsi pacienti, ki so bili nedavno operirani, "preživeti" pa pomeni, da je pacient živel še vsaj 6 tednov po operaciji.

V bolnici A izgubijo 3% ($\frac{63}{2100}$) pacientov, v bolnici B pa le 2% ($\frac{16}{800}$). Zdi se, da je bolnica B boljša izbira, če potrebuješ operacijo. ♦

	Bolnica A	**Bolnica B**
Umrli	63	16
Preživeli	2037	784
Skupaj	2100	800

Tabela 4.2: Podatki o preživetju pacientov.

Tabela 4.2 je **dvosmerna tabela**. Kadar spremenljivke le razvrstijo osebke v kategorije (na primer umrl ali preživel), ne moremo narisati razsevnega diagrama, da bi prikazali zveze med njimi. Namesto tega prikažemo števila v dvosmerni tabeli in dobimo zvezo tako, da primerjamo deleže. Primerjava deležev pacientov, ki so umrli, nam pokaže, da je stopnja umrljivosti v bolnici B nižja. Poglejmo še bolj pozorno.

Primer. (Skrita spremenljivka udari) Niso vsi kirurški posegi enako resni. V nadaljevanju vladnega poročila so podatki o posegih, razdeljeni glede na predoperativno stanje pacientov. Opišemo ga kot "dobro" ali "slabo". Ti podrobnejši podatki so povzeti v tabeli 4.3. Prepričaj se, da so vnosi v prejšnji tabeli le vsote ustreznih števil iz te tabele.

Bolnica A premaga bolnico B pri pacientih, ki so bili v dobrem stanju: izgubili so jih le 1% ($\frac{6}{600}$), medtem ko je v bolnici B izguba znašala 1,3% ($\frac{8}{600}$). Prav tako bolnica A zmaga pri pacientih v slabem stanju z le 3,8% ($\frac{57}{1500}$) proti 4% ($\frac{8}{200}$) iz bolnišnice B. Torej je bolnica A varnejša izbira za paciente ne glede na njihovo stanje. Če se torej pripravljaš na operacijo, bi bilo pametneje izbrati bolnico A. ♦

Ta primer služi kot opozorilo k statističnim dokazom, še posebej v tistih primerih, ko podatkov ne dobimo z eksperimenti. Če ne upoštevamo stanja pacientov, se zdi bolnica B boljša izbira, čeprav se v resnici bolnica A odreže bolje ne glede na stanje.

	Dobro stanje			Slabo stanje	
	Bolnica A	**Bolnica B**		**Bolnica A**	**Bolnica B**
Umrli	6	8	**Umrli**	57	8
Preživeli	594	592	**Preživeli**	1443	192
Skupaj	600	600	Skupaj	1500	200

Tabela 4.3: Podatki o preživetju pacientov glede na zdravstveno stanje.

Kako lahko gre A bolje v obeh skupinah, pa vseeno slabše v skupnem seštevku? Oglejmo si še enkrat podatke. Bolnica A je zdravstveni center, ki pritegne resno bolne paciente iz celotne regije. Sprejeli so kar 1500 pacientov v slabem stanju. V bolnici B so imeli le 200 takih pacientov. Ker je za paciente v slabem stanju bolj verjetno, da bodo umrli, je bila stopnja umrljivosti pri bolnici A večja, kljub temu, da se je odrezala bolje kot B v vsakem posameznem razredu. Tabela 4.2 je zavajujoča, ker ne upošteva stanja pacientov. Statistično sklepanje, ki bi temeljilo le na podatkih iz tabele 4.2, bi bilo prav tako zavajujoče.

Tudi če podatke pridobimo previdno in jih pravilno analiziramo, ne moremo biti absolutno prepričani v pravilnost naših zaključkov. Vedno obstaja možnost, ne glede kako majhna, da nas bo slučajna izbira pripeljala do napačnih sklepov. Moč statističnih sklepov je v tem, da poznamo možnosti za napačne zaključke in jih lahko nadzorujemo s tem, da postavimo stopnjo zaupanja tako visoko, kot se nam zdi potrebno.

Statistika torej ne proizvaja dokazov. Vendar pa je v svetu, ki vedno zahteva resnico in v katerem je večina dokazov nezanesljivih, statistika največkrat najboljša možna izbira.

4.6 Slovarček

delež vzorca (ang. sample proportion) Delež $\hat{p}$ tistih elementov iz vzorca, ki imajo neko lastnost. V primeru enostavnega slučajnega vzorca ga uporabimo kot oceno za ustrezni delež p celotne populacije, iz katere je bil izbran vzorec.

dvosmerna tabela (ang. two-way table) Tabela, ki navaja izide glede na dve spremeljivki (npr. bolniki, razporejeni glede na zdravstveno stanje in glede na uspešnost zdravljenja).

interval zaupanja (ang. confidence interval) Interval, ki ga izračunamo iz vzorca s pomočjo metode, ki z znano verjetnostjo poda interval, ki vsebuje neznani parameter. To verjetnost imenujemo *stopnja zaupanja*; običajno imajo intervali zaupanja obliko

$$\text{ocena} \pm \text{meja napake}.$$

kontrolni diagram (ang. control chart) Graf, ki prikazuje vrednost statistike pri zaporednih vzorcih (npr. en vzorec na uro ali en vzorec vsako izmeno). Graf vsebuje tudi sredinsko črto pri ciljni vrednosti procesnega parametra in kontrolni mejni vrednosti, ki jih statistika običajno ne prestopi, razen če se proces oddalji od ciljne vrednosti. Namen kontrolnega diagrama je spremljanje procesa v času in odkrivanje neobičajnih vplivov na proces.

parameter (ang. parameter) Število, ki opisuje populacijo. Običajno je cilj statističnega sklepanja oceniti neznani parameter ali pa določiti njegovo vrednost.

statistika (ang. statistic) Število, ki opisuje vzorec. Izračunamo jo lahko iz vzorca in pri tem ne potrebujemo nobenih neznanih parametrov, ki opisujejo populacijo.

vzorčno povprečje (ang. sample mean) Srednja vrednost (aritmetično povprečje) $\overline{x}$ vseh vrednosti v vzorcu. V primeru enostavnega slučajnega vzorca ga uporabimo kot oceno za neznano srednjo vrednost μ za populacijo, iz katere je bil izbran vzorec.

4.7 Dodatna literatura

- David S. Moore, *The Basic Practice of Statistics*, 2. izdaja, Freeman, New York, 1999. Šesto poglavje te knjige predstavlja podrobnosti statističnega sklepanja. V sedmem in osmem poglavju avtor obravnava praktično uporabo sklepanja pri ocenjevanju srednjih vrednosti in deležev.
- Lincoln E. Moses, *The Reasoning of statistical interference,* Perspectives on Contemporary Statistics*, Mathematical Association of America, Washington, D.C., 1992, str. 107–122.* Ta esej o naravi sklepanja zahteva nekaj znanja verjetnosti in pred tem je priporočljivo prebrati zgoraj omenjeno knjigo.

Zanimivo in poučno je videti, kako vodilne agencije opisujejo natančnost svojih raziskav. Celotna tiskovna poročila, ki jih izdajajo pri agenciji Gallup, so dostopna na www.gallup.com. Harrisova raziskava je dosegljiva na www.louisharris.com pod “*Harris Poll this week*”. Mesečna poročila o stopnji nezaposlenosti shranjuje Urad za statistiko dela na stats.bls.gov. Nacionalno združenje raziskovalnih agencij NCPP ima na svoji strani www.ncpp.org nekaj precej zanimivega materiala (*Principles of Disclosure, 20 Questions for Journalists*).

4.8 Preverjanje znanja

(1) Slučajni vzorec 10 vreč sladkorja ima povprečno težo 4,9 funta, kar je manj kot povprečna teža vseh proizvedenih vreč sladkorja, ki znaša 5,05 funta. V tem primeru je 4,9

(a) statistika.

(b) parameter.

(c) vzorec.

(2) Da bi ugotovili, koliko je zanimanja za novi park, anketiramo 300 okoliških prebivalcev. Med njimi jih je 135 naklonjenih parku. Kolikšen je delež vzorca $\hat{p}$?

(a) 0,45%

(b) 40,5%

(c) 45%

(3) Enostavni slučajni vzorec 400 prebivalcev nekega mesta povprašamo za mnenje in 144 jih je za novi gasilski dom. Kolikšen je približno standardni odklon deleža vzorca $\hat{p}$ za ta primer?

(a) 6%

(b) 2,4%

(c) približno 1%

(4) Tovarna izdeluje vreče sladkorja s povprečno težo 5,05 funta in standardnim odklonom 0,05 funta. Slučajno izberemo štiri vreče in izračunamo njihovo povprečno težo. Kolikšen je standardni odklon vzorčne porazdelitve povprečne teže?

(a) 0,0125 funta
(b) 0,025 funta
(c) 0,1 funta

(5) S slučajnim vzorcem velikosti 2000, ki ga izberemo med prebivalci Spodnjega Kašlja, ugotovimo, da jih 45 še nikoli ni imelo noric. Poišči 95% interval zaupanja za dejanski delež meščanov, ki še nikoli niso imeli noric.

(a) 2,25%± 0,3316%
(b) 2,25%± 0,6632%
(c) 2,25%± 0,2199%

(6) Vreče sladkorja imajo povprečno težo 5,05 funta in standardni odklon 0,05 funta. Tovarna uporablja 95% kontrolni limiti v kontrolnem diagramu proizvodnega procesa. Izberemo slučajni vzorec 9 vreč in jih stehtamo. Povprečna teža je 5,09 funta. Katera od naslednjih trditev je pravilna?

(a) Povprečna teža vzorca je ušla nadzoru.
(b) Povprečna teža vzorca je še pod nadzorom.
(c) Nimamo dovolj informacij.

(7) Spodaj je dvosmerna tabela, ki prikazuje sprejem gostov v privatni klub. Poišči delež sprejetih moških.

	Moški	**Ženske**
Sprejeti	20	20
Zavrnjeni	30	10

(a) 20%
(b) 25%
(c) 40%

4.9 Naloge

Ocenjevanje deleža populacije

Za vsako od debelo natisnjenih števil v nalogah 1–3 povej, ali gre za *parameter* ali *statistiko*.

(1) Za slučajni vzorec študentk določimo povprečno višino **65** inčev, kar je več kot povprečna višina vseh Američank, ki znaša **63** inčev.

(2) Raziskovalec izvede slučajeni primerjalni eksperiment na mladih podganah, da bi ugotovil vpliv toksične snovi v hrani. Kontrolna skupina dobi nekontaminirano hrano, eksperimentalna skupina pa dobi v hrani 2500 delcev na milijon toksične snovi. Po osmih tednih je povprečna pridobljena teža v kontrolni skupini **335** g, v eksperimentalni pa **289** g.

(3) Podjetje za telefonsko oglaševanje iz Los Angelesa uporablja napravo, ki slučajno izbira telefonske številke iz mesta. Od prvih 100 klicanih jih **48%** ni bilo v imeniku. To ni presenetljivo, ker vemo, da **52%** vseh telefonskih številk ni v imeniku.

(4) Tonya želi oceniti, kolikšen delež študentov v njenem študentskem domu je zadovoljnih s hrano v študentski menzi. Ankerira enostavni slučajni vzorec 50 od 680 študentov, ki stanujejo v tem domu. Od tega jih 14 meni, da je hrana dobra.

(a) Opiši populacijo in z besedami povej, kaj je parameter p.

(b) Podaj (v %) vrednost statistike $\hat{p}$, ki ocenjuje p.

(c) Recimo, da je v resnici 25% študentov zadovoljnih s hrano iz menze. Kolikšna sta srednja vrednost in standardni odklon vzorčne porazdelitve $\hat{p}$?

(5) PTC je spojina, ki ima za nekatere močan grenak okus, drugim pa se zdi brez okusa. Sposobnost zaznavanja PTC je dedna. Približno 75% Italijanov lahko zazna PTC. Oceniti želiš delež Američanov, pri katerih je vsaj eden od starih staršev iz Italije in ki lahko okusijo PTC. Predpostavi, da ocena o 75% deležu Italijanov drži za to populacijo in testiraj 500 ljudi. Skiciraj normalno krivuljo, ki prikazuje, kako se bo delež $\hat{p}$ vseh tistih, ki okusijo PTC, spreminjal pri večkratnem vzorčenju.

(6) V eni od držav na ameriškem srednjem zahodu ima 84% gospodinjstev za praznike ob koncu leta božično drevesce. V raziskavi slučajni vzorec 400 gospodinjstev vprašamo "Ste letos postavili božično drevesce?" Kakšna je vzorčna porazdelitev deleža pozitivnih odgovorov?

(7) Standardni odklon $\sigma_{\hat{p}}$ vzorčnega deleža $\hat{p}$ se spreminja skupaj z dejansko vrednostjo deleža populacije p. Na srečo se ne spreminja veliko, če p ni blizu 0% ali

100%. Recimo, da je velikost vzorca $n = 1500$. Izračunaj $\sigma_{\hat{p}}$ za $p = 30\%$, 40%, 50%, 60% in 70%. Nato izračunaj $\sigma_{\hat{p}}$ za $p = 0\%$, 10% in 20%. Na katerem delu se $\sigma_{\hat{p}}$ spreminja najhitreje, ko se spreminja p? Nariši graf $\sigma_{\hat{p}}$ v odvisnosti od p.

Intervali zaupanja

(8) Poročilo o raziskavi na vzorcu 1500 odraslih pravi: "S stopnjo zaupanja 95% med 27% in 33% vseh Američanov meni, da so droge največji problem v javnih šolah." Razloži nekomu, ki ne zna statistike, kaj v tem primeru pomeni "s stopnjo zaupanja 95%".

(9) Pri Gallupovi raziskavi so vprašali slučajni vzorec 1005 odraslih, če podpirajo dvojezične javne šole ali pa menijo, da bi se morali učenci, ki ne govorijo angleško, vključiti v angleške šole in se naučiti jezika na ta način. Pri tem je bilo 63% za enojezične šole. V tisku so objavili, da je bila meja napake te raziskave 3%. Razloži, kaj to pomeni.

(10) Recimo, da so v raziskavi iz prejšnje naloge uporabili enostavni slučajni vzorec velikosti 1005 in ugotovili, da je 60% vprašanih za enojezične šole. Podaj 95% interval zaupanja za delež odraslih, ki bi odgovorili, da so za enojezične šole.

(11) V preteklem letu je 73% vprašanih študentov prvega letnika v nacionalni raziskavi odgovorilo, da je "dobra finančna preskrbljenost" zanje pomemben cilj. Državna univerza je ugotovila, da je v enostavnem slučajnem vzorcu 200 študentov prvega letnika 132 študentov označilo ta cilj kot pomemben. Podaj 95% interval zaupanja za delež prvih letnikov na univerzi, za katere je to pomemben cilj.

(12) V ZDA razmišljajo, da bi dodatno omejili število vozil, ki lahko vstopijo v Yellowstonski nacionalni park. Da bi ocenili odziv javnosti, vprašajo enostavni slučajni vzorec 150 obiskovalcev, če podpirajo takšne ukrepe. Od teh jih 89% odgovori pritrdilno. Podaj 95% interval zaupanja za delež obiskovalcev parka, ki so za omejitev. Ali lahko s 95% stopnjo zaupanja trdiš, da jih je več kot polovica za omejitev? Odgovor utemelji.

(13) *New York Times* in *CBS News* sta izvedla nacionalno raziskavo na 1048 slučajno izbranih najstnikih med 13. in 17. letom. Med njimi jih je 692 imelo televizijski sprejemnik v svoji sobi in 189 jih je kot svoj najljubši program navedlo *Fox*. Predpostavljali bomo, da je šlo za enostavni slučajni vzorec.

(a) Podaj 95% interval zaupanja za delež vseh ljudi te starosti, ki imajo v svoji sobi TV sprejemnik, in za tiste, ki najraje gledajo *Fox*.

(b) Časopisni članek pravi, "Teoretično se v 19 primerih od 20 rezultati ne razlikujejo za več kot 3 odstotne točke od vrednosti, ki bi jo dobili, če bi anketirali vse ameriške najstnike." Pojasni, kako se tvoji izračuni ujemajo s to trditvijo.

(14) V raziskavo o enakopravnosti žensk v ZDA, ki jo je izvedla televizijska hiša MSNBC, je bilo vključenih 1019 odraslih oseb. Časopisni članek, ki je poročal o raziskavi, je navajal, "Meja napake pri rezultatih je 3 odstotne točke."

(a) Skupno je 54% vzorca (550 od 1019 ljudi) odgovorilo, da se je na tem področju naredilo dovolj. Poišči 95% interval zaupanja za delež odraslih, ki bi odgovorili pritrdilno. Ali je poročilo o meji napake približno pravilno? (Predpostavi, da je šlo za enostavni slučajni vzorec.)

(b) Časopisni članek je trdil, da 65% moških in le 43% žensk meni, da se je na tem področju storilo dovolj. Pojasni, zakaj nimamo dovolj informacij, da bi lahko podali intervale zaupanja posebej za moške in za ženske.

(c) Ali bi bila meja napake pri 95% intervalu zaupanja za ženske večja, manjša ali enaka 0,03? Zakaj? Navedbe časopisa o meji napake so očitno nekoliko zavajujoče.

Naloge 15–18 temeljijo na naslednji situaciji: Poročilo pravi, da so pri nacionalni raziskavi na 1500 slučajno izbranih odraslih ugotovili, da jih je 43% mnenja, da jim bo šlo naslednje leto slabše. V nadaljevanju poročila je bilo zapisano, da je meja napake 3 odstotne točke s 95% stopnjo zaupanja.

(15) Kateri od naslednjih virov napak so vključeni v mejo napake iz poročila?

(a) Pri raziskavi so slučajno izbirali telefonske številke, zato so pri tem izpustili ljudi, ki niso imeli telefona.

(b) Neodziv: pri nekaterih od teh številk se ni nihče oglasil ali pa je klicani zavrnil sodelovanje.

(c) Slučajne variacije pri slučajnem izbiranju telefonskih številk.

(16) Ali bi bila pri 90% intervalu zaupanja na osnovi rezultatov te ankete meja napake večja, manjša ali enaka 3 odstotnim točkam? Zakaj?

(17) Recimo, da bi pri raziskavi anketirali 1000 ljudi namesto 1500 (in spet ugotovili, da jih je 43% mnenja, da jim bo šlo prihodnje leto slabše). Ali bi bila meja napake za 95% interval zaupanja večja, manjša ali enaka 3 odstotnim točkam? Zakaj?

(18) Recimo, da smo rezultat 43% dobili s podobno metodo enostavnega slučajnega vzorčenja za vse odrasle v državi New York (z 18 milijoni prebivalcev) ne pa za celotne ZDA (z 270 milijoni prebivalcev). Ali bi bila meja napake 95% intervala zaupanja večja, manjša ali enaka 3 odstotnim točkam? Zakaj?

Ocenjevanje srednje vrednosti populacije

(19) Pošiljka mehanskih delov ima kritično dimenzijo, ki je normalno porazdeljena s srednjo vrednostjo 12 cm in standardnim odklonom 0,01 cm. Sprejemna skupina premeri slučajni vzorec 25 delov. Kakšna je vzorčna porazdelitev vzorčnega povprečja $\overline{x}$ kritičnih dimenzij za te dele?

(20) Rezultati študentov na ACT sprejemnih testih v preteklem letu so bili normalno porazdeljeni s srednjo vrednostjo $\mu = 18{,}6$ in standardnim odklonom $\sigma = 5{,}9$.

(a) Kateri interval vsebuje srednjih 95% rezultatov?

(b) Izračunamo povprečje 25 slučajno izbranih rezultatov. Kateri interval vsebuje srednjih 95% povprečij $\overline{x}$?

(21) Napake pri natančnih merjenjih so velikokrat normalno porazdeljene. Izkušnje kažejo, da se napake pri kontrolnih metodah spreminjajo, ko meritev ponavljamo, v skladu z normalno porazdelitvijo s srednjo vrednostjo 0 (se pravi, da postopek ne bo sistematično precenil ali podcenil dejanske razdalje) in standardnim odklonom 0,03 m. Geodet ponovi vsako meritev trikrat in za končno vrednost uporabi povprečje treh meritev. Napaka pri tej vrednosti je povprečna napaka $\overline{x}$ teh treh zaporednih meritev.

(a) Kakšna je porazdelitev povprečne napake $\overline{x}$, ko geodet izmeri veliko razdalj?

(b) Med katerima vrednostima leži 95% napak?

(22) Pri študiji poklicnih poti upravnikov hotelov so poslali vprašalnike enostavnemu slučajnemu vzorcu 160 hotelov iz velikih ameriških hotelskih verig. Prejeli so 114 odgovorov. Povprečen čas, v katerem je teh 114 upravnikov delalo pri svojem trenutnem podjetju, je bil $\overline{x} = 11{,}78$ let. Ne poznamo standardnega odklona populacije σ, vzorčni standardni odklon pa je enak $s = 3{,}2$ leta. Ker je vzorec velik, je s blizu σ. Podaj 95% interval zaupanja za povprečno število let, ki so jih upravniki velikih verig preživeli pri svojem trenutnem podjetju.

(23) Pri laboratorijski tehtnici je standardni odklon $\sigma = 0{,}001$ g pri večkratnih tehtanjih. Predpostavi, da so meritve pri večkratnih tehtanjih normalno porazdeljene s srednjo vrednostjo, ki je enaka dejanski teži tehtanega predmeta. Pri treh tehtanjih nekega primerka smo dobili

$$3{,}412 \qquad 3{,}414 \qquad 3{,}415$$

Podaj 95% interval zaupanja za dejansko težo primerka. Kolikšna sta ocena in meja napake pri tem intervalu?

(24) V spodnji tabeli so rezultati IQ testa 31 učenk 7. razreda iz neke ameriške šole.

114	100	104	89	102	91
114	114	103	105	108	130
120	132	111	128	118	119
86	72	111	103	74	112
107	103	98	96	112	112
93					

(a) Pričakujemo, da bo porazdelitev rezultatov blizu normalni. Napravi stebelni diagram porazdelitve teh 31 rezultatov. Ali na diagramu opaziš kakšne ubežnike, izrazito asimetričnost ali druga nenavadna odstopanja?

(b) Obravnavaj teh 31 deklic kot enostavni slučajni vzorec vseh sedmošolk iz tega okrožja. Predpostavi, da je standardni odklon rezultatov IQ testa v tej populaciji enak $\sigma = 15$. Podaj 95% interval zaupanja za povprečni rezultat v tej populaciji.

(c) V resnici pripadajo ti rezultati vsem sedmošolkam ene od šol iz tega okrožja. Natančno razloži, zakaj se ne moremo zanesti na interval zaupanja iz točke (b).

(25) Poišči mejo napake za 95% interval zaupanja iz naloge 23, če tehtamo vsak primerek dvanajstkrat in ne le trikrat. Prepričaj se, da je tvoj rezultat dvakrat manjši od meje napake iz naloge 23. Razloži, zakaj lahko že brez računanja ugotovimo, da bo nova meja napake dvakrat manjša.

(26) Pri NAEP testu je sodelovalo tudi 1077 žensk med 21. in 25. letom starosti. Njihov povprečni rezultat je bil 275. Predpostavi, da je standardni odklon posameznih rezultatov enak $\sigma = 60$.

(a) Podaj 95% interval zaupanja za povprečni rezultat μ v populaciji vseh žensk med 21. in 25. letom.

(b) Predpostavi, da so dobili enak rezultat, $\overline{x} = 275$, za vzorec 250 žensk. Podaj 95% interval zaupanja za srednjo vrednost populacije v tem primeru.

(c) Predpostavi, da so pri vzorcu 4000 žensk dobili vzorčno povprečje $\overline{x} = 275$ in še enkrat izračunaj 95% interval zaupanja za μ.

(d) Kolikšne so meje napake za vzorce velikosti 250, 1077 in 4000? Kako večja velikost vzorca vpliva na mejo napake za interval zaupanja?

(27) Naprava za predelovanje mleka spremlja število bakterij na mililiter v surovem mleku, ki ga sprejmejo v predelavo. Za slučajni vzorec 10 enomililiterskih primerkov so podatki zbrani v spodnji tabeli.

5370	4890	5100	4500	5260
5150	4900	4760	4700	4870

Predpostavimo, da je število bakterij normalno porazdeljeno in je standardni odklon $\sigma = 265$ na ml. Podaj 95% interval zaupanja za povprečno število bakterij na ml v mleku tega proizvajalca.

(28) V radijski oddaji povabijo poslušalce k razpravi o predlaganem povišanju plač mestnih svetnikov. "Kako visoke bi morale biti po vašem mnenju plače mestnih svetnikov? Pokličite nas in povejte svoje mnenje." Skupno pokliče 958 ljudi. Povprečna predlagana plača je $\overline{x} = 8740$ € na leto in standardni odklon odgovorov je $s = 1125$ €. Pri velikih vzorcih kakršen je ta je s zelo blizu neznanemu standardnemu odklonu za populacijo, σ. Radijska postaja izračuna 95% interval zaupanja za povprečno plačo μ in ugotovi, da bi bilo povprečje predlogov vseh meščanov med 8667 € in 8813 €.

(a) Pokaži, da je zgornji izračun pravilen.

(b) Kljub temu dobljeni rezultat ne velja za celotno populacijo. Pojasni, zakaj.

Statistični nadzor procesov

Pri obravnavanju kontrolnih diagramov uporabljaj hkrati signal "ene zunanje točke" in signal "niza devetih".

(29) Proizvajalec avtomobilskih klimatskih naprav vsako uro proizvodnje pregleda vzorec štirih termostatov. Termostati so nastavljeni na $75°F$ in nato vstavljeni v komoro, v kateri postopoma zvišujejo temperaturo. Zabeležijo temperaturo, pri kateri termostat vključi klimatsko napravo. Standard za srednjo vrednost je $\mu = 75°$. Pretekle izkušnje kažejo, da se odzivna temperatura pravilno nastavljenih termostatov spreminja s $\sigma = 0{,}5°$. Povprečno odzivno temperaturo $\overline{x}$ za vsakega od vzorcev vnašajo v kontrolni diagram za $\overline{x}$. Izračunaj središčno črto in mejni vrednosti za ta diagram.

(30) Širina reže, ki jo izreže frezalni stroj, je pomembna za pravilno delovanje hidravličnih sistemov pri velikih traktorjih. Proizvajalec vsako uro preverja širino z vzorcem petih zaporednih izdelkov. Povprečno širino reže za vsak vzorec vnese na kontrolni diagram. Ciljna vrednost je $\mu = 0{,}8759$ inča. Frezalni stroj pri pravilni nastavitvi izdeluje reže s povprečno širino enako ciljni vrednosti in standardnim odklonom $\sigma = 0,0012$ inča. Določi središčno črto in obe mejni vrednosti za kontrolni diagram $\overline{x}$.

(31) Laboratoriji imajo ponavadi kontrolni diagram za proces merjenja, ki temelji na večkratnih meritvah standardnega primerka. Tvoja zadolžitev je kontrolni diagram za tehtnico iz naloge 23. V rednih časovnih intervalih moraš trikrat zapored stehtati standardno 5-gramsko utež. Določi sredinsko črto in mejni vrednosti za ta kontrolni diagram.

(32) Premer potisnega ležaja v električnem motorju naj bi bil 2.205 cm. Kadar je proizvodni proces pravilno nastavljen, izdelujemo ležaje s povprečnim premerom 2,2050 cm in standardnim odklonom 0,0010 cm. Vsako uro izmerimo vzorec pet zaporedoma izdelanih ležajev. V spodnji tabeli so vzorčna povprečja $\overline{x}$ za 12 ur.

Napravi kontrolni diagram $\overline{x}$ za diameter potisnega ležaja. Uporabi oba možna signala, da oceniš, če je proces pod kontrolo. Kdaj bi morali ukrepati, da bi popravili potek procesa?

Ura	1	2	3	4
$\overline{x}$	2,2047	2,2047	2,2050	2,2049
Ura	5	6	7	8
$\overline{x}$	2,2053	2,2043	2,2036	2,2042
Ura	9	10	11	12
$\overline{x}$	2,2038	2,2045	2,2026	2,2040

Naloge 33–35 se nanašajo na naslednje podatke: Proizvajalec zdravil oblikuje tablete s pomočjo kompresije granul, ki vsebujejo aktivno sestavino in razna polnila. Za vsako skupino tablet izmerijo trdoto vzorca in tako nadzorujejo proces kompresije. Ciljni vrednosti za trdoto sta $\mu = 11{,}5$ in $\sigma = 0{,}2$. V tabeli 4.4 so tri skupine podatkov, od katerih vsaka predstavlja povprečje $\overline{x}$ za 20 zaporednih vzorcev $n = 4$ tablet.

(33) Napravi kontrolni diagram $\overline{x}$ za skupino A iz tabele 4.4. Srednja vrednost procesa μ se je med zbiranjem teh podatkov nenadoma prestavila k novi vrednosti. Ali je na diagramu opazno uhajanje izpod nadzora? Kako? Pri katerem od vzorcev se je spremenila srednja vrednost?

(34) Srednja vrednost procesa μ se je ustalila pri ciljni vrednosti $\mu = 11{,}5$ med zbiranjem podatkov iz skupine B. Vzorčna povprečja se spreminjajo, ampak le v okviru pričakovanih slučajnih odstopanj. Izdelaj kontrolni diagram $\overline{x}$ in ga komentiraj.

(35) Skupina podatkov C iz tabele 4.4 ilustrira učinek enakomernega premikanja srednje vrednosti populacije. Proces je stabilen pri prvih desetih vzorcih, nato pa se začne srednja vrednost μ vztrajno dvigati. Napravi kontrolni diagram

Vzorec	Skupina A	Skupina B	Skupina C
1	11,602	11,627	11,495
2	11,547	11,613	11,475
3	11,312	11,493	11,465
4	11,449	11,602	11,497
5	11,401	11,360	11,573
6	11,608	11,374	11,563
7	11,471	11,592	11,321
8	11,453	11,458	11,533
9	11,446	11,552	11,486
10	11,522	11,463	11,502
11	11,664	11,383	11,534
12	11,823	11,715	11,624
13	11,629	11,485	11,629
14	11,602	11,509	11,575
15	11,756	11,429	11,730
16	11,707	11,477	11,680
17	11,612	11,570	11,729
18	11,628	11,623	11,704
19	11,603	11,472	12,052
20	11,816	11,531	11,905

Tabela 4.4: Tri skupine $\overline{x}$ za 20 vzorcev velikosti 4.

$\overline{x}$ za te podatke. Ali je razvidno kakšno uhajanje izpod nadzora? Ali je na diagramu opazno plezanje μ navzgor?

Nevarnosti analize podatkov

(36) Kako na valjenje jajc vodnega pitona vpliva temperatura kačjega gnezda? Raziskovalci so razdelili novo znesena jajca v tri skupine glede na temperaturo: vroče, nevtralno in mrzlo. V prvi skupini so posnemali toploto, ki jo običajno priskrbi samica pitona, pri tretji skupini pa so simulirali odsotnost samice. V spodnji tabeli so podatki o številih jajc in njihovih usodah. (Vir: R. Shine, T. R. L. Madsen, M. J. Elphick, P. S. Harlow, The influence of nest temperatures and maternal brooding on hatchling phenotypes in water pythons, *Ecology*, 78 (1997): 1713–1721.)

	Št. jajc	**Izvaljeni**
Hladno	27	16
Nevtralno	56	38
Vroče	104	75

(a) Izdelaj dvosmerno tabelo temperature v odvisnosti od izida (izleglo ali ne).

(b) Za vsako skupino izračunaj delež jajc, iz katerih so se izlegli pitoni. Raziskovalci so pričakovali, da bo mrzlo okolje zmanjšalo število izvaljenih jajc. Ali podatki podpirajo njihova pričakovanja?

(37) V študiji vpliva kadilskih navad staršev na navade srednješolcev so raziskovalci anketirali dijake iz osmih srednjih šol v Arizoni. Rezultati so povzeti v spodnji tabeli. (Vir: S. V. Zagona, ed., *Studies and Issues in Smoking Behavior*, University of Arizona Press, Tucson, 1967, str. 157–180.)

	Študent kadi	**Študent ne kadi**
Oba starša kadita	400	1380
Eden od staršev kadi	416	1823
Nobeden od staršev ne kadi	188	1168

Opiši povezavo med kadilskimi navadami staršev in njihovih otrok, tako da izračunaš različne deleže in jih primerjaš. Nato povzemi rezultate še s preprostimi besedami.

(38) Strelno orožje je drugi najpogostejši vzrok smrti iz nezdravstvenih razlogov (prvi so motorna vozila). V tabeli so zbrani podatki za Milwaukee, Wisconsin, med leti 1990 in 1994. Primerjati želimo vrste strelnega orožja, ki so bile uporabljene v umorih in v samomorih. Predpostavljamo, da bodo šibrovke in lovske puške bolj pogosto uporabljene pri samomorih, ker jih ima veliko ljudi doma za lov. Kaj pravi o tem spodnja tabela? (Vir: S. W. Hargarten et al., Characteristics of firearms involved in fatalities, *Journal of American Medical Association*, 275 (1996): 42–45.)

	Umori	**Samomori**
Pištola	468	124
Šibrovka	28	22
Lovska puška	15	24
Nedoločeno	13	5
Skupaj	524	175

(39) V tabeli so podatki o točnih letih in zamudah dveh letalskih družb na petih različnih letališčih v obdobju enega meseca. V medijih pogosto poročajo o skupnem deležu točnih letov. Zaradi skritih spremenljivk so lahko taka poročila zavajujoča. (Vir: A. Barnett, How numbers can trick you, *Technology Review*, oktober 1994, str. 38–45.)

	Alaska Airlines		**America West**	
	Točen	Zamujal	Točen	Zamujal
Los Angeles	497	62	694	117
Phoenix	221	12	4840	415
San Diego	212	20	383	65
San Francisco	503	102	320	129
Seattle	1841	305	201	61

(a) Kolikšen delež vseh letov družbe Alaska Airlines predstavljajo tisti z zamudami? Kolikšen je ta delež pri družbi America West? To so vrednosti, ki jih običajno navajajo v medijih.

(b) Poišči delež letov z zamudami za družbo Alaska Airlines na vsakem od petih letališč. Ponovi izračune še za družbo America West.

(c) America West se odreže slabše na čisto vsakem od petih letališč, vendar jih gre v celoti bolje. Zveni nemogoče. S pomočjo podatkov razloži, zakaj pride do tega. (Vzroki tičijo v vremenskih pogojih na letališčih v Phoenixu in v Seattlu.)

(40) Kaže, da je to, ali se obsojenemu morilcu izreče smrtna kazen ali ne, odvisno od rase žrtve. Spodaj so podatki o 326 primerih, v katerih so bili obtoženi umora spoznani za krive. (Vir: M. Radelet, Racial characteristics and imposition of the death penalty, *American Sociological Review*, 46 (1981): 918–927.)

	Beli obtoženec			**Črni obtoženec**	
	Smrtna kazen			**Smrtna kazen**	
	Da	Ne		Da	Ne
Bela žrtev	19	132	Bela žrtev	11	52
Črna žrtev	0	9	Črna žrtev	6	97

(a) S pomočjo zgornjih podatkov sestavi dvosmerno tabelo rase obtoženega (bela ali črna) v primerjavi s smrtno kaznijo (da ali ne).

(b) Pokaži, da velja Simpsonov paradoks: skupno je na smrt obsojenih več belcev kot črncev, vendar pa je pri obeh rasah žrtev odstotek obsojenih črncev višji.

(c) S pomočjo podatkov razloži vzroke za paradoks tako, da te bo lahko razumel tudi sodnik.

Dodatne naloge

(41) Študent prebere, da je 95% interval zaupanja povprečnega rezultata NAEP testov za moške med 21. in 25. letom starosti med 267,8 do 276,2. Ko ga vprašamo, kaj to pomeni, pravi, da “95% moških iz te starostne skupine doseže rezultate med 267,8 in 276,2”. Ali ima prav? Odgovor utemelji.

(42) V tehnološkem kotičku *New York Timesa* so 29. maja 1995 poročali o raziskavi o uporabnikih interneta, ki je pokazala, da je med njimi kar dvakrat več moških kot žensk. To je bilo presenetljivo, saj je ena prejšnjih raziskav pokazala, da je bilo moških kar devetkrat več. V nadaljevanju članka najdemo naslednje:

> Natančne raziskave so vključevale več kot 13 000 organizacij, kjer uporabljajo internet. Prejeli so 1468 uporabnih odgovorov, g. Quaterman pa je povedal, da je meja napake 2,8 odstotka pri stopnji zaupanja 95%.

(a) Kolikšna je stopnja odziva te raziskave? (Se pravi, kolikšen delež načrtovanega vzorca je odgovoril na vprašalnik?)

(b) Ali je po tvojem mnenju majhna meja napake dobro merilo natančnosti rezultatov te raziskave? Odgovor utemelji.

(43) Anketa *New York Timesa* je zajela 1025 slučajno izbranih žensk iz ZDA (vključno z Aljasko in Havaji). Med vprašanimi jih je bilo 47% mnenja, da nimajo dovolj časa zase.

(a) Zapisali so, da je meja napake ± 3 odstotne točke pri 95% stopnji zaupanja. Poišči 95% interval zaupanja za delež odraslih žensk, ki menijo, da nimajo dovolj časa zase.

(b) Razloži nekomu, ki ne zna statistike, zakaj ne moremo trditi, da 47% odraslih žensk nima časa zase.

(c) Nazadnje jasno razloži, kaj pomeni "pri 95% stopnji zaupanja".

(44) Kadar je statistika, ki ocenjuje neznani parameter, normalno porazdeljena, ima 95% interval zaupanja za parameter obliko ocena $\pm 2\sigma_{\text{ocena}}$. Pri kompleksno načrtovanih vzorcih zahtevajo ocene in standardni odkloni zamotane izračune. Vendar pa lahko v primerih, ko sta dana ocena in standardni odklon, izračunamo interval zaupanja za μ, ne da bi pri tem poznali formule, ki so pripeljale do teh vrednosti.

Poročilo, ki temelji na redni študiji prebivalstva, ocenjuje, da je bila stopnja nezaposlenosti v ZDA junija 1998 4,7%. (To pomeni, da je bilo 4,7% civilistov, starejših od 16 let, ki so iskali zaposlitev, nezaposlenih.) Poročilo navaja tudi, da je bil standardni odklon približno 0,11%. Vzorec, ki ga uporabljajo pri teh raziskavah, je kompleksen večstopenjski vzorec približno 50 000 gospodinjstev. Vzorčna porazdelitev ocenjene stopnje nezaposlenosti je približno normalna. Podaj 95% interval zaupanja za stopnjo nezaposlenosti v tej populaciji.

(45) Enostavni slučajni vzorec študentov na univerzi Upper Wabash Tech so vprašali, če se strinjajo z omejitvijo vpisa pri bolj obleganih predmetih, katere namen je obdržati visoko kakovost predavanj. Študentska organizacija sumi, da tak načrt ne bo všeč brucem, ki še niso vpisali glavnih predmetov. V tabeli so podani odgovori, ki so jih dobili.

	Za	**Proti**
Bruc	40	160
Absolvent	80	20

(a) Podaj 95% interval zaupanja za delež brucev, ki podpirajo omejitev.

(b) Podaj 95% interval zaupanja za delež absolventov, ki podpirajo omejitev.

(46) Žveplove spojine povzročajo neprijeten vonj v vinu, zato želijo vinarji poznati *zaznavni prag*, najmanjšo koncentracijo, ki jo človek še lahko zazna. Zaznavni prag za dimetil sulfid (DMS) pri poklicnih poskuševalcih vina je približno 25 mikrogramov na liter vina (μg/l). Nevajeni nosovi potrošnikov pa so ponavadi manj občutljivi. Tole so podatki o zaznavnem pragu za 10 študentov:

31	31	43	36	23
34	32	30	20	24

Predpostavi, da je standardni odklon zaznavnega pragu pri nevajenih ljudeh $\sigma = 7$ μg/l.

(a) Izdelaj stebelni diagram in se tako prepričaj, da je porazdelitev približno simetrična in brez ubežnikov. (Večje število podatkov potrdi, da ni nobenih sistematičnih odstopanj od normalnosti.)

(b) Podaj 95% interval zaupanja za povprečni zaznavni prag DMS med študenti.

(c) Ali lahko z gotovostjo trdiš, da je povprečni zaznavni prag pri študentih višji kot 25 μg/l? Zakaj?

(47) V ZDA so zakladne menice varna naložba, seveda pa nas zanima, kolikšni so dobički. V tabeli so zbrani podatki o skupnem donosu menic (v %) med leti 1970 in 1996.

Leto	1970	1971	1972	1973
Dobiček	6,45	4,37	4,17	7,20
Leto	1974	1975	1976	1977
Dobiček	8,00	5,89	5,06	5,43
Leto	1978	1979	1980	1981
Dobiček	7,46	10,56	12,18	14,71
Leto	1982	1983	1984	1985
Dobiček	10,84	8,98	9,89	7,65
Leto	1986	1987	1988	1989
Dobiček	6,10	5,89	6,95	8,43
Leto	1990	1991	1992	1993
Dobiček	7,72	5,46	3,50	3,04
Leto	1994	1995	1996	
Dobiček	4,37	5,60	5,13	

(a) Napravi histogram teh podatkov. Stolpci naj bodo široki dve odstotni točki. Kakšno odstopanje od normalnosti opaziš? Zaradi centralnega limitnega izreka pa lahko kljub temu obravnavamo porazdelitev $\overline{x}$ kot normalno.

(b) Predpostavimo, da lahko obravnavamo podatke za teh 27 let kot slučajni vzorec donosov zakladnih menic. Podaj 95% interval zaupanja za dolgoročni povprečni donos. (Predpostavi, da poznaš standardni odklon donosov, $\sigma = 2{,}75\%$.)

(c) Stopnja inflacije v tem času je bila povprečno 5,5%. Ali smo lahko prepričani, da je povprečni donos zakladnih menic večji od 5,5%? Zakaj?

(48) Uporabili smo vzorčno porazdelitev $\hat{p}$ in pravilo 68–95–99,7, da smo poiskali 95% interval zaupanja za delež populacije p.

(a) Natančno razloži, zakaj je

$$\hat{p} \pm \sqrt{\frac{\hat{p}(100-\hat{p})}{n}}$$

68% interval zaupanja za p.

(b) Izpelji formulo za 99,7% interval zaupanja za p.

(49) Uporabi nalogo 48 (a) in podatke iz naloge 12 ter podaj 68% interval zaupanja za delež obiskovalcev Yellowstonskega parka, ki podpirajo omejitev števila vozil v parku. Primerjaj dolžino 68% intervala z dolžino 95% intervala iz naloge 12 in pojasni razliko.

(50) Uporabi nalogo 48 (b) in podatke iz naloge 9 ter podaj 99,7% interval zaupanja za delež vseh odraslih, ki podpirajo enojezične šole. Primerjaj dolžino dobljenega intervala z dolžino 95% intervala zaupanja iz naloge 10. Kaj je razlog za razliko v dolžinah?

(51) Zgornji in spodnji decil vsake normalne porazdelitve se nahajata na razdalji 1,28 standardnega odklona pod in nad srednjo vrednostjo. (Spodnji decil je točka, pod katero je verjetnost 10%, zgornji decil pa tista, pod katero je verjetnost 90%.)

(a) Uporabi to dejstvo ter izpelji formulo za 80% interval zaupanja za delež populacije p na osnovi deleža vzorca $\hat{p}$, ki je točen za velike velikosti vzorca n.

(b) Podaj 80% interval zaupanja za delež obiskovalcev iz naloge 12, ki podpirajo omejitve.

(52) Zgornji in spodnji decil vsake normalne porazdelitve se nahajata na razdalji 1,28 standardnega odklona pod in nad srednjo vrednostjo.

(a) Podaj formulo za 80% interval zaupanja za srednjo vrednost μ normalne populacije s pomočjo vzorčnega povprečja $\overline{x}$ enostavnega slučajnega vzorca velikosti n.

(b) Podaj 80% interval zaupanja za povprečen IQ iz naloge 24.

(53) Stopnja donosa delnic se spreminja iz meseca v mesec. Lahko si pomagamo s kontrolnim diagramom, da vidimo, če je vzorec spreminjanja v času stabilen ali pa so morda obdobja, v katerih so delnice posebej nestabilne v primerjavi s svojim lastnim dolgoročnim vzorcem. Spodaj so povprečne mesečne stopnje donosov (v %) Wal-Martovih delnic za 38 šestmesečnih obdobij, urejene od leve proti desni po vrsticah. Podatki zajemajo obdobje od januarja 1973 do decembra 1990.

-11,78	1,68	7,88	-11,01	19,72
1,66	1,13	2,70	-0,60	5,91
2,95	0,05	1,88	6,46	2,25
8,05	4,10	2,08	4,02	11,36
8,13	0,12	1,30	-1,27	6,59
3,01	8,68	-1,52	6,64	-3,20
2,92	0,63	3,49	2,94	5,86
-0,25	6,02	5,87		

Ta števila obravnavamo kot povprečja $\overline{x}$ iz 38 vzorcev, od katerih je vsak vključeval 6 vrednosti. Narišemo jih na kontrolni diagram $\overline{x}$.

(a) Za središčno črto uporabi povprečje vseh $\overline{x}$, ki ga ponavadi v kontroli kakovosti označijo z $\overline{\overline{x}}$. To je isto kot povprečje 228 posameznih mesečnih donosov.

(b) Standardni odklon 228 mesečnih donosov vključuje tako dolgoročne kot kratkoročne trende in je v splošnem prevelik za uspešno kontrolo. Namesto tega povprečimo standardne odklone 38 vzorcev in dobimo $\overline{s} = 9450$. Predpostavi, da je $\overline{\overline{x}}$ približek za μ in $\overline{s}$ približek za σ. Določi mejni vrednosti za kontrolni diagram $\overline{x}$.

(c) Izdelaj kontrolni diagram. Ali opaziš kakšno zunanjo točko ali pa niz devetih? Ali so bila v teh 19 letih obdobja, ko so donosi ušli izpod kontrole?

(54) Spodaj so dane vsote po vrsticah in stolpcih dvosmerne tabele z dvema vrsticama in dvema stolpcema.

a	b	50
c	d	50
60	40	100

Poišči dva *različna* nabora števil a, b, c in d, ki bi dala enake vsote po vrsticah in stolpcih. Ta primer nam pokaže, da ne moremo razbrati zveze med dvema spremenljivkama le iz posameznih porazdelitev.

(55) Nedavne študije so pokazale, da so prejšnja poročila podcenila tveganje, ki ga predstavlja prekomerna telesna teža. Do napake je prišlo, ker so spregledali skrite spremenljivke. Tako na primer kajenje hkrati zmanjšuje težo in vodi v zgodnjo smrt. Ilustriraj nevarnosti, ki jih predstavljajo skrite spremenljivke, s poenostavljeno verzijo zgornje situacije. Napravi tabelo, ki bo vključevala prekomerno težo (da ali ne), zgodnjo smrt (da ali ne) in kadilske navade (da ali ne), poleg tega pa bo iz nje razvidno naslednje:

- Kadilci in nekadilci s prekomerno telesno težo umrejo prej kot tisti, ki niso predebeli.
- Ko kadilce in nekadilce združimo v dvosmerno tabelo s podatki o teži in umrljivosti, vidimo, da tisti z normalno težo umrejo prej.

4.10 Tehnološki kotiček

Testiranje generatorja naključnih števil

Velikokrat smo se že zanašali na generator naključnih števil v programu za delo s preglednicami kot nadomestek za pošteni kovanec, pošteno kocko ali dobro premešan kupček kart. Ali so bile te simulacije res poštene in nepristranske? Na to vprašanje lahko odgovorimo z uporabo vzorcev in intervalov zaupanja.

Met kovanca simuliramo tako, da izberemo slučajno število med 0 in 1, kar storimo z ukazom `=RandBetween(0,1)`. Če predpostavimo, da je program pošten, potem je verjetnost, da izbere 1, enaka $p = 0{,}5$. Če izberemo 10 vzorcev, je v 95% primerov dobljeni $\hat{p}$ v 95% intervalu zaupanja, $0{,}5 \pm 0{,}316$, torej med 0,184 in 0,816. To pomeni, da bi med 2 in 8 vzorcev moralo biti vsaj v 95% primerov enakih 1.

Naloga 1. Izvedi eksperiment, ki smo ga pravkar opisali. Izračunaj vsoto 10 števil, naključno generiranih z ukazom `=RandBetween(0,1)`. Ponovi ta eksperiment skupno dvajsetkrat. Kolikokrat je število enic padlo iz sprejemljivih mej?

Naloga 2. Če izberemo večje vzorce, se velikost 95% intervala zaupanja zmanjša. Povečaj velikost vzorca na 100 in izračunaj pripadajoči interval zaupanja. Nato izračunaj vsoto 100 naključnih števil, dobljenih z zgoraj navedenim ukazom. Ponovi ta eksperiment dvajsetkrat. Kako pogosto je število enic padlo iz sprejemljivih mej?

Ocena μ z uporabo vzorčnega povprečja $\overline{x}$

Velikokrat uporabljamo vzorčno povprečje $\overline{x}$ za oceno dejanske srednje vrednosti populacije μ. V preglednici na sliki 4.7 je naštetih 5 tež vzorcev (v gramih) za nek izdelek. Vsaka posamezna teža je natančna s standardnim odklonom $\sigma = 0{,}05$ g. Na osnovi teh vrednosti izračunamo vzorčno povprečje in vzorčni standardni odklon $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$. Krajišči 95% intervala zaupanja za srednjo vrednost dobimo tako, da prištejemo in odštejemo $\sigma_{\overline{x}}$ vzorčnemu povprečju.

Naloga 3. Napravi tabelo, podobno tisti na sliki 4.7, za naslednjih pet podatkov:

38,82 38,76 38,75 38,75 38,81

	A	B	C	D
1	38,85		sigma	0,05
2	38,76		vzorčno povp.	38,804
3	38,80		n	5
4	38,83		sigma/sqrt(n)	0,022361
5	38,78			
6			interval min	38,75928
7	194,02	vsota	interval max	38,84872

Slika 4.7: Izračun intervala zaupanja s pomočjo preglednice.

Naloga 4. Uporabi funkcijo `= 25,5 + RandBetween(5,25)/50`, da generiraš nove "meritve". Standardni odklon je približno $\sigma = 0{,}1$. Napravi pet vzorcev in izračunaj 95% interval zaupanja za "dejansko" težo.

Raziskovanje

Recimo, da moraš po pogodbi zagotoviti, da je vzorčna varianca $\sigma_{\overline{x}}$ manjša od 0,01. Vsako tehtanje te stane nekaj časa in denarja, vendar pa so bolj natančne tehtnice ponavadi bolj drage. Kako poiskati kompromis med stroški, ki jih predstavlja večje število tehtanj, in ceno bolj natančne tehtnice? Zamisli si scenarij, v katerem se lotiš tega problema.

4.11 Pisni projekti

(1) Kako agencije opisujejo natančnost svojih raziskav? Pod žarometom je opisano, kako so to naredili pri New York Timesu. Pojdi na spletno stran agencije Gallup (`www.gallup.com`) in preberi objavo kakšne nedavne ankete. Nato pojdi na stran Urada za statistiko dela (`stats.bls.gov/newsrels.htm`) in poišči najnovejše podatke o zaposlenosti (*Employment Situation*) v razdelku *Employment & Unemployment*.

V obeh objavah poišči tisti del, ki opisuje metode vzorčenja, mejo napake rezultata in vire napake, ki v to mejo niso vključni. V drugem primeru se posveti samo podatkom iz raziskave prebivalstva (*Current Population Survey*). Napiši kratko primerjavo obeh primerov. Zakaj je po tvojem mnenju Gallup precej bolj redkobeseden kot Urad za statistiko dela?

(2) Meja napake za raziskavo na vzorcu upošteva naključne variacije, ki so posledica slučajnega vzorčenja. V praksi pa so lahko rezultati napačni iz drugih razlogov. Nekaterih izbranih ni mogoče kontaktirati, drugi se zlažejo ali pa se ne spomnijo, formulacija vprašanj vpliva na odgovore. To so "praktične ovire", ki jih omenjamo pod žarometom.

Napiši kratko razpravo o najbolj pomebnih praktičnih ovirah, na katere naletimo pri anketiranju in drugih raziskavah na človeški populaciji. Nekaj materiala je v *Statistics: Concepts and Controversies* (pod priporočeno literaturo v prvem poglavju) in v članku P.E. Converse, M.W. Traugott, Assessing the accuracy of polls and surveys, *Science*, 234 (1986): 1094–1098.

(3) Intervalov zaupanja ne uporabljamo le v situacijah, ki smo jih opisali v tem poglavju. Eden od primerov je ocenjevanje razlike med dvema deležema populacije, p_1 in p_2. Na primer, p_1 bi lahko bil delež žensk in p_2 delež moških, ki ponoči ostajajo doma, ker jih je strah kriminala. Preberi poročilo o intervalih zaupanja za $p_1 - p_2$ v knjigah o statističnih metodah, na primer v razdelku 8.2 Moorove *Basic Practice of Statistics*. Nato natančno razloži, kako dobimo ta novi interval s pomočjo enakega sklepanja kot tiste, ki smo jih že spoznali: poišči oceno, opazi, da je vzorčna porazdelitev ocene približno normalna, ugotovi, kakšen je standardni odklon te normalne porazdelitve in se za dva standardna odklona odmakni od ocene, da dobiš 95% interval zaupanja.