II STATISTIKA

Predstavili bomo osnove opisne oziroma deskriptivne statistike, ki se ukvarja z zbiranjem in opisovanjem podatkov ter nekatere elemente inferenčne oziroma matematične, analitične, sklepne, indukcijske statistike, ki nas uči, kako o celoti sklepamo iz podatkov dobljenih na manjših vzorcih, ki so le del celote.

Kazalo

Uvod

- Pridobivanje podatkov
- Obdelava podatkov
- Statistično sklepanje

Osnovni pojmi opisne statistike

Slikovni prikazi

Mere za povprečje in razpršenost

Shanonov indeks heterogenosti

Različne matematične sredine

Vzorčenje

Cenilke oziroma statistike

- Vzorčno povprečje
- Vzorčna varianca
- Vzorčna relativna frekvenca
- Razlika povprečij dveh populacij

Razlika povprečij dveh populacij

Porazdelitveni zakon srednje vrednosti vzorcev

Intervalne cenilke

- Intervalna cenilka srednje vrednosti
- Intervalna cenilka relativne frekvence
- Intervalna razlika povprečij dveh populacij

Preverjanje hipotez oziroma domnev

- Formalen postopek za testiranje hipotez
- Test za primerjanje porazdelitvenih zakonov
- Test za primerjanje razlike povprečij

UVOD

Statistika je znanost zbiranja, organiziranja in interpretiranja numeričnih dejstev, ki jih imenujemo podatki. Podatki so pomembni pri delu mnogih, zato je izobraževanje na področju statistike izredno pomembno pri številnih poklicih. Ekonomisti, finančni svetovalci, vodstveni kader v politiki in gospodarstvu preučujejo najnovejše podatke o nezaposlenosti in inflaciji. Zdravniki morajo razumeti izvor in zanesljivost podatkov, ki so objavljeni v medicinskih revijah. Poslovne odločitve so običajno zasnovane na raziskavah tržišč, ki razkrijejo želje kupcev in njihovo obnašanje. Večina akademskih raziskav uporablja številke in izkorišča statistične metode.

Tako kot so besede na papirju brez pomena za nepismenega ali slabo izobraženega človeka, tako so lahko tudi podatki privlačni, zavajajoči ali enostavno nesmiselni. Statistična pismenost, to je sposobnost sledenja in razumevanja argumentov, ki izhajajo iz podatkov, je pomembna za sleherno osebo.

Pridobivanje podatkov

Novice so polne številk. Televizijski napovedovalec pove, da se je stopnja nezaposlenosti zmanjšala na 4,7%. Raziskava trdi, da je 45% Američanov zaradi kriminala strah ponoči zapustiti domove. Od kod pridejo te številke? Ne vprašamo vseh ljudi, če so zaposleni ali ne. Raziskovalne agencije vprašajo le nekaj posameznikov, če zaradi strahu pred ropi ostajajo ponoči doma. Vsak dan se v novicah pojavi nov naslov. Eden od teh trdi: Aspirin preprečuje srčne infarkte. Nadaljnje branje razkrije, da je raziskava obravnavala 22 tisoč zdravnikov srednjih let. Polovica zdravnikov je vsak drugi dan vzela aspirin, druga polovica pa je dobila neaktivno tableto. V skupini, ki je jemala aspirin, je 139 zdravnikov doživelo srčni infarkt. V drugi skupini je bilo v enakem časovnem obdobju 239 infarktov. Ali je ta razlika dovolj velika, da lahko trdimo, da aspirin res preprečuje srčne infarkte? Da bi ubežali neprijetnostim kot sta nezaposlenost in srčni infarkt, prižgimo televizijo. V pogovorni oddaji voditelj povabi gledalce, da sodelujejo v anketi. Tema pogovora je dobrodelnost in voditelja zanima, če gledalci redno prispevajo denar ali oblačila v dobrodelne namene. Med oddajo sprejmejo 50 tisoč klicev in 83% gledalcev trdi, da redno sodelujejo v tovrstnih akcijah. Ali je res, da smo tako zelo humanitarno osveščeni? Zanesljivost teh številk je v prvi vrsti odvisna od njihovega izvora. Podatkom o nezaposlenosti lahko zaupamo, v tistih 83% iz pogovorne oddaje pa najbrž lahko utemeljeno podvomimo. Naučili se bomo prepoznati dobre in slabe metode pridobivanja podatkov. Razumevanje metod, s katerimi lahko pridobimo zaupanja vredne podatke, je prvi (in najpomembnejši) korak k pridobivanju sposobnosti odločanja o pravilnosti sklepov, ki jih izpeljemo na osnovi danih podatkov. Izpeljava zaupanja vrednih metod za pridobivanje podatkov je področje, kjer vstopimo v svet statistike, znanosti o podatkih.

Obdelava podatkov

Za sodobno družbo je značilna poplava podatkov. Podatki, ali numerična dejstva, so bistveni pri odločanju na skoraj vseh področjih življenja in dela. Kot druge velike poplave nam poplava podatkov grozi, da nas bo pokopala pod sabo. Moramo jo kontrolirati s premišljeno organizacijo in interpretacijo podatkov. Baza podatkov kakšnega podjetja na primer vsebuje velikansko število podatkov: o zaposlenih, prodaji, inventarju, računih strank, opremi, davkih in drugem. Ti podatki so koristni le v primeru, ko jih lahko organiziramo in predstavimo tako, da je njihov pomen jasen. Posledice neupoštevanja podatkov so lahko hude. Veliko bank je izgubilo na milijarde dolarjev pri nedovoljenih špekulacijah njihovih zaposlenih, ki so ostale skrite med goro podatkov, ki jih odgovorni niso dovolj pozorno pregledali.

Verjetnost: matematika naključij

Ste se kdaj vprašali, zakaj so igre na srečo, ki so za nekatere rekreacija ali pa droga, tako dober posel za igralnice? Vsak uspešen posel mora iz uslug, ki jih ponuja, kovati napovedljive dobičke. To velja tudi v primeru, ko so te usluge igre na srečo. Posamezni hazarderji lahko zmagajo ali pa izgubijo. Nikoli ne morejo vedeti, če se bo njihov obisk igralnice končal z dobičkom ali z izgubo. Igralnica pa ne kocka, pač pa dosledno dobiva in država lepo služi na račun loterij in drugih oblik iger na srečo. Presenetljivo je, da lahko skupni rezultat več tisoč naključnih izidov poznamo s skoraj popolno gotovostjo. Igralnici ni potrebno obtežiti kock, označiti kart ali spremeniti kolesa rulete. Ve, da ji bo na dolgi rok vsak stavljeni evro prinesel približno pet centov dobička. Splača se ji torej osredotočiti na brezplačne predstave ali poceni avtobusne vozovnice, da bi privabili več gostov in tako povečali število stavljenega denarja. Posledica bo večji dobiček. Igralnice niso edine, ki se okoriščajo z dejstvom, da so velikokratne ponovitve slučajnih izidov napovedljive. Na primer, čeprav zavarovalnica ne ve, kateri od njenih zavarovancev bodo umrli v prihodnjem letu, lahko precej natančno napove, koliko jih bo umrlo. Premije življenjskih zavarovanj postavi v skladu s tem znanjem, ravno tako kot igralnica določi glavne dobitke.

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. Po domače statistično sklepanje velikokrat temelji na grafični predstavitvi podatkov. Formalno sklepanje, tema tega predmeta, uporablja verjetnost, da pove, do kakšne mere smo lahko prepričani, da so naši zaključki pravilni.

Statistika je veda, ki preučuje množične pojave. Ljudje običajno besedo statistika povezujejo z zbiranjem in urejanjem podatkov o nekem pojavu, izračunom raznih značilnosti iz teh podatkov, njih predstavitvijo in razlago. To je najstarejši del statistike in ima svoje začetke že v antiki – z nastankom večjih združb (držav) se je pojavila potreba po poznavanju stanja – 'računovodstvo', astronomija, . . . Sama beseda statistika naj bi izvirala iz latinske besede *status* – v pomenu država. Tej veji statistike pravimo opisna statistika.

Druga veja, inferenčna statistika, poskuša spoznanja iz zbranih podatkov posplošiti (razširiti, podaljšati, napovedati, . . .) in oceniti kakovost teh posplošitev. Inferenčno statistiko lahko vidimo tudi kot drugo plat verjetnostnega računa: le-ta računa verjetnostni kompleksnih dogodkov, ko pozna verjetnostno porazdelitev slučajnih spremenljivk; statistika pa iz verjetnosti posameznih dogodkov skuša razbrati porazdelitev slučajnih spremenljivk.

Osnovni pojmi opisne statistike

Statistična enota – posamezna proučevana stvar ali pojav. Primer: redni študent na Univerzi v Ljubljani v študijskem letu 2008/09.

Populacija – množica vseh proučevanih enot; pomembna je natančna opredelitev populacije (npr. časovno in prostorsko). Primer: vsi redni študentje na UL v študijskem letu 2008/09

Vzorec – podmnožica populacije, na osnovi katere ponavadi sklepamo o lastnostih celotne populacije. Primer: vzorec 300 slučajno izbranih rednih študentov na UL v l. 2008/09.

Spremenljivka – lastnost enot; označujemo jih npr. z X, Y , Z. Primer: spol, uspeh iz matematike v zadnjem razredu srednje šole.

Parameter – značilnost populacije; običajno jih označujemo z malimi grškimi črkami.

Statistika – značilnost vzorca; običajno jih označujemo z malimi latinskimi črkami. Vrednost statistike je lahko za različne vzorce različna.

Eno izmed osnovnih vprašanj statistike je, kako z uporabo ustreznih statistik oceniti vrednosti izbranih parametrov.

Slikovni prikazi

Stolpčni prikaz: Na eni osi prikažemo (urejene) razrede. Nad vsakim naredimo stolpec/črto višine sorazmerne frekvenci razreda.

Krožni prikaz: Vsakemu razredu priredimo krožni izsek

Histogram: drug poleg drugega rišemo stolpce – pravokotnike, katerih ploščina je sorazmerna frekvenci v razredu. Če so razredi enako široki, je višina sorazmerna tudi frekvenci.

Poligon: v koordinatnem sistemu zaznamujemo točke (x_i, f_i), kjer je x_i sredina i–tega razreda in f_i njegova frekvenca. Točke zvežemo z daljicami.

Glej Excel

Mere za povprečje in razpršenost

Naj imata populacija N elementov in vzorec n elementov. Lahko opredelimo:

Povprečje populacije

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Povprečje vzorca

$$\bar{X}_n = \frac{\sum_{i=1}^{n} x_i}{n}$$

Varianca populacije

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Vzorčna varianca

$$S_n^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{X}_n)^2}{n}$$

Prilagojena vzorčna varianca

$$\hat{S}_n^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{X}_n)^2}{n-1}$$

Standardni odklon oziroma standardna deviacija

Standardni odklon je pozitivni kvadratni koren variance.

Koeficient razpršenosti

Uporabljamo tudi koeficient razpršenosti, ki je razmerje med standardnim odklonom in srednjo vrednostjo $\frac{\sigma}{\mu}$. Le ta je koristen pri primerjanju različnih populacij.

Shanonov indeks heterogenosti

Zanimivo merilo heterogenosti populacije je nedvomno indeks, ki ga je prvi obravnaval Shannon (1948) v svoji teoriji informacije in ki ga v fiziki uporabljamo za merjenje entropije.

$$S = -\sum_{i=1}^{N} p_i \log_a(p_i)$$

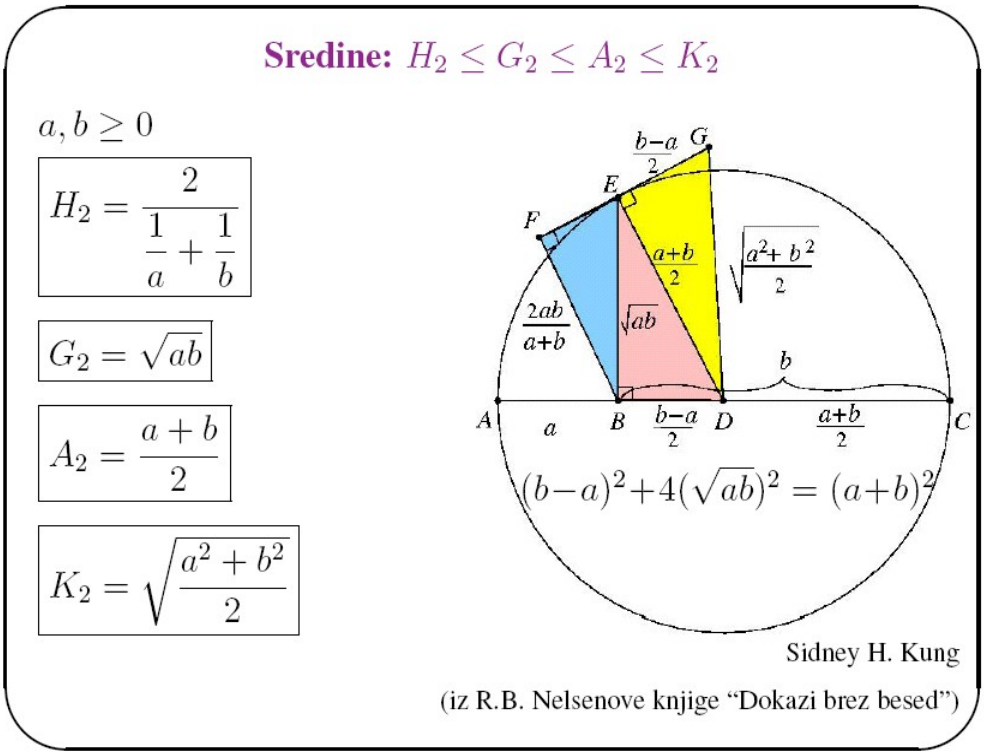
kjer so p_i relativne frekvence. V primeru, da so relativne frekvence med sabo enake $p_i = \frac{1}{N}$, sledi da je indeks maksimalen in enak

$$S = -\sum_{i=1}^{N} \frac{1}{N} \log_a\left(\frac{1}{N}\right) = -N \frac{1}{N} \log_a\left(\frac{1}{N}\right) = \log_a(N).$$

Indeks meri količino informacije, ki je potrebna, da nam je rezultat stohastičnega pojava popolnoma znan: tako bo poštena kocka imela velik Shannonov indeks; kocka, ki nam da skoraj vedno dve piki, pa zelo majhen indeks. Majhen Shannonov indeks populacije pomeni veliko homogenost le-te. Zanimivo je računanje indeksa, ko primerjamo letni dohodek državljanov v različnih državah, saj državi z enakima povprečnima dohodkoma lahko imata zelo različna indeksa.

Različne matematične sredine

A. Jurišić in V. Batagelj: Verjetnostni račun in statistika



VZORČENJE

Analitična statistika je veja statistike, ki se ukvarja z uporabo vzorčnih podatkov, da bi z njimi naredili zaključek (inferenco) o populaciji.

Zakaj vzorčenje?

- cena
- čas
- destruktivno testiranje

Glavno vprašanje statistike je: kakšen mora biti vzorec, da lahko iz podatkov zbranih na njem veljavno sklepamo o lastnostih celotne populacije. Kdaj vzorec dobro predstavlja celo populacijo?

Preprost odgovor je:

- vzorec mora biti izbran nepristransko,
- vzorec mora biti dovolj velik.

Omenili bomo dve vrsti vzorčenja:

Bernoullijevo vzorčenje, kjer iz populacije velikosti N na slepo žrebamo *n* elementov z vračanjem in vsi elementi imajo enako verjetnost, da bodo izbrani. Populacija velikosti *N* ima N^n vzorcev z *n* elementi.

Blok vzorčenje, kjer na slepo izberemo množico *n-tih* elementov iz populacije brez vračanja. Populacija velikosti *N* ima

$$\frac{N!}{n!(N-n!)}$$

različnih vzorcev.

CENILKE OZIROMA STATISTIKE

Parameter je število, ki sintetično opisuje lastnosti populacijo: na primer srednja vrednost in varianca populacije sta parametra populacije. Ponavadi parametrov ne poznamo, ker je populacija prevelika, zato poiščemo ustrezne lastnosti vzorcev, **cenilke**, in iz njih skušamo dobiti parametre populacije. Vrednost cenilk je lahko za različne vzorce različna. Med vsemi cenilkami bomo obravnavali: vzorčno povprečje oziroma aritmetično srednjo vrednost vzorca, vzorčno varianco, vzorčno relativno frekvenco in razliko povprečij vzorcev.

Če ima populacija *N* elementov $\{x_{1,} x_{2,...}, x_N\}$, si lahko predstavljamo izbiro enega elementa iz populacije, kot slučajno spremenljivko X, ki ima kot zalogo vrednosti elemente populacije, njena porazdelitev pa je odvisna od načina vzorčenja; za Bernoullijevo vzorčenje, je porazdelitev spremenljivke konstantna. Izbiro vzorca z *n* elementi si predstavljamo kot zaporedje neodvisnih slučajnih spremenljivk $X_{1,} X_{2,...} X_n$. Slučajnim spremenljivkam *X* pravimo **vzorčne slučajne spremenljivke.**

Glej Excel.vzorci

Statistika neke populacije je vsaka funkcija vzorčnih slučajnih spremenljivk $X_{1,} X_{2,\ldots} X_n$

$$T_n = f(X_{1,} X_{2,\ldots}, X_n)$$

Cenilke so vrednosti slučajne spremenljivke T_n

T_n je slučajna spremenljivka in torej lahko govorimo o njenem upanju. Pravimo, da je **statistika nekega parametra** φ **pravilna** , če je matematično upanje statistike enako parametru populacije, oziroma, če

$$E[T_n] = \varphi$$

Se lahko dokaže, da je vzorčna srednja vrednost pravilna statistika za srednjo vrednost populacije, vzorčna standardna deviacija pa ni pravilna statistika za standardno deviacijo populacije in jo je potrebno primerno popraviti.

Vzorčno povprečje

Če opredelimo statistiko **aritmetična srednja vrednost** kot

$$\bar{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

lahko izračunamo upanje te spremenljivke, in sicer

$$E[\bar{X}] \frac{1}{n}(E[X_1] + E[X_2] + \ldots + E[X_n]) = \mu$$

zato lahko rečemo, da je aritmetična srednja vrednost vzorca pravilna statistika populacije.

Vzorčna varianca

Če sedaj opredelimo statistiko **varianca** kot

$$S_n^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2}{n} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$$

pa se pokaže, da v primeru Bernoullijevega vzorčenja dobimo

$$E[S_n^2] = \frac{(n-1)}{n} \sigma^2 \quad ;$$

v primeru blok vzorčenja pa dobimo

$$E[S_n^2] = \frac{(n-1)}{n} \frac{N}{N-1} \sigma^2 \, ,$$

zato v obeh primerih varianca, ni pravilna statistika parametra, v kolikor njeno upanje ni enako varianci populacije. Treba je opredeliti **prilagojeno varianco** in sicer:

$$\hat{S}_n^2 = \frac{n}{n-1} S_n^2$$

za Bernoullijevo vzorčenje in

$$\hat{S}_n^2 = \frac{n}{n-1} \frac{N-1}{N} S_n^2$$

za blok vzorčenje.

Glej Excel.varianca

Iz poglavij verjetnostnega računa, vemo že, da je

$$\sigma^2 = V[X] = E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 =$$
$$= E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$$

zato lahko izračunamo varianco vzorcev tudi na ta način in jih nato ustrezno popravimo.

Pomemben indikator statistike je tudi njena varianca, ki meri razpršenost statistike v populaciji. Se lahko dokaže, da je varianca statistike srednja vrednost vzorca enaka

$$V[\bar{X}_n] = \frac{\sigma^2}{n}$$

za Bernoullijevo vzorčenje in

$$V[\bar{X}_n] = \frac{\sigma^2}{n} \frac{N-n}{N-1}$$

za blok vzorčenje.

Vzorčna relativna frekvenca

Tudi slučajna spremenljivka $F_n = \frac{\text{št. Koristnih izidov}}{n} = \frac{k}{n}$, ki predstavlja **relativno frekvenco** nekega izida v vzorcu, je pravilna statistika; to pomeni, da $E[F_n] = p$, kjer je p relativna frekvenca izidov v populaciji. Dokazali bomo trditev v primeru Bernoullijevega vzorčenja. Če je p verjetnost, da se uresniči koristen izid v populaciji in *(1-p)* verjetnost, da se ne uresniči, potem je v vsakem vzorcu z n elementi verjetnost, da dobimo k koristnih izidov enaka

$$P(k) = \binom{n}{k} p^k (1-p)^k$$

in $E[K]$ =np ter $V[K] = np(1-p)$. Zato za spremenljivko relativna frekvenca velja

$$E[F_n] = E\left[\frac{K}{n}\right] = E\frac{[K]}{n} = p \quad \text{in} \quad V[F_n] = V\left[\frac{K}{n}\right] = V\frac{[K]}{n^2} = \frac{p(1-p)}{n.}$$

Relativna frekvenca vzorca je torej korektna statistika za parameter relativne statistike populacije in standardna deviacija statistike je $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{p(1-p)}{n}}$.

Primer: Na šoli z N=200 dijaki slučajno izberemo vzorec z n=50dijaki in po opravljeni anketi ugotovimo, da jih 10 kadi. Lahko zaključimo, da je relativna frekvenca kadilcev populacije $p = \frac{10}{50} = 0{,}2$ in standardna napaka $\sigma = \sqrt{\frac{0{,}2(1-0{,}2)}{50}} = 0{,}056 = 5{,}6\%$. Iz tega sledi, da si z veliko verjetnostjo v populaciji šole lahko pričakujemo, da je število kadilcev med 37 in 42.

Razlika povprečnih vrednosti dveh vzorcev populacije

Razlika povprečij dveh populacij

Pogosto je zanimivo primerjati dve populaciji glede nekega skupnega parametra, recimo srednje vrednosti. Če sta dani populaciji X in Y z ustreznima parametroma srednja vrednost μ_1 in μ_2 ter ustreznima variancama σ_1^2 in σ_2^2, ter z ustreznima Bernoullijevima vzorčenjema z n_1 in n_2 elementi, potem lahko opredelimo vzorčni povprečji $E[\bar{X}_{n_1}]$ in $E[\bar{Y}_{n_2}]$. Se lahko dokaže, da velja

$$E[\bar{X}_{n_1}-\bar{Y}_{n_2}]=\mu_1-\mu_2$$
$$V[\bar{X}_{n_1}-\bar{Y}_{n_2}]=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

TOČKASTE CENILKE

Iz prejšnjega poglavja sledi, da srednja vrednost vzorca, prilagojena varianca vzorca in relativna frekvenca vzorca so pravilne statistike za ustrezne parametre populacije. Pravimo jim **točkaste cenilke** v kolikor ocenimo le vrednost in standardno napako, Če želimo poznati tudi, verjetnost, da bo ocenjena vrednost padla v določeno območje, se bomo posluževali intervalnih cenilk.

PORAZDELITVENI ZAKON SREDNJE VREDNOSTI VZORCEV

Izrek: Naj ima populacija katerokoli porazdelitev s srednjo vrednostjo μ in varianco σ^2. Potem sledi, da slučajna spremenljivka $\bar{X}_n$ srednja vrednost je porazdeljena normalno z upanjem μ in varianco $\frac{\sigma^2}{n}$.

Dokaz: (direktna posledica centralnega limitnega izreka)

Tale izrek ima zanimive posledice v statistiki; oglejmo si primer.

Primer: Vemo, da na določenem teritoriju imajo žitna polja srednjo vrednost produktivnosti na hektar $\mu=25\text{q}$ z varianco $\sigma=4\text{q}$ .Določi verjetnost, da na slučajno izbranem vzorcu desetih polj je srednja vrednost produktivnosti večja od 27q.

Prejšnji izrek nas uči, da je slučajna spremenljivka aritmetična srednja vrednost vzorca porazdeljena normalno z upanjem μ in varianco $\frac{\sigma^2}{n}$. Vemo, da je za Bernoullijevo vzorčenje upanje srednjih vrednosti vzorcev enako srednji vrednosti populacije, zato $E[\bar{X}_{10}]=\mu=25$ in $V[\bar{X}_n]=\frac{\sigma^2}{10}=1,6$. Sledi da, da je standardni odklon srednje vrednosti vzorca $\sigma=1,26$. Uporabljali bomo standardizirano slučajno spremenljivko Z,

$$Z=\frac{\bar{X}_{10}-\mu}{\sigma}=\frac{\bar{X}_{10}-25}{1,26}$$

ki ima normalno porazdelitev *N(0,1)*. Torej

$$P(\bar{X}_{10}>27)=P(Z>1.58)=1-P(Z\le 1,58)=1-0,9429=0,0571\approx 6\%$$

Glej.geoGebra.normalna.html

Smo že ugotovili, da aritmetična srednja vrednost vzorca in popravljena varianca vzorca sta pravilna cenilca oziroma pravilni statistiki. Oglejmo si primer v Excelu.

Primer: Iz populacije 140 dijakov neke šole izberemo slučajno vzorec 50 dijakov in jih vprašamo, koliko ur posvetijo učenju. Rezultat je podan v tabeli

Ure učenja	1	2	3	4	5
Št. Dijakov	4	11	22	9	4

Določi srednjo vrednost vzorca, varianco in prilagojeno varianco, prilagojeno standardno deviacijo in standardno napako. Nato izračunaj še kakšna je verjetnost, da bi dobili v slučajnem vzorcu desetih dijakov aritmetično srednjo vrednost vzorca večjo od dva?

$$P(\bar{X}_{10}>2)=P(Z>-2{,}92)=1-P(Z<-2{,}92)=1-P(Z>2{,}92)=0{,}54=54\%$$

Glej.Excel.cenilci

INTERVALNE CENILKE

Do sedaj smo se ukvarjali s točkastimi cenilkami: srednja vrednost, varianca , relativna frekvenca. Sedaj se bomo vprašali, kako lahko s pomočjo cenilk vzorca, določimo v katerem intervalu bo ležal iskani parameter populacije z zaželeno verjetnostjo. Na primer lahko poiščemo v kateri interval bo padla srednja vrednost populacije z dano verjetnostjo *95%: t*o pomeni, da če izračunamo srednjo vrednost kateregakoli vzorca populacije, imamo verjetnost *95%*, da bo ta padla v ocenjeni interval. Za intervalno določanje cenilk , je potrebno poznati porazdelitveni zakon populacije.

Intervalna cenilka srednje vrednosti

Obravnavali bomo tri primere:

a) sta znani bodisi μ kot σ^2 populacije,

b) je znana le μ populacije in vzorec je dovolj velik (n>30), da lahko uporabljamo prilagojeno varianco $\hat{S}_n^2$;

c) je znana le μ populacije in vzorec je premajhen za prilagojeno varianco.

Navedene metode so koristne pri vrednotenju kakovosti produkcij tovarn in procesov, kjer poznamo lastnosti posameznih strojev in lahko izračunamo srednjo vrednost in varianco populacije izdelkov in nato iščemo interval srednje vrednosti *dobrega* vzorca.

a) Če je μ srednja vrednost populacije in σ^2 varianca populacije, vemo, da je srednja vrednost vzorca z n elementi porazdeljena normalno s srednjo vrednostjo μ in varianco $\frac{\sigma^2}{n}$. Sprašujemo se s katero verjetnostjo, bo srednja vrednost vzorca $\bar{X}_n$ ležala v intervalu $(\mu-\epsilon,\mu+\epsilon)$?

Standardizirana slučajna spremenljivka

$$Z=\frac{\bar{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}}$$

je porazdeljena *N(0,1)*. Če želimo zvedeti za verjetnost, da Z pade v interval $(-\epsilon,\epsilon)$, napišemo

$$P(-\epsilon<Z<\epsilon)=2\mathrm{P}(0<z<\epsilon)=2(P(Z<\epsilon)-\frac{1}{2})=\phi$$

oziroma

$$P(Z<\epsilon)=\frac{1+\phi}{2}$$

kjer je ϕ želena natančnost. Na primer, če je $\phi=0.95$, (95%) potem je

$$P(Z<\epsilon)=\frac{1+0,95}{2}=0,975$$

in iz tabel razberemo, da je $\epsilon=1,96$. Torej vrednosti slučajni spremenljivke Z ležijo z natančnostjo 95% v intervalu širine ϵ. Ocenili smo torej pripadnost srednje vrednosti vzorca intervalu z želeno širino

$$P(-\epsilon<Z<\epsilon)=P(\mu-\epsilon\frac{\sigma}{\sqrt{n}}<\bar{X}_n<\mu+\epsilon\frac{\sigma}{\sqrt{n}})$$

Glej.geoGebra.cenilke.html

b) Če variance ne poznamo in je vzorec dovolj velik, dobimo prilagojeno varianco $\hat{S}_n^2$ in jo uporabljamo namesto variance populacije.

c) Če variance ne poznamo in je vzorec majhen, moramo uporabljati bolj razpršeno porazdelitev in sicer se lahko dokaže, da slučajna spremenljivka

$$t=\frac{\bar{X}_n-\mu}{\frac{\hat{S}_n}{\sqrt{n}}}$$

ima **t-student** porazdelitev z **n-1 stopnjami svobode**. Število stopenj svobode se zmanjša za ena, v kolikor, ko pri standardni deviaciji seštevamo razlike odstopanj od srednje vrednosti, vemo , da je vsota razlik enaka nič; to pomeni da imamo eno enačbo, ki povezuje n neznank in seveda je število svobodnih neznank enako n-1. Gostota te porazdelitve je bolj stlačena in se približuje standardizirani normalni porazdelitvi, ko n raste in se prilagojena varianca vzorca približuje varianci populacije. Vrednosti za ploščine pod grafom t-student porazdelitve dobimo v tabelah ali s pomočjo ustrezne programske opreme (npr Excel). Porazdelitev t-student je leta 1908 prvi opisal W.S.Gosset (1876-1937), ki je uporabljal pseudonim student. Za velike n se studentova porazdelitev približuje normalni porazdelitvi.

Intervalna cenilka relativne frekvence

Smo že ugotovili, da za slučajno spremenljivko F_n, ki predstavlja relativno frekvenco določenega pojava na vzorcu z n elementi je pravilna statistika parametra relativne frekvence istega pojava v populaciji in smo izračunali njeno srednjo vrednost in varianco.

$$E[F_n]=p \quad \text{in} \quad V[F_n]=\frac{p(1-p)}{n}$$

F_n ima binomsko porazdelitev $B\left(p,\sqrt{\frac{p(1-p)}{n}}\right)$ zato, če je velikost vzorca n dovolj velika in se p ne preveč oddaljuje od ene polovice, jo lahko aproksimiramo z normalno porazdelitvijo $N\left(p,\sqrt{\frac{p(1-p)}{n}}\right)$. Iz tega sledi, da bo slučajna spremenljivka

standardno normalno porazdeljena N(0,1). Sedaj lahko nadaljujemo kot v prejšnjem poglavju. Če želimo zvedeti za verjetnost, da Z pade v interval $(-\epsilon,\epsilon)$, napišemo

$$P(-\epsilon<Z<\epsilon)=2\mathrm{P}(0<z<\epsilon)=2(P(Z<\epsilon)-\frac{1}{2})=\phi$$

oziroma

$$P(Z<\epsilon)=\frac{1+\phi}{2}$$

kjer je ϕ želena natančnost. Na primer, če je $\phi=0.95$, (95%) potem je

$$P(Z<\epsilon)=\frac{1+0,95}{2}=0,975$$

in iz tabel razberemo, da je $\epsilon=1,96$. Torej vrednosti slučajni spremenljivke *Z* ležijo z verjetnostjo 95% v intervalu širine ϵ. Ocenili smo torej pripadnost srednje vrednosti vzorca intervalu z želeno širino

$$P(-\epsilon<Z<\epsilon)=P(p-\epsilon\sqrt{\frac{p(1-p)}{n}}<F_n<p+\epsilon\sqrt{\frac{p(1-p)}{n}})=\phi$$

Če sedaj iz enačbe

$$F=p+\epsilon\sqrt{\frac{p(1-p)}{n}}$$

dobimo *p* v odvisnosti od *F*, lahko napišemo, da

$$p=F+\epsilon\sqrt{\frac{F(1-F)}{n}}$$

in zato

$$P\left(F-\epsilon\sqrt{\frac{F(1-F)}{n}}<p<F+\epsilon\sqrt{\frac{F(1-F)}{n}}\right)=\phi=0,95 \quad .$$

To je intervalna cenilka za parameter relativna frekvenca določenega dogodka.

Primer: Na neki šoli smo izbrali vzorec *n=250* dijakov in ugotovili, da jih 68% kadi; to pomeni, da $F_{250}=0,68$. Parameter *relativna frekvenca kadilcev šole* leži z verjetnostjo *95%* v intervalu

$$\left(0,68-1,96\sqrt{\frac{0,68(1-0,68)}{n}}<p<0,68+1,96\sqrt{\frac{0,68(1-0,68)}{n}}\right)=\mathbf{(0,62<p<0,74)} \quad .$$

Primer: Imamo kocko, za katero ne vemo, če je poštena ali ne. Kocko vržemo 100 krat in dobimo 13 krat šestico. Oceni interval, v katerem bo z zanesljivostjo 95% ležala relativna frekvenca.

Met kocke predstavlja slučajna spremenljivka X, ki ima seveda le dva možna izida: 1, ko pride šestica, 0, ko ne pride šestica. Iz populacije metov smo slučajno izbrali (smo opravili eksperiment) vzorec z 100 meti kocke z relativno frekvenco ugodnega dogodka $p=\frac{13}{100}$ .Slučajna spremenljivka F_n , relativna frekvenca vzorca z n elementi ima binomsko porazdelitev $B\left(p,\sqrt{\frac{p(1-p)}{n}}\right)$ zato, če je velikost vzorca *n* dovolj velika in se *p* ne preveč oddaljuje od ene polovice, jo lahko aproksimiramo z normalno porazdelitvijo

$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$. Parameter *relativna frekvenca šestic* leži z verjetnostjo *95%* v intervalu

$$\left(0{,}13-1{,}96\sqrt{\frac{0{,}13(1-0{,}13)}{n}}<p<0{,}13+1{,}96\sqrt{\frac{0{,}13(1-0{,}13)}{n}}\right)=(\mathbf{0{,}127<p<0{,}164})$$

Ker je relativna frekvenca izida šestice poštene kocke 0,167, lahko z gotovostjo 95% zaključimo, da naša kocka ni poštena.

Intervalna razlika povprečij dveh populacij

Dani sta populaciji X in Y z ustreznima parametroma srednja vrednost μ_1 in μ_2 ter ustreznima variancama σ_1^2 in σ_2^2, ter z ustreznima Bernoullijevema vzorčenjema z n_1 in n_2 elementi, potem lahko opredelimo vzorčni povprečji $E[\bar{X}_{n_1}]$ in $E[\bar{Y}_{n_2}]$. Vemo že, da

$$E[\bar{X}_{n_1}-\bar{Y}_{n_2}]=\mu_1-\mu_2$$
$$V[\bar{X}_{n_1}-\bar{Y}_{n_2}]=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

Se lahko pokaže, da za dovolj velike vzorce je slučajna spremenljivka razlike dveh vzorčnih povprečij $\bar{X}_{n_1}-\bar{Y}_{n_2}$ porazdeljena normalno $N\left(\mu_1-\mu_{2,}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$. Za računanje ustreznih intervalov postopamo kot zgoraj saj je spremenljivka

$$Z=\frac{(\bar{X}_{n_1}-\bar{Y}_{n_2})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

porazdeljena standardno normalno *N(0,1).*V primeru pa, da ne poznamo variance populacij, le-te ocenimo z vzorčnimi variancami $\hat{S}_{n_1}$ in $\hat{S}_{n_2}$ ter upoštevamo, da je v primeru manjših vzorcev slučajna spremenljivka

$$Z=\frac{(\bar{X}_{n_1}-\bar{Y}_{n_2})-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)\hat{S}_{n_1}^2+(n_2-1)\hat{S}_{n_2}^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

porazdeljena t-student z n_1+n_2-2 svobodnimi stopnjami.

Primer: Želeli bi primerjati uspeh dijakov dveh šol. V prvi šoli smo izbrali vzorec desetih dijakov in dobili povprečje ocen 7,2 s prilagojeno varianco 1,2; v drugi šoli pa smo dobili povprečje ocen 8,0 s prilagojeno varianco 1,8.

Preverjanje hipotez oziroma domnev

Teorijo preizkušanja domnev sta v 20. in 30. letih prejšnjega stoletja razvila J. Neyman in E.S. Pearson. Statistična domneva (ali hipoteza) je vsaka domneva o porazdelitvi slučajne spremenljivke X na populaciji. Ponavadi je domneva o parametru populacije, na podlagi vzorčnih podatkov: na primer domneva o določeni vrednosti srednje vrednosti v primeru,

da poznano vzorčno srednjo vrednost in varianco.

V mnogih primerih ne odločamo na osnovi sigurnega znanja, ampak moramo odločati le na osnovi znanja, ki smo ga pridobili s statistično analizo. Seveda, v takih okoliščinah, ne bomo prepričani, da bo naša odločitev pravilna, lahko pa jo sprejmemo oziroma zavrnemo z dano verjetnostjo. Taka analiza ugotovi odstopanja od *norme* in morebitno prisotnost nepričakovanih *sistematičnih napak.*

Formalen postopek za testiranje hipotez

1. Postavi hipotezi:

– ničelna H_0

– alternativna H_1

2. Za parameter poiščemo kar se da dobro cenilko .

3. Določi odločitveno pravilo.

Izberemo stopnjo značilnosti α. Na osnovi stopnje značilnosti in porazdelitve statistike določimo kritično območje.

4. Zberi/manipuliraj podatke ter na vzorčnih podatkih izračunaj (eksperimentalno) vrednost testne statistike.

5. Primerjaj in naredi zaključek.

• *če eksperimentalna vrednost pade v kritično območje, ničelno domnevo zavrni in sprejmi osnovno domnevo ob stopnji značilnosti .*

• *če eksperimentalna vrednost ne pade v kritično območje, pa pravimo da vzorčni podatki kažejo na statistično neznačilne razlike med parametrom in vzorčno oceno.*

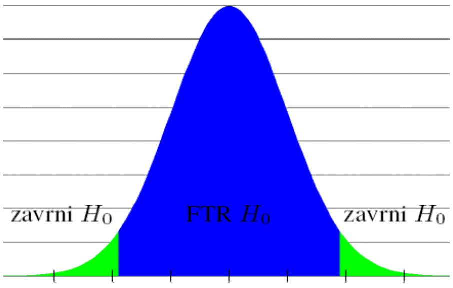
Primer: Proizvajalec prenosnih telefonov proizvaja telefone, ki brezhibno delujejo t=48 mesecev s standardnim odklonom $\sigma=5\,\text{mesecev}$. Proizvodnjo ustavimo bodisi v primeru da se telefoni predčasno pokvarijo, saj bo kupec nejevoljen, bodisi v primeru da je čas brezhibnega delovanja predolg, saj bi se tako zmanjšal nakup novih telefonov. V tovarni izberejo slučajen vzorec *n=17* novih telefonov in dobijo, da je vzorčna srednja vrednost trajanja telefonov $\bar{X}=45\,\text{mesecev}$. Če predpostavljamo, da je trajanje telefonov porazdeljeno normalno povej, če bodo proizvodnjo telefonov ustavili s stopnjo značilnosti $\alpha=5\%$?

Najprej postavimo hipotezi:

$H_0 : \mu = 48$;

$H_1 : \mu \neq 48$.

Predpostavljamo, da je H_0 resnična. Že vemo, da je slučajna spremenljivka $Z = \frac{\bar{X} - \mu}{\frac{\bar{S}}{\sqrt{n}}}$ porazdeljena normalno.



Izračunajmo vrednost $t = \frac{45 - 48}{\frac{5}{\sqrt{17}}} = -2{,}47$ in jo primerjamo s testno statistiko Z Ker je stopnja značilnosti $\alpha = 5\% = 0.05$, sledi da

$$P(-\epsilon < Z < \epsilon / H_0) = 1 - 0.05 = 0.95$$

Velja tudi

$$P(-\epsilon < Z < \epsilon) = 1 - 2P(Z < -\epsilon) = 1 - 2(1 - P(Z < \epsilon)) = -1 + 2\mathrm{P}(Z < \epsilon) = 0.95$$

zato iz tabele za normalno porazdelitev razberemo $\epsilon = 1.96$ za verjetnost

$$P(T < \epsilon) = \frac{1 + 0.95}{2} = 0.975$$

Razvidno je, da izračunana vrednost pade v kritično območje, zato hipotezo H_0 zavrnemo in sprejmemo alternativno hipotezo: proizvodnjo je treba ustaviti.

Primer: Tovarna proizvaja fotografski film. Vemo, da je srednja vrednost populacije $\mu = 5\text{mm}$ in standardna deviacija $\sigma = 1{,}4\ \text{mm}$. Občasno želimo preveriti kvaliteto izdelkov, zato slučajno izberemo Bernoullijev vzorec *n=100* filmov in dobimo srednjo vrednost $\bar{X}_{100} = 5{,}6\ \text{mm}$. Se sprašujemo, če je s stopnjo značilnosti *5%* kvaliteta filmov dobra.

H_0 : opazovani vzorec je del populacije

H_1 : opazovani vzorec ni del populacije

Preveriti moramo, če je z verjetnostjo *95%* opazovani vzorec del populacije. Vemo, da je slučajna spremenljivka aritmetična srednja vrednost vzorcev porazdeljena normalno s srednjo vrednostjo μ in varianco $\frac{\sigma^2}{n}$. Standardizirana slučajna spremenljivka

$$Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

je porazdeljena *N(0,1)*. Smo že videli, da

$$P(-\epsilon<Z<\epsilon)=P(\mu-\epsilon\frac{\sigma}{\sqrt{n}}<\bar{X}_n<\mu+\epsilon\frac{\sigma}{\sqrt{n}})=\phi=0{,}95=1-\alpha$$

in v našem primeru je zaželeni interval

$$\left(5-1{,}96\frac{1{,}4}{\sqrt{100}}<\bar{X}_{100}<5+1{,}96\frac{1{,}4}{\sqrt{100}}\right)=(4.73<\bar{X}_{100}<5.27)\,.$$

Ker opazovani vzorec pade v kritično območje. Bomo hipotezo H_0 zavrnili. Lahko zaključimo, da je z verjetnostjo *95%* vzorec slab: napake v vzorcu niso le vsota slučajnih napak, a prisotna je tudi kaka sistematična napaka; morda se je kak stroj pokvaril ali kak delavec ne opravlja pravilno svojega dela.

Če variance populacije ne poznamo, jo moramo aproksimirati z slučajno spremenljivko $\hat{S}_n^2$ in vemo, da slučajna spremenljivka

$$t=\frac{\bar{X}_n-\mu}{\frac{\hat{S}_n}{\sqrt{n}}}$$

ima porazdelitev t-student z (n-1) stopnjami svobode.

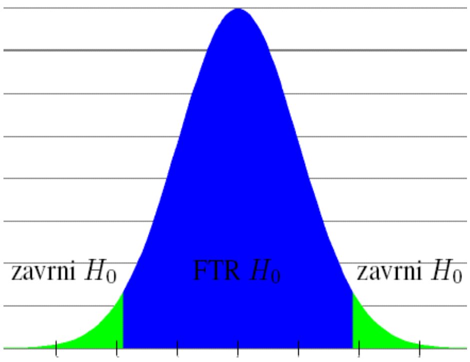
Primer: Želimo preučiti srednjo vrednost starosti ljudi s srčnimi težavami. Slučajno izberemo vzorec *n=17* ljudi in dobimo, da je vzorčna srednja vrednost $\bar{X}=48.59$ in prilagojena vzorčna varianca $\bar{S}=245.55$. Lahko trdimo, da je srednja vrednost starosti obolelih $\mu=48$ s stopnjo značilnosti $\alpha=5\,\%$?

Najprej postavimo hipotezi:

$H_0\colon \mu=48$;

$H_1\colon \mu\neq 48$.

Predpostavljamo, da je H_0 resnična. Že vemo, da je slučajna spremenljivka $T=\frac{\bar{X}-\mu}{\frac{\bar{S}}{\sqrt{n}}}$ porazdeljena kot t-student s 16 stopnjami prostosti.



Izračunajmo vrednost $t=\frac{48.59-48}{\frac{\sqrt{245.55}}{\sqrt{17}}}=0.155$ in jo primerjamo s testno statistiko T. Ker je stopnja značilnosti $\alpha=5\,\%=0.05$, sledi da

$$P(-\epsilon<T<\epsilon/H_0)=1-0.05=0.95$$

Velja tudi

$$P(-\epsilon<T<\epsilon)=1-2P(T<-\epsilon)=1-2(1-P(T<\epsilon))=-1+2\mathrm{P}(T<\epsilon)=0.95$$

zato iz tabele za t-student porazdelitev razberemo $\epsilon=2.12$ za verjetnost

$$P(T<\epsilon)=\frac{1+0.95}{2}=0.975$$

Razvidno je, da izračunana vrednost ne pade v kritično območje, zato hipoteze H_0 ne moremo zavrniti.

Test za primerjanje porazdelitvenih zakonov

Želimo sedaj preveriti hipotezo, da neka populacija ima določeno porazdelitev. Iz določenega na slepo izbranega vzorca bomo izračunali *empirične relativne frekvence* $\bar{p}_i$, iz postavljene hipoteze bomo dobili *teoretične relativne frekvence* p_i. Nato opredelimo sledečo slučajno spremenljivko

$$T_n=n\sum_{i=1}^{m}\frac{(\bar{p}_i-p_i)^2}{p_i}$$

Se lahko pokaže (Pearsonov izrek), da se porazdelitev T_n približuje porazdelitvi spremenljivke $\chi^2_\alpha(m-1)$, kjer je α zaželena stopnja gotovosti. Postavljeno hipotezo sprejmemo, ko je $T_n>\chi^2_\alpha(m-1)$.

Primer: Hočemo oceniti z gotovostjo 95% ali je nek kovanec pošten ali ni. Kovanec vržemo n=2000 krat in zabeležimo izide v tabelo

1	2	3	4	5	6
388	322	314	316	344	316

Opravili bomo Pearsonov test s hipotezo, da je kocka poštena.

H_0: kocka je poštena
H_1: kocka ni poštena

V tabelo izpišemo teoretične (verjetnostni račun) in eksperimentalne relativne frekvence.

	1	2	3	4	5	6
p_i	0,167	0,167	0,167	0,167	0,167	0,167
$\bar{p}_i$	0,194	0,161	0,157	0,158	0,172	0,158

Izračunajmo nato vrednost Pearsonove spremenljivke

$$T_n=2000\sum_{i=1}^{6}\frac{(\bar{p}_i-0{,}167)^2}{0{,}167}=12{,}6$$

Iz tabel ali s programsko opremo ugotovimo,da je $\chi^2_{0{,}95}(5)=11{,}07$. Ker izračunana vrednost pade v kritično območje, hipotezo H_0 zavrnemo in sprejmemo alternativno hipotezo: kocka ni poštena s stopnjo značilnosti 5%. Hipoteze ne bi zavrnili s stopnjo značilnosti 10% saj je $0\chi^2_{0,90}(5)=9.27$. Torej lahko rečemo, da je kocka poštena s stopnjo značilnosti 10% (če smo dovolj tolerantni je lahko vsaka kocka poštena).

Test za primerjanje razlike povprečij

Smo že videli, da za dovolj velike neodvisne vzorce (večje od 30) je slučajna spremenljivka

$$Z=\frac{(\bar{X}_{n_1}-\bar{Y}_{n_2})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

porazdeljena standardno normalno *N(0,1)*.V primeru pa, da ne poznamo variance populacij, le-te ocenimo z vzorčnimi variancami $\hat{S}_{n_1}$ in $\hat{S}_{n_2}$ ter upoštevamo, da je v primeru manjših vzorcev slučajna spremenljivka

$$Z=\frac{(\bar{X}_{n_1}-\bar{Y}_{n_2})-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)\hat{S}^2_{n_1}+(n_2-1)\hat{S}^2_{n_2}}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}¿$$

porazdeljena t-student z n_1+n_2-2 svobodnimi stopnjami.

Primer: Preveriti hočemo domnevo, da so dekleta na maturi boljša od fantov. To domnevo preverimo tako, da izberemo slučajni vzorec 36 deklet in slučajni vzorec 36 fantov, za katere imamo izpitne rezultate: $\bar{X}_F=7$; $\bar{X}_D=7{,}2$; $\hat{S}^2_D=1$; $\hat{S}^2_F=1$. Domnevo želimo preveriti pri *95%* stopnji zanesljivosti.

Vemo, da za večje vzorce je slučajna spremenljivka

$$Z=\frac{(\bar{X}_{n_1}-\bar{Y}_{n_2})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

porazdeljena standardno normalno.

Postavimo ničelno hipotezo H_0 : fantje in dekleta so enako dobri, oziroma $m_F-\mu_D=0$.

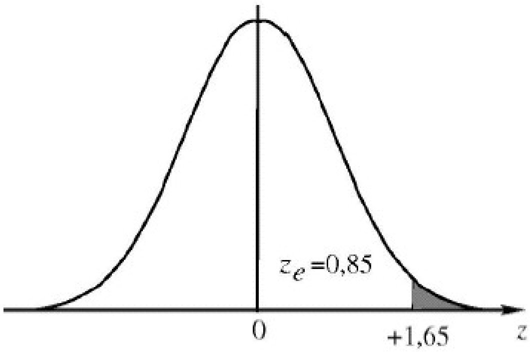
Postavimo alternativno hipotezo H_1 : dekleta so boljša od fantov, oziroma $H=m_F-\mu_D<0$.

Če predpostavljamo, da je hipoteza H_0 resnična, sledi da

$$Z=\frac{(7-7{,}2)-(0)}{\sqrt{\frac{1}{36}+\frac{1}{36}}}=0{,}848$$

Normalna porazdelitev nam da vrednost 1,65 za verjetnost 0.95, saj je

$$P(T<\epsilon)=0.95$$



zato smo izven kritičnega območja in ničelne hipoteze ne moremo zavrniti: *povprečni uspeh deklet in fantov ni statistično značilno različna.*

Če bi Z padel v kritično območje, bi ničelno hipotezo zavrnili in sprejeli alternativno hipotezo.

Če primer ponovimo z novimi podatki: $\bar{X}_F=7$ in $\bar{X}_D=7,5$, potem Z=2,12 in leži v kritičnem območju; v tem primeru ničelno hipotezo o enakosti povprečij zavrnemo in z *95%* stopnjo zanesljivosti sprejmemo alternativno hipotezo, da so dekleta boljša od fantov.

ANALITIČNO KAZALO

Bernoullijevo vzorčenje....8
Blok vzorčenje....8
Cenilke....8
dohodek....7
empirične relativne frekvence....19
hipoteza....15
inferenčna....4
Intervalne cenilke....12
Parameter....5
Populacija....5
Povprečje....5
Prilagojena vzorčna varianca....6
Shannon....6
Standardizirana....12
Statistična enota....5
Statistična domneva....15
Statistika....3
Statistika....5
stohastičnega....6
t-student....13
točkaste cenilke....11
Varianca....6
Verjetnost....4
Vzorec....5
W.S.Gosset....13