4.3 Osnovni pojmi pri statističnem preverjanju domnev

Želimo preveriti ali je kovanec pošten. Naredili smo poizkus, kjer smo 10-krat vrgli kovanec in dobili, da je grb padel 7-krat.

- Zapišite ničelno domnevo za vaš primer. Ali je ničelna domneva enostavna ali sestavljena? Zapišite testno statistiko, označite jo z X - kakšna je njena porazdelitev pod ničelno domnevo?
  $H_0 : p = 0{,}5$. S tem je porazdelitev slučajne spremenljivke pod ničelno domnevo natanko določena, zato pravimo, da je ničelna domneva enostavna. Testna statistika X je število grbov: porazdelitev testne statistike pod ničelno domnevo je binomska $B(0{,}5, 10)$.

- Denimo, da je vaša alternativna domneva $H_A : p > 0{,}5$. Ali je ta domneva enostavna ali sestavljena? Pri kakšnih vrednostih X boste zavrnili ničelno domnevo v prid alternativni? Ali je domneva enostranska ali dvostranska?
  Domneva je sestavljena, saj zajema več parametrov iste porazdelitve. Zavračali bomo pri velikih vrednostih X. Domneva je enostranska, saj nas bo zanimal le desni rep porazdelitve.

- V našem primeru je $X = 7$. Kolikšna je verjetnost, da se na vzorcu zgodi ta dogodek, če ničelna domneva drži?
  Če ničelna domneva drži ($p = 0{,}5$), je $P(X = 7) = 0{,}117$.

- Denimo, da je območje zavrnitve sestavljeno iz vrednosti $\{10\}$. Kakšna je stopnja značilnosti α v tem primeru? Kakšna je stopnja značilnosti, če je območje zavrnitve sestavljeno iz vrednosti $\{6,7,8,9,10\}$?
  Če ničelna domneva drži ($p = 0{,}5$), je $P(X = 10) = 0{,}001$.
  Če ničelna domneva drži ($p = 0{,}5$), je $P(X \geq 6) = 0{,}377$.

- Določite območje zavrnitve pri stopnji značilnosti $\alpha = 0{,}05$. Ali lahko na podlagi dobljenih podatkov zavrnete ničelno domnevo pri stopnji značilnosti $\alpha = 0{,}05$?
  Najmanjša vrednost k, tako da velja $P(X \geq k) \leq 0{,}05$, je enaka 9. Stopnja značilnosti je v tem primeru enaka 0,01. Ničelne domneve seveda ne moremo zavrniti.

- Kolikšna je moč testa pri tej vrednosti α, če predpostavimo, da je prava vrednost parametra $p = 0{,}6$? Kaj pa pri $p = 0{,}7$? Kolikšna je v teh

primerih napaka druge vrste?
Moč testa je majhna - 0,046 oziroma 0,149. Pri tako majhnem vzorcu in majhni stopnji značilnosti bomo ničelno domnevo le redko uspešno zavrnili. Napako druge vrste izračunamo kot 1-moč.

- Predpostavite sedaj, da je vaša alternativna domneva $H_A : p \neq 0{,}5$. Ali je ta domneva enostavna ali sestavljena? Ali je domneva enostranska ali dvostranska?
  Alternativna domneva je še vedno sestavljena, zdaj je tudi dvostranska.

- Kakšno bo sedaj območje zavrnitve, če želite, da je $\alpha \leq 0{,}05$? Kakšna natanko bo stopnja značilnosti za to območje?
  Območje zavrnitve bo sestavljeno iz vrednosti $\{0{,}1{,}9{,}10\}$. Stopnja značilnosti za to območje je enaka $\alpha = 0{,}02$.

- Izračunajte še moč testa v tem primeru.
  Pri moči testa moramo upoštevati, da bomo sedaj ničelno domnevo zavrnili tudi če bo X enak 0 ali 1. Ker pa je verjetnost teh dveh vrednosti za $p = 0{,}6$ oz. $p = 0{,}7$ zelo majhna, se moč praktično ne spremeni.

p\k	0	1	2	3	4	5	6	7	8	9	10
0,5	0,001	0,011	0,055	0,172	0,377	0,623	0,828	0,945	0,989	0,999	1
0,6	0,000	0,002	0,012	0,055	0,166	0,367	0,618	0,833	0,954	0,994	1
0,7	0,000	0,000	0,002	0,011	0,047	0,150	0,350	0,617	0,851	0,972	1

Tabela 1: Kumulativne verjetnosti za binomsko porazdelitev pri $n = 10$ ($P(X \leq k|p)$)

Predlogi za vaje v R-u:

- 1000x ponovite poskus v katerem po 10x mečete kovanec. Oglejte verjetnost zavrnitve za posamezno območje.

- Spremenite verjetnost s katero pade grb in si oglejte moč testa.

- Povečajte vzorec in si oglejte, kako se spreminja moč testa.

4.4 Moč testa

Iz literature lahko povzamemo, da se športnikovo povprečje hemoglobina ob vsaj 14-dnevnem bivanju na višini nad 1500m zviša za 2 g/l, medtem ko višinski treningi ne vplivajo na varianco njegovih vrednosti. Ob običajnih treningih se posameznikove vrednosti porazdeljujejo normalno, $X \sim N(\mu_1, 5^2)$, kjer je μ_1 športnikovo povprečje.
Športnik pogosto opravlja višinske treninge, vendar v krajših intervalih. Zanima ga, ali se njegovo povprečje hemoglobina v obdobju višinskih treningov kljub temu zviša. V sezoni opravi 12 meritev, 8 med obdobjem višinskih priprav in 4 sicer. Cilj naloge je ugotoviti, kakšna bo moč njegovega testa, če bo pri sklepanju uporabil stopnjo značilnosti $\alpha = 0{,}05$?

- Kaj je športnikova ničelna in kaj alternativna domneva?

  Zapišimo porazdelitev hemoglobina v obdobju višinskih priprav z $N(\mu_2, 5^2)$.
  Ničelna domneva je:
  H_0: Povprečje hemoglobina v obeh obdobjih je enako, $\mu_1 = \mu_2$.
  Alternativna domneva je, da je $\mu_2 > \mu_1$, zanima ga torej le enostranski test.

- Predlagajte testno statistiko. Izračunajte njeno porazdelitev pod ničelno domnevo.

  Športnik bo izračunal razliko povprečij na vzorcih, ki je porazdeljena kot

  $$R = \bar{X}_2 - \bar{X}_1 \sim N\left(\mu_2 - \mu_1, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$$

  Testna statistika

  $$Z = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

  je torej pod ničelno domnevo porazdeljena standardno normalno. Ker ga zanima le enostranska alternativna domneva, bo ničelno domnevo zavrnil, kadar bo $Z > z_\alpha$, torej $Z > 1{,}64$.

- Izračunajte moč testa, torej verjetnost, da bo ničelno domnevo uspel zavrniti, če se mu povprečje hemoglobina v obdobju višinskih priprav zares poveča za 2 g/l?

Zanima nas $P(Z > 1{,}64)$, torej $P(R > 1{,}64 \cdot \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$, torej v našem primeru $P(R > 5{,}02)$. Pod alternativno domnevo je $R \sim N\left(2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$ in zato

$$P\left(R > 1{,}64 \cdot \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) \quad = \quad P\left(\frac{\bar{X}_2 - \bar{X}_1 - 2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} > 1{,}64 - \frac{2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}\right)$$

$$P\left(U > 1{,}64 - \frac{2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right),$$

kjer je U standardna normalna spremenljivka. V našem primeru:

$$P\left(U > 1{,}64 - \frac{2}{5\sqrt{\frac{1}{8} + \frac{1}{4}}}\right) = P(U > 0\ 99) = 0{,}16 \tag{1}$$

Moč testa je zelo majhna - pri tako majhnem številu meritev je le majhna verjetnost, da bo športnik zavrnil ničelno domnevo (četudi se mu povprečje dejansko zares zviša za 2 g/l).

- Kako bi se moč testa spremenila, če bi imel na voljo enako število meritev v vsakem obdobju?

  Če bi imel po 6 meritev v vsakem obdobju, bi bila moč enaka

$$P\left(U > 1{,}64 - \frac{2}{5\sqrt{\frac{1}{6} + \frac{1}{6}}}\right) = P(U > 0{,}95) = 0{,}17$$

- Kako je moč testa odvisna od variance posameznikovih meritev in kako od dejanske velikosti razlike v populaciji?

  Iz enačbe (1) je očitno, da večja razlika pomeni večjo moč - če je dejanska razlika med obdobjema večja, jo bomo lažje opazili na podatkih. Če bi bila varianca posameznikovih meritev manjša, bi imeli manjšo standardno napako in zato večjo moč testa.

4.5 Posplošeni test razmerja verjetij

Zanima nas ali imajo zares vsi športniki enako variabilnost hemoglobina. Primerjati želimo meritve k športnikov, naj bodo vrednosti i-tega športnika ($i = 1, \ldots, k$) porazdeljene normalno, torej $X_{ij} \sim N(\mu_i, \sigma_i^2)$, kjer $j = 1, \ldots, n_i$ označujejo meritve pri posamezniku. Predpostavimo, da so vse meritve med seboj neodvisne.

- Zapišite ničelno in alternativno domnevo

  Ničelna domneva:
  $$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$

  Alternativna domneva:
  $$H_1 : \sigma_i^2 \text{ niso vse enake}$$

- Najprej vzemimo, da imamo le enega športnika in n njegovih meritev. Kako bi ocenili njegova parametra μ in σ^2 z metodo največjega verjetja? Funkcija verjetja je enaka
  $$L(x, \mu, \sigma) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(x_j - \mu)^2}{2\sigma^2},$$
  del njenega logaritma v katerem nastopata parametra, ki ju želimo oceniti pa je enak
  $$\log L(x, \mu, \sigma) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{j=1}^{n} (x_j - \mu)^2.$$
  Poiščimo maksimum po μ:
  $$\begin{aligned} \frac{\partial \log L(x, \mu, \sigma)}{\partial \mu} &= 0 \\ -\frac{1}{2\widehat{\sigma}^2} \sum_{j=1}^{n} (x_j - \widehat{\mu})(-2) &= 0 \\ \sum_{j=1}^{n} (x_j - \widehat{\mu}) &= 0 \\ \widehat{\mu} &= \frac{1}{n} \sum_{j=1}^{n} x_j \end{aligned}$$

Pa še za varianco:

$$\begin{aligned}
\frac{\partial \log L(x,\mu,\sigma)}{\partial \sigma} &= 0 \\
-\frac{n}{\widehat{\sigma}} - \frac{1}{2}\sum_{j=1}^{n}(x_j-\widehat{\mu})^2\frac{-2}{\widehat{\sigma}^3} &= 0 \\
-\widehat{\sigma}^2 n + \sum_{j=1}^{n}(x_j-\widehat{\mu})^2 &= 0 \\
\widehat{\sigma}^2 &= \frac{1}{n}\sum_{j=1}^{n}(x_j-\widehat{\mu})^2
\end{aligned}$$

- Vrnimo se h k športnikom. Utemeljite, da so pod alternativno domnevo ocene parametrov enake

  $$\begin{aligned}
  \hat{\mu}_i &= \frac{1}{n_i}\sum x_{ij} \\
  \hat{\sigma}_i^2 &= \frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij}-\hat{\mu}_i)^2
  \end{aligned}$$

  Funkcija verjetja pod alternativno domnevo je enaka

  $$L(x,\mu,\sigma) = \prod_{i=1}^{k}\prod_{j=1}^{n_i}\frac{1}{\sqrt{2\pi}\sigma_i}\exp -\frac{(x_{ij}-\mu_i)^2}{2\sigma_i^2}$$

  Vsak člen vsote, ki jo dobimo po logaritmiranju gornje funkcije, je sestavljen le iz parametrov enega posameznika, ko odvajamo po tistem parametru torej ostanejo le členi, ki so vezani na tistega posameznika. Za ocenjevanje parametrov za nekega posameznika i torej potrebujemo izključno njegove vrednosti, parametre posameznikov torej ocenimo povsem neodvisno drug od drugega.

- Kakšna je ocena povprečij pod ničelno domnevo?
  Pod ničelno domnevo je σ_i enak za vse i, zato ga v logaritmu funkcije verjetja lahko izpostavimo in ne vpliva na našo oceno posameznih povprečij. Ocena posameznih povprečij je zato enaka kot pod alternativno domnevo.

- Kakšna je ocena variance pod ničelno domnevo?
  Del logaritma funkcije verjetja, ki nas zanima, je enak

$$\log L(x, \mu, \sigma) = -\sum_{i=1}^{k} n_i \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2.$$

  Odvod po σ izenačimo z 0 in dobimo

$$\hat{\sigma}_0^2 = \frac{1}{\sum\limits_{i=1}^{k} n_i} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2$$

- Kako bi ničelno domnevo preverili s testom razmerja verjetij?

  Zapišemo Wilksov Λ (zgoraj je funkcija verjetja pod alternativno domnevo, spodaj pod ničelno):

$$\begin{aligned}
\Lambda &= \frac{\prod\limits_{i=1}^{k} \prod\limits_{j=1}^{n_i} \left(\frac{1}{\sqrt{2\pi}\widehat{\sigma}_i} \exp\{-\frac{(x_{ij}-\widehat{\mu}_i)^2}{2\widehat{\sigma}_i^2}\} \right)}{\prod\limits_{i=1}^{k} \prod\limits_{j=1}^{n_i} \left(\frac{1}{\sqrt{2\pi}\widehat{\sigma}_0} \exp\{-\frac{(x_{ij}-\widehat{\mu}_{0i})^2}{2\widehat{\sigma}_0^2}\} \right)} \\
&= \frac{\left(\prod\limits_{i=1}^{k} \prod\limits_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\widehat{\sigma}_i} \right) \prod\limits_{i=1}^{k} \exp\{-\frac{\sum_{j=1}^{n_i} (x_{ij}-\widehat{\mu}_i)^2}{2\widehat{\sigma}_i^2}\}}{\left(\prod\limits_{i=1}^{k} \prod\limits_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\widehat{\sigma}_0} \right) \exp\{-\sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{(x_{ij}-\widehat{\mu}_{0i})^2}{2\widehat{\sigma}_0^2}\}}
\end{aligned}$$

  Vstavimo ocene za variance v eksponent in tako v števcu kot tudi v imenovalcu dobimo $\exp\{-\frac{1}{2} \sum_{i=1}^{k} n_i\}$, ki se zato pokrajša. Logaritem Λ

je enak

$$
\begin{aligned}
\log\Lambda &= -\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log(\widehat{\sigma}_i)\right) + \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log(\widehat{\sigma}_0)\right) \\
&= \left(\sum_{i=1}^{k} n_i\log(\widehat{\sigma}_0)\right) - \left(\sum_{i=1}^{k} n_i\log(\widehat{\sigma}_i)\right) \\
&= \sum_{i=1}^{k} n_i\left[\log(\widehat{\sigma}_0) - \log(\widehat{\sigma}_i)\right]
\end{aligned}
$$

Dvakratna vrednost logaritma verjetij je porazdeljena kot χ^2_{k-1}, saj smo pod alternativno domnevo ocenili $k-1$ parametrov več kot pod ničelno.