

1.6 Porazdelitev povprečja

Vrnimo se spet k primeru odkrivanja dopinga. Izkaže se, da ima vsak posameznik sebi lastno povprečje hemoglobina in da se te vrednosti med posamezniki precej razlikujejo. Da bi dosegli večjo občutljivost testa, zato uvedemo polletno testno obdobje, v katerem vsakega športnika testiramo petkrat. Povprečje teh petih meritev bomo vzeli kot oceno za posameznikovo povprečje pri testih v prihodnosti (meje bomo postavljali glede na to povprečje). Recimo, da vemo, da se vrednosti vsakega športnika okrog njemu lastnega povprečja porazdeljujejo normalno z varianco $\sigma^2 = 5^2$, in da so posamezne vrednosti med seboj neodvisne.

- Naj bodo X_i , $i = 1, \dots, n$, neodvisne, enako porazdeljene slučajne spremenljivke. Kaj lahko rečete o pričakovani vrednosti in varianci njihovega povprečja? Označite $E(X_i) = \mu$ in $\text{var}(X_i) = \sigma^2$ za vsak i .
- Izračunajte meje okrog ocenjenega povprečja, znotraj katerih naj bi pri šesti meritvi nedopingiran športnik ostal z verjetnostjo 0,99 (prvih pet meritev uporabimo za oceno športnikovega povprečja).
Namig: uporabite rezultat, da je vsota neodvisnih normalno porazdeljenih spremenljivk spet normalna.
- Ali je povprečje neodvisnih enako porazdeljenih slučajnih spremenljivk vedno znova porazdeljeno z isto porazdelitvijo?

Predlogi za vaje v R-u:

- Predpostavite, da so povprečja športnikov normalno porazdeljena z $N(148, 7,5^2)$ in zgenerirajte povprečne vrednosti za 100 športnikov. Nato uporabite normalno porazdelitev $N(0, 5^2)$, ki predstavlja odmike od osebnega povprečja vsakega športnika - zgenerirajte po 6 vrednosti na posameznika. Ocenite osebna povprečja s pomočjo prvih petih vrednosti ter primerjajte varianco teh povprečij s teoretično vrednostjo. Oglejte si porazdelitev odstopanja šeste vrednosti od prvih petih.

1.7 Pogojna pričakovana vrednost in varianca

Raziskovalci na področju športa so dokazali, da je pri kolesarjih hemoglobin izven tekmovalnega obdobja porazdeljen kot $N(150, 7^2)$, med tekmovalnim obdobjem pa kot $N(140, 11^2)$. Vzemimo, da tekmovalno obdobje traja 9 mesecev. Zanimata nas pričakovana vrednost in standardni odklon za naključno odvzeti vzorec.

Namig: Zanima nas slučajna spremenljivka Y , vemo $\{Y|X = 0\} \sim N(150, 7^2)$ in $\{Y|X = 1\} \sim N(140, 11^2)$, $P(X = 1) = 0,75$

- Skicirajte porazdelitev Y , kaj lahko rečete o pričakovani vrednosti ter standardnem odklonu?
- Na danem primeru razložite formulo $E(Y) = E[E(Y|X)]$. Je $E(Y|X)$ slučajna spremenljivka ali konstanta? Izračunajte pričakovano vrednost spremenljivke Y .
- Izračunajte varianco Y
- Izrazite varianco v splošnem ($\{Y|X = 0\} \sim N(\mu_0, \sigma_0^2)$, $\{Y|X = 1\} \sim N(\mu_1, \sigma_1^2)$, $P(X = 1) = p$)
- Izračunajte kovarianco X in Y . Izrazite je splošno ($\{Y|X = 0\} \sim N(\mu_0, \sigma_0^2)$, $\{Y|X = 1\} \sim N(\mu_1, \sigma_1^2)$, $P(X = 1) = p$). Kako je kovarianca odvisna od parametrov? Kaj pa korelacija?

Predlogi za vaje v R-u:

- Zgenerirajte veliko število vrednosti in si oglejte njihovo porazdelitev:

```
> set.seed(1)
> a <- rnorm(1000000, mean=140, sd=11) #generiram vrednosti Y|X za tek. obd.
> b <- rnorm(1000000, mean=150, sd=7)  #vrednosti Y|X za ne-tek. obd.
> x <- sample(0:1, size=1000000,      #obdobje - porazdelitev X
+ replace=T, prob=c(0.25, 0.75))
> y <- a*x+b*(1-x)                    #slučajna spremenljivka Y
> hist(y, prob=T)                     #narisemo spremenljivko
> mean(y)                             #ocena povprecja
> var(y)                              #ocena variance
```

- Poizkusite preveriti vsakega od rezultatov še z R-om. Primerjajte teoretične vrednosti z njihovimi ocenami.