

VIS - Labosi (drugi ciklus)

Pouzdaní intervali i testiranje hipoteza u R-u

Vanessa Keranović, Kristijan Kilassa Kvaternik, Mate Puljiz, Stjepan Šebek, Josip Žubrinić

Uvod

Labosi na predmetu “Vjerojatnost i statistika” izvode se u programskom jeziku R, radnoj okolini RStudio, u obliku R Markdown izvještaja koji kombiniraju pisanje teksta s programskim kodom i rezultatima izvođenja koda. Predznanje ovih alata nije nužno za izvedbu jer se kroz labose demonstriraju ključne funkcionalnosti. Kao dodatne materijale preporučamo udžbenik “Programirajmo u R-u” doc. dr. sc. Damira Pintara, dostupan na stranicama vještine “Osnove programskog jezika R” (<https://www.fer.unizg.hr/predmet/opjr>).

RStudio u oblaku

RStudio moguće je isprobati i bez lokalne instalacije na vašem računalu. Dovoljno je otići na stranicu <https://rstudio.cloud>, registrirati se i započeti s programiranjem.

R Markdown

R Markdown dokument sastavljen je od isječaka koda u R-u i teksta oko njih. Trenutnu liniju koda izvodimo kombinacijom tipaka **CTRL+ENTER**, a cijeli isječak kombinacijom **CTRL+SHIFT+ENTER**. Iz R Markdown dokumenta moguće je stvoriti izvještaj u PDF, HTML, DOCX ili drugim formatima (output parametar u zaglavlju dokumenta) kombinacijom tipaka **CTRL+SHIFT+K**.

Auditorne vježbe (12. tjedan)

U nastavku ćemo pokazati kako se neki zadaci rješavaju u R-u. Uz kôd, pronaći ćete i klasična rješenja istih zadataka. Napominjemo da ćete **za ispit morati znati rješavati zadatke klasičnom metodom (papir, olovka, tablice, kalkulator)**. Dio završnog ispita će se također sastojati od kratkih pitalica gdje ćemo ispitati vašu sposobnost da iščitajte output kraćeg R programa i na temelju toga donesete valjane statističke zaključke.

Zadatak 1.

Registrirana su vremena (u minutama) između uzastopnih poziva u telefonskoj centrali: 8, 12, 7, 10, 5. Ako znamo da se vrijeme između uzastopnih poziva ravna po eksponencijalnoj razdiobi s parametrom λ , pomoću kriterija najveće izglednosti odredite procjenu za parametar λ . Kolika je vjerojatnost da će se na sljedeći poziv čekati više od 5 minuta?

Rješenje

$$\begin{aligned}f(\lambda, x) &= \lambda e^{-\lambda x}, \quad x > 0, \\L(\lambda, x_1, \dots, x_n) &= f(\lambda, x_1) \cdots f(\lambda, x_n) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \\ \frac{\partial L}{\partial \lambda} &= n\lambda^{n-1} e^{-\lambda \sum_{i=1}^n x_i} - \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \sum_{i=1}^n x_i = \lambda^{n-1} e^{-\lambda \sum_{i=1}^n x_i} (n - \lambda \sum_{i=1}^n x_i) \\ \frac{\partial L}{\partial \lambda} &= 0, \text{ za } \lambda = 0 \text{ ili } \lambda = \frac{n}{\sum_{i=1}^n x_i} \\ \hat{\lambda} &= \frac{1}{\bar{x}} = \frac{1}{\frac{8+12+7+10+5}{5}} = 0.119 \\ X &\sim \mathcal{E}(0.119), \quad \mathbb{P}(X > 5) = e^{-5\lambda} = 0.55\end{aligned}$$

Gornji zadatak možemo riješiti i u R-u. Trebat će nam funkcija `fitdistr` iz paketa `MASS` koja procjenjuje parametre distribucije metodom najveće izglednosti (MLE od *engl.* maximum likelihood estimation).

```
sample = c(8, 12, 7, 10, 5)
fit = MASS::fitdistr(sample, 'exponential')
```

Estimirani λ koristimo za računanje tražene vjerojatnosti koristeći funkciju `pexp`.

```
lambda_est = fit$estimate
1 - pexp(5, rate = lambda_est)
```

```
## [1] 0.5514313
```

Sve standardne distribucije su implementirane u R-u. Primjerice `dexp` daje funkciju gustoće, `pexp` funkciju razdiobe (kumulativnu funkciju distribucije), `qexp` daje kvantile, a `rexp` generira slučajan uzorak iz eksponencijalne distribucije.

Analogno imamo funkcije `dnorm`, `pnorm`, `qnorm` i `rnorm` za normalnu distribuciju; `dpois`, `ppois`, `qpois` i `rpois` za Poissonovu; `dbinom`, `pbinom`, `qbinom` i `rbinom` za binomnu distribuciju i slično.

Zadatak 2.

Uzorak x_1, x_2, \dots, x_n izvučen je iz populacije koja se ravna po geometrijskoj razdiobi. Pomoću kriterija najveće izglednosti odredite procjenu za parametar p .

Rješenje

$$\begin{aligned} f(p, x) &= p(1-p)^{x-1}, \quad x = 1, 2, 3, \dots \\ L(p, x_1, \dots, x_n) &= f(p, x_1) \cdots f(p, x_n) = p^n (1-p)^{\sum_{i=1}^n x_i - n} \\ \ln L &= n \ln p + \left(\sum_{i=1}^n x_i - n \right) \ln(1-p) \\ \frac{\partial \ln L}{\partial p} &= \frac{n}{p} - \frac{\sum_{i=1}^n x_i - n}{1-p} \\ \frac{\partial \ln L}{\partial p} &= 0, \text{ za } p = \frac{n}{\sum_{i=1}^n x_i} \\ \hat{p} &= \frac{1}{\bar{x}} \end{aligned}$$

Ovaj zadatak ne možemo riješiti u R-u ali možemo provjeriti da dobiveno rješenje odgovara onome što bismo izračunali u R-u za neku konkretnu situaciju.

Generirat ćemo uzorak duljine 1000 iz $\mathcal{G}(0.2)$, izračunati \bar{x} i provjeriti da `MASS::fitdistr` daje $\frac{1}{\bar{x}}$. No trebamo biti pažljivi jer je u R-u geometrijska distribucija implementirana da počinje od 0 te stoga **donji kôd daje krivi rezultat**.

```
sample = rgeom(1000, 0.2)
1 / mean(sample)
```

```
## [1] 0.2443793
```

```
MASS::fitdistr(sample, 'geometric')
```

```
##      prob
## 0.196386489
## (0.005567179)
```

Ispod je modificirani kôd. Vidimo da sada dobivamo iste brojeve kao što smo i očekivali.

```
sample = rgeom(1000, 0.2)
1 / mean(sample + 1)
```

```
## [1] 0.1919754
```

```
MASS::fitdistr(sample, 'geometric')
```

```
##      prob
## 0.19197543
## (0.00545705)
```

(DZ) Pokažite da je procjenitelj najveće izglednosti (MLE) za parametar p geometrijske razdiobe koja počinje od 0 jednak $\frac{1}{\bar{x}+1}$.

Zadatak 3.

Iz generalnog skupa koji ima normalnu razdiobu s devijacijom $\sigma = 0.5$ i nepoznatim očekivanjem, izvučen je uzorak volumena 8: 16, 16, 16, 16.2, 16.2, 16.2, 16.5, 16.5. Odredite procjenu i 90% pouzdani interval za matematičko očekivanje a .

Rješenje

$$X \sim \mathcal{N}(a, \sigma^2), \quad \sigma = 0.5$$

Interval povjerenja reda $1 - \alpha$ za očekivanje normalne razdiobe, uz poznati σ^2 :

$$\mathbb{P}\left(\bar{x} - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < a < \bar{x} + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha = 0.9,$$

$$\alpha = 0.1, \quad 1 - \alpha/2 = 0.95, \quad u_{1-\alpha/2} = 1.645, \quad \bar{x} = 16.2, \quad n = 8$$

$$\mathbb{P}(15.9092 < a < 16.4908) = 0.9$$

Ovaj zadatak u R-u rješavamo isto kao i na papiru uz razliku da ne moramo posezati za tablicama jer R ima ugrađene funkcije za kvantile.

```
sample = c(16, 16, 16, 16.2, 16.2, 16.2, 16.5, 16.5)
sigma = 0.5
confidence = 0.9
n = length(sample)
mean = mean(sample)
alpha = 1 - confidence
d = sigma * qnorm(1 - alpha / 2) / sqrt(n)
c(mean - d, mean + d)
```

```
## [1] 15.90923 16.49077
```

Zadatak 4.

Iz populacije koja se podvrgava normalnom zakonu izvučen je sljedeći uzorak:

x_j	110	115	120	125	130	135
n_j	2	3	6	5	2	2

Izračunajte procjenu i 90% pouzdani interval za matematičko očekivanje i disperziju.

Rješenje

$$X \sim \mathcal{N}(a, \sigma^2)$$

Interval povjerenja reda $1 - \alpha$ za očekivanje normalne razdiobe, uz nepoznati σ^2 :

$$\mathbb{P}\left(\bar{x} - t_{n-1, 1-\alpha/2} \frac{\hat{s}}{\sqrt{n}} < a < \bar{x} + t_{n-1, 1-\alpha/2} \frac{\hat{s}}{\sqrt{n}}\right) = 1 - \alpha = 0.9,$$

$$\alpha = 0.1, \quad n = 20, \quad t_{n-1, 1-\alpha/2} = t_{19, 1-\alpha/2} = 1.729,$$

$$\bar{x} = \sum_{j=1}^6 n_j x_j = 122, \quad \hat{s}^2 = \frac{1}{n-1} \sum_{j=1}^6 n_j x_j^2 - n\bar{x}^2 = 51.05263, \quad \hat{s} = 7.145112$$

$$\mathbb{P}(119.2374 < a < 124.7626) = 0.9$$

I ovdje kao i u prethodnom zadatku bismo mogli čitav ovaj postupak prepisati u R uz korištenje funkcije `qt` za kvantile Studentove t -distribucije (**napravite to za DZ**). No moguće je dobiti i traženi interval za a u jednom koraku koristeći činjenicu da je kritično područje t -testa (koji je implementiran u R-u) zapravo traženi interval pouzdanosti. Više o vezi intervala pouzdanosti i statističkih testova možete pročitati u dodatku na kraju ovoga dokumenta.

U sljedećem kodu smo iskoristili funkciju `rep` koja ponavlja vrijednosti iz vektora `xi` s frekvencijama `fi` čime smo iz frekvencijske tablice kreirali uzorak iz kojeg je dobivena ta tablica.

```
xi = c(110, 115, 120, 125, 130, 135)
fi = c(2, 3, 6, 5, 2, 2)
sample = rep(xi, fi)
sample
```

```
## [1] 110 110 115 115 115 120 120 120 120 120 120 125 125 125 125 125 130
## [18] 130 135 135
```

Sada provedemo t -test uz razinu značajnosti 0.1 (koja odgovara pouzdanosti od 90%) i očitamo traženi pouzdani interval. Ovdje nije bitno što uzimamo za nul-hipotezu budući da nas zanima samo interval pouzdanosti.

```
t.test(sample, conf.level = 0.9)
```

```
##
## One Sample t-test
##
## data: sample
## t = 76.36, df = 19, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
## 119.2374 124.7626
## sample estimates:
## mean of x
## 122
```

Napomenimo još da t -test nismo mogli koristiti u zadatku 3 budući da je tamo disperzija bila poznata. Za zadatak 3 je bio potreban u -test (koji se u literaturi učestalije zove Z -testom) a koji pak nije implementiran u R-u.

Sada računamo interval povjerenja reda $1 - \alpha$ za disperziju normalne razdiobe, uz nepoznato očekivanje a :

$$\mathbb{P}\left(\frac{(n-1)\hat{s}^2}{\chi_{n-1,1-\alpha/2}^2} < \sigma^2 < \frac{(n-1)\hat{s}^2}{\chi_{n-1,\alpha/2}^2}\right) = 1 - \alpha = 0.9,$$

$$\chi_{n-1,1-\alpha/2}^2 = \chi_{19,0.95}^2 = 30.144, \quad \chi_{n-1,\alpha/2}^2 = \chi_{19,0.05}^2 = 10.117$$

$$\mathbb{P}(32.179 < \sigma^2 < 95.878) = 0.9,$$

Pouzdana interval za disperziju (varijancu) u R-u možemo računati “pješke”.

```
df = length(sample) - 1
var = var(sample)
confidence = 0.9
alpha = 1 - confidence
lower = var * df / qchisq(1 - alpha / 2, df)
upper = var * df / qchisq(alpha / 2, df)
c(lower, upper)
```

```
## [1] 32.17938 95.87810
```

No, postoji i funkcija koji može računati taj interval direktno, ali preko testa kojeg nećemo obrađivati u ovom kolegiju. Potrebna funkcija `varTest` se nalazi u paketu `EnvStats` i pristupamo joj naredbom `EnvStats::varTest`.

```
EnvStats::varTest(sample, conf.level = 0.9)
```

```
##
## Chi-Squared Test on Variance
##
## data: sample
## Chi-Squared = 970, df = 19, p-value < 2.2e-16
## alternative hypothesis: true variance is not equal to 1
## 90 percent confidence interval:
## 32.17938 95.87810
## sample estimates:
## variance
## 51.05263
```

Zadatak 5.

Kontrolom 100 žarulja iz određene velike pošiljke ustanovljeno je da ima 6 loših žarulja.

- Odredite 95%-tni interval povjerenja za postotak loših žarulja.
- Koliki broj žarulja n treba kontrolirati da bi se s pouzdanosti 0.95 moglo tvrditi da u čitavoj pošiljci nema *više od 9%* loših žarulja? Pretpostavite da se opet dobije isti točkovni procjenitelj $\hat{p} = 0.06$.

Rješenje

U ovom zadatku koristimo pouzdane intervale za vjerojatnost događaja p . Oni nisu egzaktni već se temelje na aproksimaciji binomne slučajne varijable normalnom distribucijom koja proizlazi iz centralnog graničnog teorema. Konkretno, za $X \sim \mathcal{B}(n, p)$ vrijedi $X \approx \mathcal{N}(np, np(1-p))$ odnosno

$$\frac{1}{n}X \approx \mathcal{N}\left(p, \frac{p(1-p)}{n}\right).$$

Ova aproksimacija nije najbolja ukoliko je n mali, ili ako je p preblizu 0 ili 1. Običaj je zahtijevati da je $np > 5$ i $n(1-p) > 5$.

- Označimo li s X *slučajan broj* loših žarulja od 100 ispitanih onda je $X \sim \mathcal{B}(100, p)$ gdje je p pravi postotak loših žarulja u čitavoj velikoj pošiljci. Procjenitelj $\hat{p} = \frac{1}{100}X$ je onda približno distribuiran kao $\mathcal{N}\left(p, \frac{p(1-p)}{100}\right)$ a 95%-pouzdana interval računamo kako slijedi:

$$\mathbb{P}(p_1 \leq p \leq p_2) = 1 - \alpha, \quad p_{1,2} = \hat{p} \mp u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} = \frac{m}{n},$$

$$1 - \alpha = 0.95, \quad \alpha = 0.05, \quad 1 - \alpha/2 = 0.975, \quad u_{1-\alpha/2} = 1.95996, \quad \hat{p} = 0.06, \quad n = 100$$

$$p_1 = 0.0134535, \quad p_2 = 0.106546$$

Ili u R-u:

```

xi = c('D', 'L')
fi = c(94, 6)
sample = rep(xi, fi)
n = length(sample)
confidence = 0.95
alpha = 1 - confidence
p_est = sum(sample == 'L') / n
d = qnorm(1 - alpha / 2) * sqrt(p_est * (1 - p_est) / n)
c(p_est - d, p_est + d)

```

```
## [1] 0.01345343 0.10654657
```

Već smo rekli da su ovi intervali pouzdanosti aproksimativni što znači da su izvedeni uz pretpostavku da promatrana varijabla X ima normalnu distribuciju. U R-u se mogu dobiti i pravi intervali pouzdanosti gdje je uzeta u obzir prava (binomna) distribucija varijable X . Trebamo samo iskoristiti funkciju `binom.test`

```

no_of_suceses = sum(sample == 'L')
binom.test(no_of_suceses, 100, alternative = 'two.sided', conf.level = confidence)

```

```

##
## Exact binomial test
##
## data: no_of_suceses and 100
## number of successes = 6, number of trials = 100, p-value < 2.2e-16
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.02233489 0.12602993
## sample estimates:
## probability of success
##                0.06

```

- (b) Zadatak nas pita da odredimo n takav da uz $\alpha = 0.05$ možemo odbaciti hipotezu $p \geq 0.09$. Očigledno moramo uzeti da je nul-hipoteza $\mathcal{H}_0 \dots p = 0.09$ uz alternativu $\mathcal{H}_1 \dots p < 0.09$. Radi se, jasno, o testu o proporciji. Testna statistika je

$$U = (\hat{p} - 0.09) \sqrt{\frac{n}{0.09 \cdot 0.91}}, \quad \hat{p} = 0.06$$

a \mathcal{H}_0 odbacujemo ukoliko je $U < -u_{1-\alpha}$. To znači da mora vrijediti

$$0.03 \sqrt{\frac{n}{0.09 \cdot 0.91}} > u_{0.95} = 1.644854$$

odnosno $n > 246.2045$, pa moramo ispitati barem 247 žarulja.

U R-u ne možemo direktno riješiti ovaj problem, ali možemo koristeći `binom.test` provjeriti da se naš n slaže s onim što bismo dobili testiranjem uzorka od 247 žarulja u koje smo prebrojali 15 loših ($247 \cdot 0.06 = 14.82$).

```
binom.test(15, 247, alternative = 'less', p = 0.09)
```

```

##
## Exact binomial test
##
## data: 15 and 247
## number of successes = 15, number of trials = 247, p-value =
## 0.06151
## alternative hypothesis: true probability of success is less than 0.09
## 95 percent confidence interval:

```

```
## 0.00000000 0.09197296
## sample estimates:
## probability of success
## 0.06072874
```

Vidimo da je p -vrijednost veća od $\alpha = 0.05$ što znači da ne bismo mogli odbaciti \mathcal{H}_0 , što bi onda upućivalo da smo krivo riješili zadatak i da je potreban veći n . No sjetite se da smo pri rješavanju koristili test proporcija u kojem je statistika samo aproksimativno normalna, dok R koristi egzaktnu distribuciju pa pogrešku možemo pripisati lošoj aproksimaciji.

Metodom pokušaja i pogreške (ili jednostavno zavrtimo petlju) možemo pronaći koji bi n bio dovoljan da egzaktni binomni test odbaci \mathcal{H}_0 .

```
for(n in 1:500){
  test = binom.test(ceiling(n*0.06), n, alternative = 'less', p = 0.09)
  if (test$p.value<0.05) {
    print(n)
    break
  }
}
```

```
## [1] 266
```


Auditorne vježbe (13. tjedan)

Zadatak 1.

Slučajna varijabla X je normalno distribuirana s nepoznatim očekivanjem i nepoznatom disperzijom. Uzorak od $n = 50$ mjerenja (zapisanih u datoteci `zad1.csv`) dao je srednju vrijednost $\bar{x} = 24.2$ i $\hat{s} = 4.8$. Uz nivo značajnosti $\alpha = 0.05$ testirajte hipotezu $\mathcal{H}_0 \dots a = 25$ uz alternativu $\mathcal{H}_1 \dots a < 25$.

Rješenje

Prvo učitajmo podatke iz datoteke i uvjerimo se da su \bar{x} i \hat{s} točno izračunati.

```
sample = read.csv('zad1.csv')$xi
n = length(sample)
c(mean(sample), sqrt(var(sample)))
```

```
## [1] 24.2 4.8
```

Koristimo t -test, uz statistiku

$$T = \frac{\bar{x} - a}{\hat{s}/\sqrt{n}} = \frac{24.2 - 25.0}{4.8/\sqrt{50}} = -\frac{5\sqrt{2}}{6} \approx -1.178511$$

Radi se o jednostranoj alternativni pa je kritično područje oblika $T < -t_{49,0.95}$. To znači da bismo, ukoliko je $T < -t_{49,0.95}$, nul-hipotezu morali odbaciti. Budući da je

$$T > -1.66 > -t_{60,0.95} > -t_{49,0.95}$$

onda \mathcal{H}_0 **ne odbacujemo**.

Umjesto korištenja tablica, u R-u bismo mogli direktno računati $-t_{49,0.95} = t_{49,0.05}$.

```
-qt(0.95,49)
```

```
## [1] -1.676551
```

No moguće je i sprovesti čitav test direktno iz podataka koristeći funkciju `t.test`.

```
alpha = 0.05
confidence = 1 - alpha
t.test(sample, mu = 25, alternative = 'less', conf.level = confidence)
```

```
##
## One Sample t-test
##
## data: sample
## t = -1.1785, df = 49, p-value = 0.1221
## alternative hypothesis: true mean is less than 25
## 95 percent confidence interval:
##      -Inf 25.33808
## sample estimates:
## mean of x
##      24.2
```

Uočite da je dobivena ista testna statistika $T = -1.1785$. No R neće zapisati kritični interval, već nam izračuna p -vrijednost. Sjetimo se da je **p -vrijednost broj koji označava najmanju razinu značajnosti uz koju bismo odbacili nul-hipotezu**. Dakle, ukoliko je p -vrijednost manja od zadane razine značajnosti α , odbacujemo \mathcal{H}_0 , u suprotnom je ne odbacujemo.

Drugi način kako se može definirati p -vrijednost jest da je to **vjerojatnost (uz pretpostavku da vrijedi \mathcal{H}_0) da se za testnu statistiku dobije upravo vrijednost koju smo dobili ili neku još ekstremniju vrijednost koja bi još više išla u prilog alternativu.**

U našem slučaju bi ekstremnije vrijednosti bile $T < -1.1785$ pa je p -vrijednost upravo broj p takav da je $t_{49,p} = -1.1785 = T$ i može ga se dobiti preko funkcije razdiobe t -distribucije.

```
pt(-1.1785, 49)
```

```
## [1] 0.1221444
```

Budući da je p -vrijednost veća od zadane razine značajnosti α ($0.1221 > 0.05$) nul-hipotezu \mathcal{H}_0 **ne odbacujemo**.

Uočite da što je p -vrijednost manja, to s većom pouzdanošću (tj. manjim α) možemo odbaciti nul-hipotezu. Obratno, što je p -vrijednost veća, to su dani podaci više u skladu s nul-hipotezom.

Zadatak 2.

Pseudoslučajnim generatorom simulirano je bacanje novčića 10000 puta. Pismo je registrirano 5120 puta. S kojim nivoom značajnosti možemo potvrditi hipotezu o ispravnosti generatora?

Rješenje

Koristit ćemo test o proporciji. Označimo li s p vjerojatnost da novčić padne na pismo, a s P broj pisama dobivenih u 10000 bacanja, vrijedi

$$P \sim \mathcal{B}(10000, p).$$

Postavimo sada nul-hipotezu $\mathcal{H}_0 \dots p = \frac{1}{2}$ u odnosu na (dvostranu) alternativu $\mathcal{H}_1 \dots p \neq \frac{1}{2}$.

Računamo testnu statistiku

$$U = \left(\frac{5120}{10000} - \frac{1}{2} \right) \sqrt{\frac{10000}{(1/2) \cdot (1/2)}} = 2.4$$

Budući da imamo dvostranu alternativu, kritično područje je oblika $|U| > u_{1-\alpha/2}$. Stoga ćemo potvrditi tezu o ispravnosti generatora (prihvatiti \mathcal{H}_0) ukoliko je $|2.4| \leq u_{1-\alpha/2}$ odnosno ukoliko je $0.992 \leq 1 - \alpha/2$, tj. $\alpha \leq 0.016$. Dakle, za nivo značajnosti $\alpha \leq 0.016$ možemo prihvatiti tezu o ispravnosti generatora. Ekvivalentno, za svaki $\alpha > 0.016$ bismo odbacili \mathcal{H}_0 pa je 0.016 ujedno i p -vrijednost ovog testa.

U R-u bismo sve ove korake mogli postepeno računati, no možemo i direktno sprovesti test. Prvo upišemo podatke:

```
xi = c('P', 'G')
fi = c(5120, 4880)
sample = rep(xi, fi)
#sample
```

a zatim koristeći `binom.test` testiramo našu hipotezu.

```
no_of_suceses = sum(sample == 'P')
n = length(sample)
binom.test(no_of_suceses, n)
```

```
##
## Exact binomial test
##
## data: no_of_suceses and n
## number of successes = 5120, number of trials = 10000, p-value =
```

```
## 0.01684
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
## 0.5021507 0.5218423
## sample estimates:
## probability of success
## 0.512
```

R koristi egzaktni binomni test dok smo mi koristili aproksimaciju binomne distribucije normalnom. Unatoč tome, uočite da se dobivena p -vrijednost 0.01684 približno podudara s našom.

Dodatno, ispisan je 95%-pouzdan interval za vjerojatnost pisma. Uočite da on ne sadržava vrijednost 0.5. To je stoga što bismo na razini značajnosti $\alpha = 0.05 = 1 - 0.95$ morali odbaciti \mathcal{H}_0 u korist \mathcal{H}_1 .

Zadatak 3.

Rezultati mjerenja slučajne varijable X dani su u tablici:

j	0	1	2	3	4
n_j	132	48	20	3	2

Pomoću χ^2 -testa provjerite hipotezu da se ovi podaci ravnaју po Poissonovoj razdiobi uz nivo značajnosti 0.05.

Rješenje

Provjeravamo hipotezu $X \sim \mathcal{P}(\lambda)$. Budući da je $\mathbb{E}X = \lambda$ uzimamo procjenitelj $\hat{\lambda} = \bar{x}$. Može se pokazati da je to procjenitelj najveće izglednosti (MLE) (**napravite to za DZ**).

$$\hat{\lambda} = \bar{x} = \frac{\sum j n_j}{\sum n_j} = \frac{105}{205} \approx 0.5122$$

```
xi = 0:4
fi = c(132, 48, 20, 3, 2)
n=sum(fi)
sample = rep(xi, fi)
lambda_est = mean(sample)
lambda_est
```

```
## [1] 0.5121951
```

ili

```
MASS::fitdistr(sample, 'poisson')
```

```
##      lambda
## 0.51219512
## (0.04998513)
```

Koristit ćemo χ^2 -test prilagodbe razdiobama. Sastavljamo tablicu

j	n_j	p_j	$n p_j$	$\Delta = n_j - n p_j$	$\Delta^2 / (n p_j)$
0	132	0.5992	122.83	9.17	0.68
1	48	0.3069	62.91	-14.91	3.54
2	20	0.0786	16.11	3.89	0.94
3	3	0.0134	2.75	0.25	0.02
≥ 4	2	0.0019	0.39	1.61	6.60

gdje je $p_j = \mathbb{P}(X = j) = \frac{\lambda^j e^{-\lambda}}{j!}$.

Brojeve u gornjoj tablici možemo računati u R-u ovako

```
probs = dpois(xi, lambda_est)
probs[5] = 1 - sum(probs[1:4])

fi

## [1] 132  48  20   3   2
probs

## [1] 0.599178863 0.306896491 0.078595443 0.013418734 0.001910468
n*probs

## [1] 122.831667  62.913781  16.112066   2.750840   0.391646
fi-n*probs

## [1]   9.1683330 -14.9137807   3.8879342   0.2491595   1.6083540
(fi-n*probs)^2/(n*probs)

## [1] 0.68433761 3.53532805 0.93818091 0.02256781 6.60494990
```

Sjetmo se da je statistika u ovom slučaju aproksimativno χ^2 i da bi ta aproksimacija bila valjana običaj je zahtijevati da *očekivane* (a ne opažene) frekvencije u svakom razredu budu barem 5. U našem slučaju to znači da moramo zajedno grupirati posljednja tri razreda u jedan. Nova tablica izgleda ovako:

j	n_j	p_j	$n p_j$	$\Delta = n_j - n p_j$	$\Delta^2 / (n p_j)$
0	132	0.5992	122.83	9.17	0.68
1	48	0.3069	62.91	-14.91	3.54
≥ 2	25	0.0939	19.25	5.75	1.72

Sada dobivamo χ^2 statistiku $\chi^2_q = 5.94$ koju uspoređujemo s $\chi^2_{1,0.95} \approx 3.841$ jer je broj stupnjeva slobode $f = 3 - 1 - 1 = 1$. Budući da je $5.94 > 3.841$ na razini značajnosti 0.05 **odbacujemo \mathcal{H}_0 da podaci dolaze iz Poissonove razdiobe**.

U R-u bismo ovaj test sproveli na sljedeći način.

```
chisq.test(fi, p = probs)

## Warning in chisq.test(fi, p = probs): Chi-squared approximation may be
## incorrect
##
## Chi-squared test for given probabilities
##
## data:  fi
## X-squared = 11.785, df = 4, p-value = 0.01902

Uočite da nas i R upozorava da aproksimacija može biti pogrešna jer nismo grupirali posljednja tri razreda u
jedan.

c(fi[1:2], sum(fi[3:5]))

## [1] 132  48  25
n*c(probs[1:2], sum(probs[3:5]))

## [1] 122.83167  62.91378  19.25455
```

```
chisq.test(c(fi[1:2], sum(fi[3:5])), p = c(probs[1:2], sum(probs[3:5])))
```

```
##
## Chi-squared test for given probabilities
##
## data: c(fi[1:2], sum(fi[3:5]))
## X-squared = 5.9341, df = 2, p-value = 0.05146
```

Sada vidimo da upozorenja više nema. No još uvijek imamo jedan problem. Naime, broj stupnjeva slobode df bi u našem slučaju trebao biti $3 - 1 - 1 = 1$, a ne 2 kako gore piše, budući da smo procjenjivali parametar λ iz podataka. Nažalost, funkcija `chisq.test` nema podržanu tu opciju pa ćemo morati sami provesti test do kraja zanemarujući gore dobivenu p -vrijednost (po kojoj **krivo** ispada da ne odbacujemo \mathcal{H}_0).

Ono što ipak možemo koristiti jest gore izračunata statistika $\chi_q^2 = 5.9341$ (koja se podudara s prije izračunatom). Pravu p -vrijednost onda dobijemo koristeći funkciju `pchisq` uz parametar $df=1$.

```
1-pchisq(5.9341, df = 1)
```

```
## [1] 0.01485065
```

Ekvivalentno, mogli smo odrediti kritično područje $(\chi_{1,0.95}^2, +\infty)$ koristeći funkciju `qchisq` opet uz $df=1$.

```
qchisq(0.95, df = 1)
```

```
## [1] 3.841459
```

Budući da je testna statistika u kritičnom području $5.9341 > \chi_{1,0.95}^2$ (ili ekvivalentno jer je p -vrijednost manja od zadane razine značajnosti $0.01485 < \alpha$) na zadanoj razini značajnosti **odbacujemo nul-hipotezu** da se podaci ravnaju po Poissonovoj razdiobi.

Zadatak 4.

U Mendelovim eksperimentima s graškom ispitano je 560 zrna i dobiveno je:

- 317 okruglih i žutih
- 109 okruglih i zelenih
- 102 smežuranih i žutih
- 32 smežurana i zelena.

Prema njegovoj teoriji o naslijeđivanju, ovi brojevi bi morali biti u omjeru $9 : 3 : 3 : 1$. S nivoom značajnosti 0.05 odredite treba li prihvatiti ovu pretpostavku.

Rješenje

Opet ćemo koristiti χ^2 -test prilagodbe razdiobama. Nul-hipoteza \mathcal{H}_0 jest da su vjerojatnosti dobivanja različitih tipova graška u omjeru $9 : 3 : 3 : 1$. Računamo tablicu

j	n_j	p_j	$n p_j$	$\Delta = n_j - n p_j$	$\Delta^2 / (n p_j)$
OY	317	9/16	315	2	0.013
OZ	109	3/16	105	4	0.152
SY	102	3/16	105	-3	0.086
SZ	32	1/16	35	-3	0.257

```
xi = c('OY', 'OZ', 'SY', 'SZ')
fi = c(317, 109, 102, 32)
n=sum(fi)
```

```

sample = rep(xi, fi)

probs = c(9,3,3,1)/16

fi

## [1] 317 109 102  32

probs

## [1] 0.5625 0.1875 0.1875 0.0625

n*probs

## [1] 315 105 105  35

fi-n*probs

## [1]  2  4 -3 -3

(fi-n*probs)^2/(n*probs)

## [1] 0.01269841 0.15238095 0.08571429 0.25714286

sum((fi-n*probs)^2/(n*probs))

## [1] 0.5079365

```

Dobivamo $\chi^2_q \approx 0.5079$, uz $f = 4 - 1 = 3$ stupnja slobode imamo $\chi^2_{3,0.95} = 7.8147$ pa **ne odbacujemo već prihvaćamo nul-hipotezu o omjerima**.

U R-u test izgleda ovako:

```

chisq.test(fi, p=probs)

##
##  Chi-squared test for given probabilities
##
## data:  fi
## X-squared = 0.50794, df = 3, p-value = 0.9171

```

Iz njega čitamo da je p -vrijednost veća od $\alpha = 0.05$ pa **ne odbacujemo već prihvaćamo \mathcal{H}_0** . Dodatno uočite da je p -vrijednost jako visoka što znači da se podaci jako dobro slažu s nul-hipotezom (a što je bilo vidljivo i iz tablice).

Zadatak 5.

4 kovana novčića bačena su istovremeno 96 puta i svaki puta je zabilježen broj grbova:

i	0	1	2	3	4
f_i	5	26	34	24	7

S nivoom značajnosti 0.05 odredite slažu li se dobiveni rezultati s hipotezom o ispravnosti svih novčića.

Rješenje

Označimo li s X broj grbova kad bacimo 4 novčića, zadatak nas traži da provjerimo da je $X \sim \mathcal{B}(4, 0.5)$. Za nul-hipotezu uzimamo $\mathcal{H}_0 \dots p_0 = 0.5$ gdje je $X \sim \mathcal{B}(4, p_0)$, a alternativu uzimamo dvostranu $\mathcal{H}_1 \dots p_0 \neq 0.5$.

Tablica glasi:

j	n_j	p_j	$n p_j$	$\Delta = n_j - n p_j$	$\Delta^2/(n p_j)$
0	5	1/16	6	-1	1/6
1	26	4/16	24	2	1/6
2	34	6/16	36	-2	1/9
3	24	4/16	24	0	0
4	7	1/16	6	1	1/6

gdje je $p_j = \mathbb{P}(X = j) = \frac{1}{2^4} \binom{4}{j} = \frac{3}{2 j! (4-j)!}$.

Dobivamo $\chi_q^2 = 11/18 \approx 0.611$, a uz $f = 5 - 1 = 4$ stupnja slobode imamo $\chi_{4,0.95}^2 = 9.4877$ pa vrijedi $\chi_q^2 < \chi_{4,0.95}^2$ pa na razini značajnosti od 0.05 **ne možemo odbaciti nul-hipotezu da su svi novčići ispravni**.

U R-u bismo test proveli ovako:

```
xi=0:4
fi = c(5, 26, 34, 24, 7)
probs = dbinom(xi, 4, 0.5)
chisq.test(fi, p=probs)
```

```
##
## Chi-squared test for given probabilities
##
## data:  fi
## X-squared = 0.61111, df = 4, p-value = 0.9618
```

Kako je p -vrijednost $0.9618 > 0.05$ to opet na razini značajnosti $\alpha = 0.05$ **ne odbacujemo (već prihvaćamo) nul-hipotezu da su svi novčići ispravni**.

Zadatak 6.

220 puta bačeno je 5 novčića i blježen je broj X pojavljivanja grbova:

x_j	0	1	2	3	4	5
n_j	6	32	71	69	35	7

Pomoću χ^2 -testa provjerite hipotezu da X ima binomnu razdiobu s parametrom $p = \frac{1}{2}$ uz nivo značajnosti $\alpha = 0.1$.

Rješenje

Zadatak je posve analogan prethodnome. Sada je $X \sim \mathcal{B}(5, p_0)$. Za nul-hipotezu uzimamo $\mathcal{H}_0 \dots p_0 = 0.5$, a za dvostranu alternativu $\mathcal{H}_1 \dots p_0 \neq 0.5$.

Tablica glasi:

j	n_j	p_j	$n p_j$	$\Delta = n_j - n p_j$	$\Delta^2/(n p_j)$
0	6	1/32	6.875	-0.875	0.111
1	32	5/32	34.375	-2.375	0.164
2	71	10/32	68.750	2.250	0.074
3	69	10/32	68.750	0.250	0.001
4	35	5/32	34.375	0.625	0.011
5	7	1/32	6.875	0.125	0.002

gdje je $p_j = \mathbb{P}(X = j) = \frac{1}{2^4} \binom{4}{j} = \frac{3}{2 j! (4-j)!}$.

Dobivamo $\chi_q^2 \approx 0.364$, a uz $f = 6 - 1 = 5$ stupnjeva slobode imamo $\chi_{5,0.9}^2 = 9.236$ pa vrijedi $\chi_q^2 < \chi_{5,0.9}^2$ te na razini značajnosti $\alpha = 0.1$ **ne odbacujemo nul-hipotezu da X ima binomnu razdiobu $\mathcal{B}(5, 0.5)$.**

U R-u bismo test proveli ovako:

```
xi=0:5
fi = c(6, 32, 71, 69, 35, 7)
probs = dbinom(xi, 5, 0.5)
chisq.test(fi, p=probs)

##
## Chi-squared test for given probabilities
##
## data:  fi
## X-squared = 0.36364, df = 5, p-value = 0.9963
```

Kako je p -vrijednost $0.9963 > 0.1$ to na razini značajnosti $\alpha = 0.1$ **ne odbacujemo nul-hipotezu da X ima binomnu razdiobu $\mathcal{B}(5, 0.5)$.** Uočimo još da je p -vrijednost jako velika što govori o jako dobrom podudaranju podataka s nul-hipotezom (što i dalje ne znači da je nul-hipoteza točna).

Napomena: „Ne odbacujemo \mathcal{H}_0 ” nasuprot „prihvaćamo \mathcal{H}_0 ”

Iako su ove dvije formulacije naizgled logički ekvivalentne, kod statističkog testiranja ipak preferiramo prvu. *Statističkim testom samo možemo odbacivati hipoteze koje nisu u skladu s opažanjima.*

Čitava stvara pomalo podsijeca na dokaz metodom kontradikcije. Prvo pretpostavimo da vrijedi \mathcal{H}_0 . Zatim logički dobijemo da to vodi na kontradikciju (dobijemo atipičnu vrijednost neke testne statistike, odnosno malu p -vrijednost), iz čega zaključujemo da tvrdnja od koje smo krenuli mora biti kriva i zato odbacujemo \mathcal{H}_0 .

No ukoliko ne dodemo do kontradikcije (ako vrijednost testne statistike nije dovoljno atipična, tj. ako p -vrijednost nije mala), to nije dokaz da je \mathcal{H}_0 točna. Možda se samo nismo dovoljno potrudili da nađemo kontradikciju. U tom slučaju zapravo ne možemo zaključiti ništa, pa je ispravnije reći da ne isključujemo mogućnost da je \mathcal{H}_0 točna, odnosno da ne odbacujemo \mathcal{H}_0 .

Primjer.

Pogledajmo još jednom Zadatak 2 u kojem se novčić baca 10000 puta, od kojih je pismo registrirano 5120 puta. Označimo s p vjerojatnost da novčić padne na glavu. Na razini značajnosti $\alpha = 0.05$ napraviti ćemo dva testa.

Test 1	Test 2
$\mathcal{H}_0: p = 0.51$	$\mathcal{H}_0: p = 0.52$
$\mathcal{H}_1: p \neq 0.51$	$\mathcal{H}_1: p \neq 0.52$
p -vrijednost = $0.6965 > \alpha$	p -vrijednost = $0.1093 > \alpha$
ne odbacujemo $\mathcal{H}_0: p = 0.51$	ne odbacujemo $\mathcal{H}_0: p = 0.52$

Uočite da bismo, koristeći drugu formulaciju, rekli da zbog Testa 1 prihvaćamo $p = 0.51$ i istovremeno zbog Testa 2 prihvaćamo $p = 0.52$, što nema smisla. Točnije je reći da na temelju gornja dva testa uz značajnost 0.05 ne možemo odbaciti ni $p = 0.51$ ni $p = 0.52$.

```
xi = c('P', 'G')
fi = c(5120, 4880)
sample = rep(xi, fi)

no_of_suceses = sum(sample == 'P')
```



```

n = length(sample)
binom.test(no_of_succeses, n, p=0.51)

##
## Exact binomial test
##
## data: no_of_succeses and n
## number of successes = 5120, number of trials = 10000, p-value =
## 0.6965
## alternative hypothesis: true probability of success is not equal to 0.51
## 95 percent confidence interval:
## 0.5021507 0.5218423
## sample estimates:
## probability of success
## 0.512

binom.test(no_of_succeses, n, p=0.52)

##
## Exact binomial test
##
## data: no_of_succeses and n
## number of successes = 5120, number of trials = 10000, p-value =
## 0.1093
## alternative hypothesis: true probability of success is not equal to 0.52
## 95 percent confidence interval:
## 0.5021507 0.5218423
## sample estimates:
## probability of success
## 0.512

```

Dodatak: O intervalima pouzdanosti

Intervali pouzdanosti su intervalni procjenitelji parametara modela.

Pretpostavimo da nam je dan neki konačan uzorak (x_1, x_2, \dots, x_n) za koji pretpostavljamo da predstavlja n **nezavisnih realizacija iste slučajne varijable** X .

Primjerice X može biti temperatura zraka, danas u 22 sata, slučajno (uniformno) izabrane lokacije na teritoriju Hrvatske. Ključno je uočiti da ta slučajna varijabla ima dobro definiranu distribuciju, očekivanje, varijancu (disperziju), ... Te parametre bismo mogli točno odrediti ako bismo izmjerili temperaturu u 22 sata za svaku točku teritorija. No radi jednostavnosti mi odredimo nasumce 5 lokacija i pošaljemo petoro meteorologa da provjere temperature u 22 sata, time dobivamo uzorak (x_1, x_2, \dots, x_5) .

Iz tog uzorka ne možemo izračunati koja je, primjerice, prava srednja vrijednost temperature nad hrvatskom (matematičko očekivanje od X), no srednja vrijednost uzorka $\frac{x_1+x_2+\dots+x_5}{5}$ je dobra točkovna procjena za tu vrijednost. Postavlja se pitanje koliko je dobra ta procjena.

Primjerice, ako se dogodi da smo sve lokacije izabrali nad Velebitom, vjerojatno će naša procjena biti daleko niža od pravog prosjeka. Važno je uočiti da je šansa za to mala ako smo slučajno birali lokacije. Vjerojatnije da će lokacije biti raspoređene svuda po karti.

S tim u vezi definiraju se intervalne procijene. $(1 - \alpha) \cdot 100\%$ -pouzdan interval za neki parametar od X je zapravo algoritam kojim kreiramo interval iz slučajnog uzorka tako da dobiveni interval u (barem) $(1 - \alpha) \cdot 100\%$ slučajeva prekriva pravu vrijednost tog parametra.

Obično se radi jednostavnosti još dodatno pretpostavi da je nepoznata distribucija (u našem slučaju distribucija temperatura X) iz neke klase poznatih distribucija parametriziranih nekim parametrima. Na primjer, možda je razumno pretpostaviti da je $X \sim \mathcal{N}(\mu, \sigma^2)$ za neke $\mu \in \mathbb{R}$ i $\sigma > 0$. U tom slučaju 90% pouzdani interval za μ je

$$\left(\bar{x} - t_{n-1, 0.95} \frac{\hat{s}}{\sqrt{n}}, \bar{x} + t_{n-1, 0.95} \frac{\hat{s}}{\sqrt{n}} \right).$$

Ključno je uočiti da su μ i σ fiksni iako nama nepoznati. Zamislimo sada da u 100 paralelnih svemira s identičnim temperaturama pošaljemo u svakom svemiru naših pet meteorologa da mjere vrijednosti temperature u 22 sata po Hrvatskoj ali u svakom svemiru nezavisno izaberemo 5 lokacija na koje ih šaljemo. Možda smo u nekom svemiru sve poslali na Velebit, ali u većini svemira će ih ipak biti posvuda. I u svakom od svemira izmjerimo pripadni 90%-pouzdan interval. Time smo dobili 100 različitih (konkretnih, numeričkih) intervala. Interval koji smo dobili iz svemira gdje su svi meteorolozi otišli na Velebit možda neće sadržavati pravu vrijednost μ no ipak bi u otprilike 90ak svemira dobiveni interval trebao pokrivati konkretan μ . Ta proporcija u koliko posto svemira dobiveni interval zbilja sadrži pravu vrijednost parametra jest upravo što znači pouzdanosti intervala.

Ponekad se kaže da je $p = 1 - \alpha$ vjerojatnost da parametar koji procjenjujemo „upadne” u taj interval, no trebalo bi biti jasno iz gornje interpretacije da je stvar zapravo obrnuta. $p = 1 - \alpha$ je vjerojatnost da interval pokrije pravu vrijednost parametra (koja je fiksna iako nama nepoznata). Interval je taj koji je „slučajan” i varira od realizacije do realizacije, od jednog do drugog paralelnog svemira, dok je prava vrijednost parametra fiksna iako nama nepoznata.

Pogledajmo kako ove stvari izgledaju na primjeru. Izgenerirat ćemo 100 uzoraka, svaki od po pet realizacija iz distribucije $\mathcal{N}(23, 4^2)$. Sve realizacije i uzorci su međusobno nezavisni. Zamislite da svakom od stotinu svojih prijatelja date po set od 5 uzoraka i zatražite ih da pogode koji ste μ koristili za generiranje uzorka. Pri tom im samo kažete da ste uzorak generirali iz jedne normalne distribucije. Svaki od njih na gornji način odredi svoj 90%-pouzdan interval za μ . Pogledajmo što bi se moglo dogoditi

```
total_in = 0
repetition = 100
df = data.frame(1:5, row.names = 1)
ci_df = data.frame(1:3, row.names = 1)
```

```

for (i in 1:repetition) {
  sample = rnorm(5, 23, 4)
  df[i] = sample
  n = length(sample)

  mean_est = mean(sample)
  s_est = sqrt(var(sample))
  d = qt(0.95, n - 1) * s_est / sqrt(n)
  is_in = 23 < mean_est + d && 23 > mean_est - d

  ci_df[i] = c(mean_est - d, mean_est + d, is_in)
  total_in = total_in + is_in

  #print(sample)
  #print(c(mean_est - d, mean_est + d))
  #print(is_in)
}
total_in

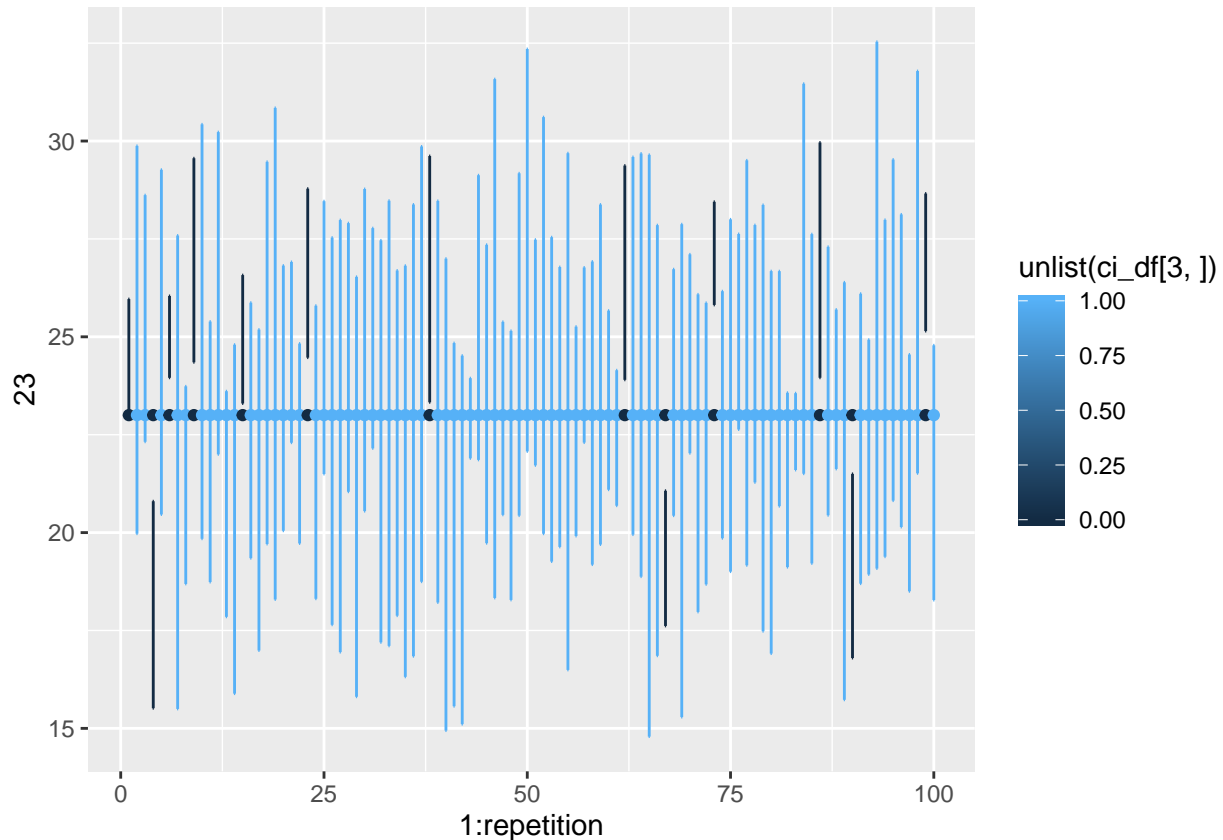
```

```
## [1] 87
```

```

library(ggplot2)
qplot(x = 1:repetition, y = 23,
      color = unlist(ci_df[3, ])) + geom_errorbar(aes(
        ymin = unlist(ci_df[1, ]),
        ymax = unlist(ci_df[2, ]), width = 0.15
      ))

```



Na gornjoj slici iscrtali smo intervale pouzdanosti koje je svaki od 100 prijatelja izračunao. Crno su označeni intervale koji ne sadrže vrijednost $\mu = 23$. Uočite da je takvih intervala po prilici 10%.

Uočite još jednu zanimljivu stvar, svaki od naših prijatelja će tvrditi s 90 postotnom sigurnošću da baš njegov interval sadrži pravu vrijednost μ , no otprilike 10 njih od 100 će zapravo biti u krivu (baš zato što su to 90%-pouzdana intervali).

Analogno, zamislamo sada da uzmemo 100 znanstvenih članaka koji su dokazali svoju hipotezu na razini značajnosti $\alpha = 0.1$ (odnosno s pouzdanošću od 90%), onda je, kao i gore, skoro pa garantirano da oko 10ak od tih 100 članaka sadrži krive zaključke, čak i ukoliko su znanstvenici koji su ih pisali u potpunosti korektno proveli svoje statističke testove.

Veza statističkih testova i intervala pouzdanosti

Statistički testovi su usko povezani s intervalima pouzdanosti. Nadovežimo se na gornji primjer. Već spomenutoj stotini svojih prijatelja dali ste po pet nezavisnih uzoraka iz $\mathcal{N}(23, 4^2)$ bez da ste im otkrili koji ste μ i σ koristili za generiranje, no rekli ste im da su podaci došli iz neke normalne razdiobe. Zatim ih zatražite da t -testom testiraju hipotezu $\mathcal{H}_0 \dots \mu = 23$ u odnosu na alternativu $\mathcal{H}_1 \dots \mu \neq 23$ na razini značajnosti $\alpha = 0.1$.

Svaki od njih će izračunati T statistiku:

$$T = \frac{\bar{x} - 23}{\hat{s}/\sqrt{n}}$$

i provjeriti da li ona upada u kritično područje koje je za dvostrani t -test dano s

$$|T| > t_{n-1, 1-\alpha/2} \quad \text{tj.} \quad |T| > t_{4, 0.95}.$$

U slučaju da upada, oni će odbaciti \mathcal{H}_0 a u suprotnom, ako dobiju $|T| \leq t_{4, 0.95}$, neće odbaciti već prihvatiti \mathcal{H}_0 . Uočite da vrijedi $|T| \leq t_{4, 0.95}$ ako i samo ako je

$$\bar{x} - t_{n-1, 0.95} \frac{\hat{s}}{\sqrt{n}} \leq 23 \leq \bar{x} + t_{n-1, 0.95} \frac{\hat{s}}{\sqrt{n}}.$$

Drugim riječima, točno oni prijatelji čiji 90%-pouzdana intervali pokrivaju broj 23, sada su dobili da prihvaćaju nul-hipotezu $\mu = 23$. Istovremeno, oni prijatelji čiji 90%-pouzdana interval nije pokrio broj 23, su (*krivo!*) odbacili nul-hipotezu $\mu = 23$ u odnosu na alternativu $\mu \neq 23$.

Uočite da je ovdje \mathcal{H}_0 bila zbilja ispravna, no oko 90% prijatelja je počinilo pogrešku prve vrste jer je odbacilo točnu nul-hipotezu.