SP1 - šeste vježbe

Prilagodba modela podacima (2)

Za svaki skup podataka najprije navedite što želite provjeriti. Zatim pregledajte odgovarajući Q–Q graf, navedite hipoteze formalnog testa, prikažite p-vrijednost i zaključite na razini 0.05. Razlikujte neodbacivanje normalnosti od potvrde normalnosti. U simulacijama zadržite zadani početni broj kako biste mogli usporediti rezultate.

Zadatak 1. Generirajte 100 opažanja iz eksponencijalne distribucije s parametrom `rate = 2`. Provedite `ks.test(x, "pexp", rate = 2)`, a zatim testirajte dolaze li podaci iz Exp(1) distribucije. Navedite obje nul-hipoteze i zaključke.

Zadatak 2.

```
set.seed(621)
a <- rnorm(150, 20, 4)
b <- -rexp(150)
c <- rt(150, df = 4)
d <- runif(150, -3, 3)
```

1. Prije crtanja predvidite oblik svakog normalnog Q–Q grafa.
2. Nacrtajte grafove i provedite Lillieforsove testove. Jesu li grafovi suglasni s odlukama?
3. Usporedite Q–Q grafove za `a` i `10 + 3*a`. Mijenja li linearna transformacija normalnost i Lillieforsovu p-vrijednost?

Zadatak 3.

```
set.seed(622)
x <- rnorm(120, mean = 2, sd = 1)
```

1. Testirajte tvrdnju da je distribucija $N(0, 1)$ odgovarajućim KS-testom. Zatim testirajte pripadnost obitelji normalnih distribucija.
2. Objasnite kako je moguće odbaciti prvu, a ne odbaciti drugu hipotezu. Usporedite to s Q–Q grafom.
3. Objasnite zašto `ks.test(x, "pnorm", mean(x), sd(x))` nije ispravan odgovor na drugo pitanje iz prvog zadatka.

Zadatak 4.

Sljedeće linije koda 1000 puta ponavljaju lillie test (funkcija `replicate`), a onda izvršavaju to za različite iznose n (duljine uzorka) – 10, 30, 100 i 300 (uz pomoć funkcije `sapply`).

```
set.seed(623)
B <- 1000
udio <- function(n) {
  mean(replicate(B,
    lillie.test(rexp(n))$p.value <= 0.05))
}
sapply(c(10, 30, 100, 300), udio)
```

Što procjenjuje svaki rezultat? Jesu li podaci ovdje normalno distribuirani? Ponovite postupak zamjenom `rexp(n)` s `rnorm(n)`. Što sada procjenjujete, kako se promijenio rezultat funkcije udio? Objasnite razliku između povećavanja n i povećavanja B.

Zadatak 5. Na 40 nezavisno odabranih uređaja mjerimo vrijeme izvođenja prije i poslije promjene. Par čine dva mjerenja istog uređaja.

```
set.seed(624)
prije <- 100 + rexp(40, rate = 0.05)
razlika <- rnorm(40, mean = 3, sd = 2)
poslije <- prije - razlika
```

1. Nacrtajte tri Q–Q grafa: za `prije`, `poslije` i `prije - poslije`. Koji je važan za upareni t-test?
2. Provedite Lillieforsov test na relevantnom nizu. Znači li nenormalnost prvog niza (`prije`) da upareni t-test nije primjeren?
3. Testirajte skraćuje li promjena očekivano vrijeme izvođenja. Navedite smjer razlike i alternativu prije poziva `t.test`.

Zadatak 6.

```
set.seed(625)
x <- rnorm(60, 10, 1)
y <- rnorm(60, 15, 1)
```

Pregledajte Q–Q grafove za `x`, `y` i `c(x,y)`. Koji su grafovi relevantni za usporedbu očekivanja Welchovim testom? Provedite test i objasnite zašto oblik spojenog uzorka ne pobija normalni model unutar skupina.

Zadatak 7. Za uzorak od 12 opažanja Lillieforsov test daje p-vrijednost 0.28, a Q–Q graf pokazuje jednu vrlo udaljenu točku. Za drugi uzorak od 5000 opažanja p-vrijednost je 0.002, ali graf ima tek blagu zakrivljenost. Možemo li iz samih p-vrijednosti zaključiti u kojem je slučaju t-test pouzdaniji? Što biste još provjerili? Kako bi se oprez promijenio kada bismo umjesto očekivanja uspoređivali varijance F-testom?