c5

filippo

18 May 2018

Es 5. 'Women who are union members earn 2.50$ per hour more than women who are not union members'.

(The Wall Street Journal, July 26, 1994). Sembrerebbe quindi che per le donne statunitensi sia conveniente far parte di un sindacato. Per verificare l'affermazione del Wall Street Journal abbiamo scelto due campioni indipendenti di lavoratrici del settore industriale. Il primo campione è costituito da 15 lavoratrici iscritte ad un sindacato, mentre il secondo campione è costituito da 20 lavoratrici che non fanno parte di nessun sindacato. Su ciascuna unità statistica (la lavoratrice) abbiamo misurato il salario orario (in $). I dati sono contenuti nel dataset `workers.txt` (variabili `salary` e `union`).

```
## Loading required package: MASS
```

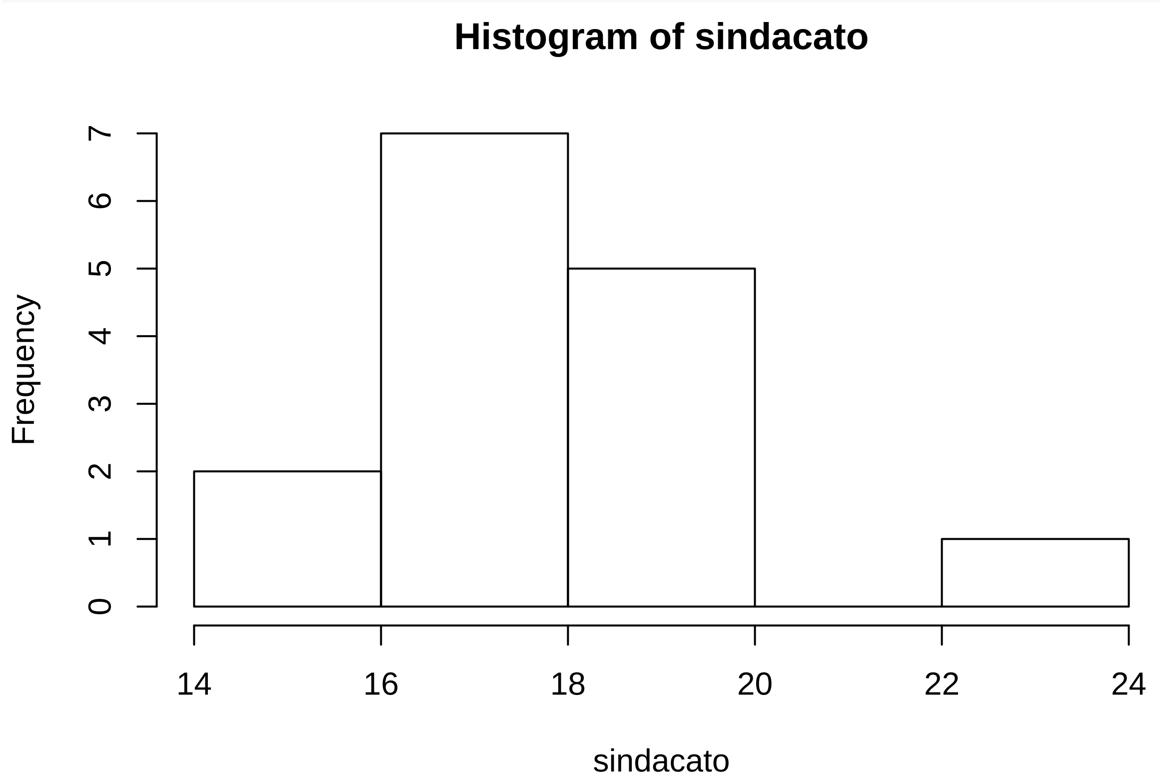
```
## Loading required package: survival
```

(a)

Considerate separatamente le due classi di lavoratrici e verificate se ci sono dei valori anomali.

```
sindacato <- workers[workers$status == "union",1]
senza.sindacato <- workers[workers$status != "union",1]

hist(sindacato)
```

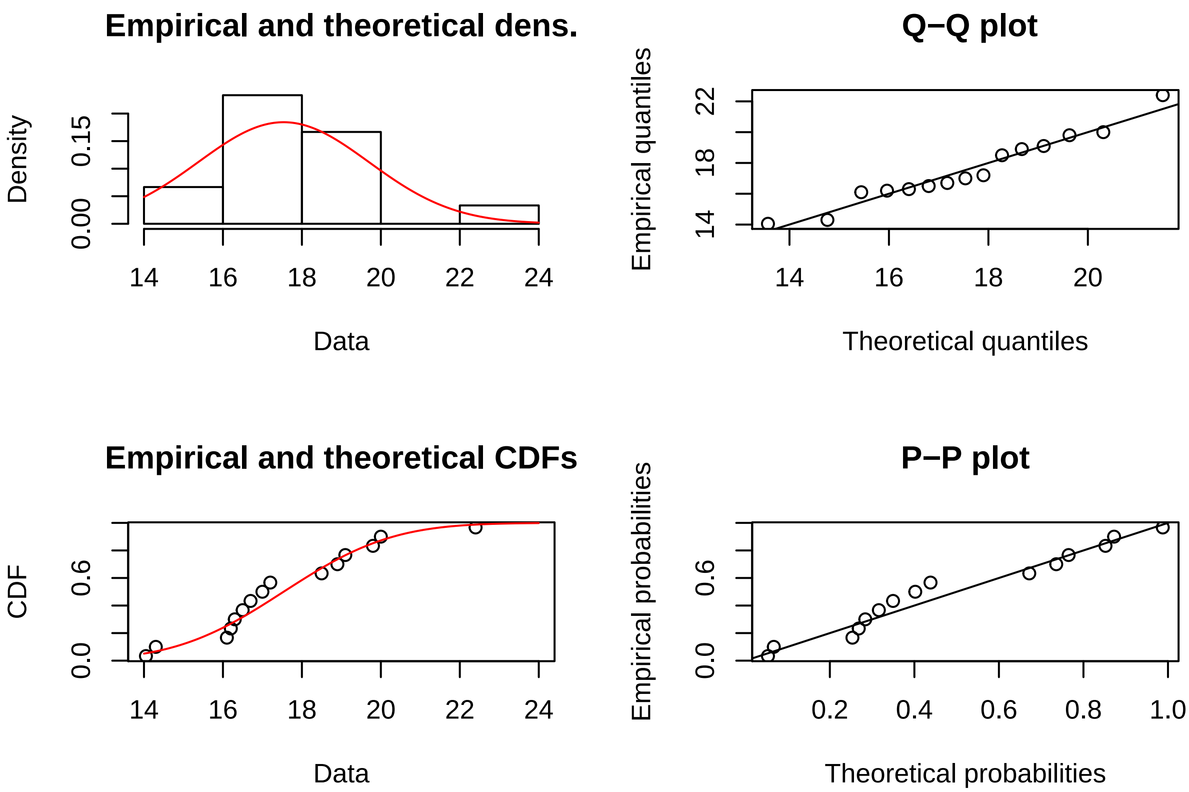


```r
hist(senza.sindacato)
```

(b)

Verificate se potete supporre che i dati siano assimibilabili ad un campione da una v.c. normale

```r
fit.norm.mle <- fitdist(sindacato,distr = "norm")
plot(fit.norm.mle)
```




```
fit.norm.mle
```

```
## Fitting of the distribution ' norm ' by maximum likelihood
## Parameters:
##       estimate Std. Error
## mean 17.536667  0.5588175
## sd    2.164291  0.3951433
```

(c)

Se $X \sim F_X$ rappresenta la variabile casuale salario delle donne iscritte e la distribuzione della e $Y \sim F_Y$ rappresenta la variabile casuale salario delle donne non iscritte, si proponga uno stimatore per sostituzione del parametro $\theta = E_{F_X}(X) - E_{F_Y}(Y)$.

$$\hat{\theta}_n = \hat{\mu_X} - \hat{\mu_Y}$$

(d)

Sia $\hat{\theta}_n$ lo stimatore ottenuto al punto precedente. Si scriva una prodedura bootstrap per il calcolo della varianza dello stimatore.

```
B <- 1000

sindacato.values <- matrix(sample(sindacato, B*length(sindacato),replace = TRUE), ncol = length(sindaca
no.sindacato.values <- matrix(sample(senza.sindacato, B*length(senza.sindacato),replace = TRUE), ncol =

boot.sindacato.estimatore <- apply(sindacato.values,1,mean)
```

```r
boot.no.sindacato.estimatore <- apply(no.sindacato.values,1,mean)

boot.estimatore <- boot.sindacato.estimatore - boot.no.sindacato.estimatore

v.boot <- mean(boot.estimatore^2) - mean(boot.estimatore)^2
v.boot
```

```
## [1] 0.556178
```

(e)

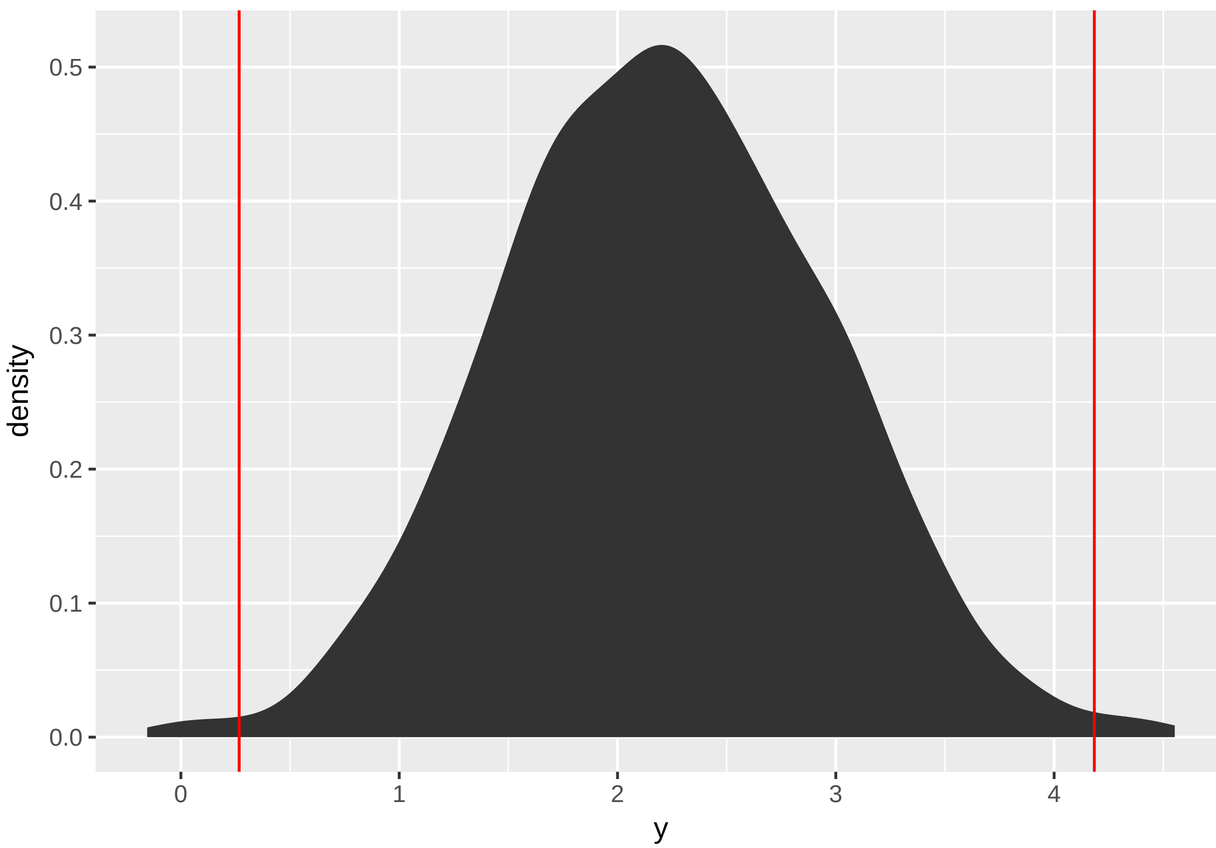
Si calcoli un intervallo di confidenza per θ di livello 0.99.

```r
boot.theta <- boot.sindacato.estimatore-boot.no.sindacato.estimatore

alpha <- 0.01
limiti <- quantile(boot.theta,prob=c(alpha/2,1-alpha/2))
limiti
```

```
##      0.5%     99.5%
## 0.2668833 4.1840875
```

```r
p <- ggplot(data.frame(y = boot.theta), aes(y))
p <- p + stat_density()
p <- p + geom_vline(xintercept = limiti,col="red")
p
```



(f)

Si stimi con il metodo bootstrap la probabilità che $P(\hat{\theta}_n > 0)$.

```
mean(boot.theta>0)
```

```
## [1] 0.999
```