

# B13

*filippo*

*5 April 2018*

## Contents

<b>Consegna</b>	<b>1</b>
(a) Importate il file in R. . . . .	1
(b) Disegnate un istogramma dei dati e confrontatelo con una stima della funzione di densità. . . .	1
(c) Proponete un intervallo di confidenza per la spesa media. . . . .	4
<b>lim.inf &lt;-</b>	<b>4</b>
(d) Quali potrebbero essere gli elementi di debolezza della vostra scelta ? . . . . .	4
(e) Trasformate ora i dati secondo queste tre opzioni: a) $\log(y_i)$ , b) $\sqrt{y_i}$ , c) $y_i^{1/3}$ . . . . .	4
(f) A vostro parere quale di queste trasformazioni rende la distribuzione dei dati trasformati più simile a quella di una v.c. normale ? . . . . .	4

## Consegna

La stagione di vendita al dettaglio delle vacanze del 2009, che ha segnato il 27 novembre 2009 (il giorno successivo al Giorno del Ringraziamento), è stata contrassegnata da una spesa per consumi auto-segnalata leggermente inferiore a quella osservata nel periodo comparabile del 2008. Per ottenere una stima della spesa dei consumatori, sono stati intervistati 436 adulti americani campionati a caso. E' stata esaminata la spesa giornaliera dei consumatori per il periodo di sei giorni successivo al Ringraziamento, che si è estesa al weekend del Black Friday e al Cyber Monday e i dati sono contenuti nel file `thanksgiving_spend.csv`.

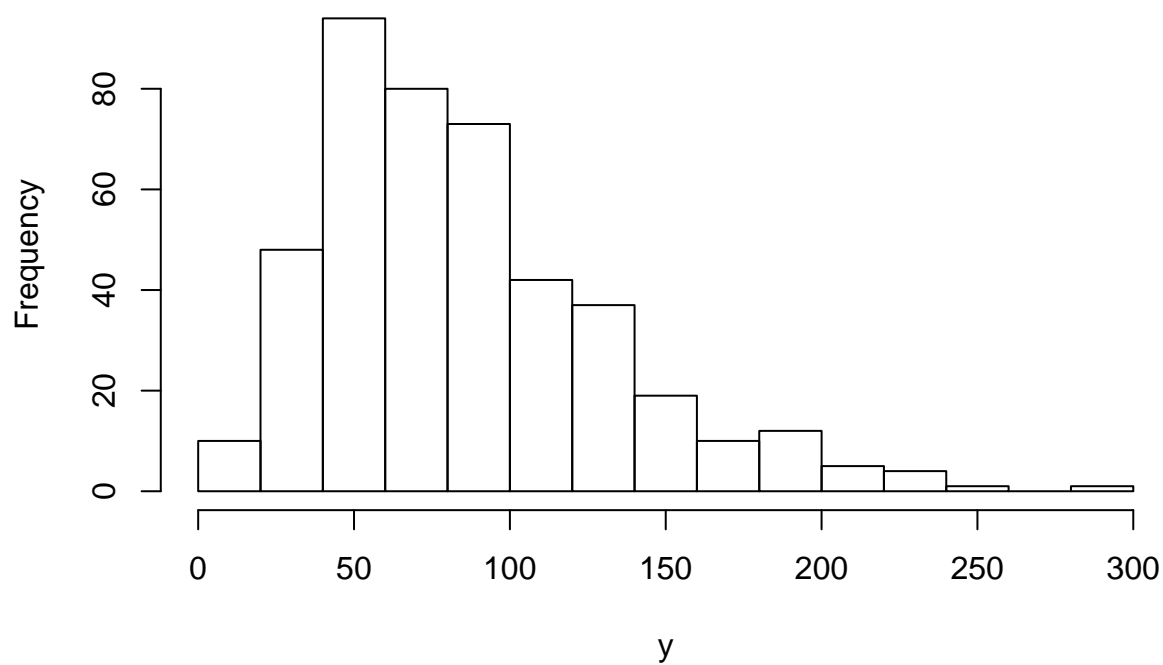
### (a) Importate il file in R.

```
data <- read.csv("thanksgiving_spend.csv")
y <- data$spending
```

### (b) Disegnate un istogramma dei dati e confrontatelo con una stima della funzione di densità.

```
hist(y)
```

**Histogram of y**

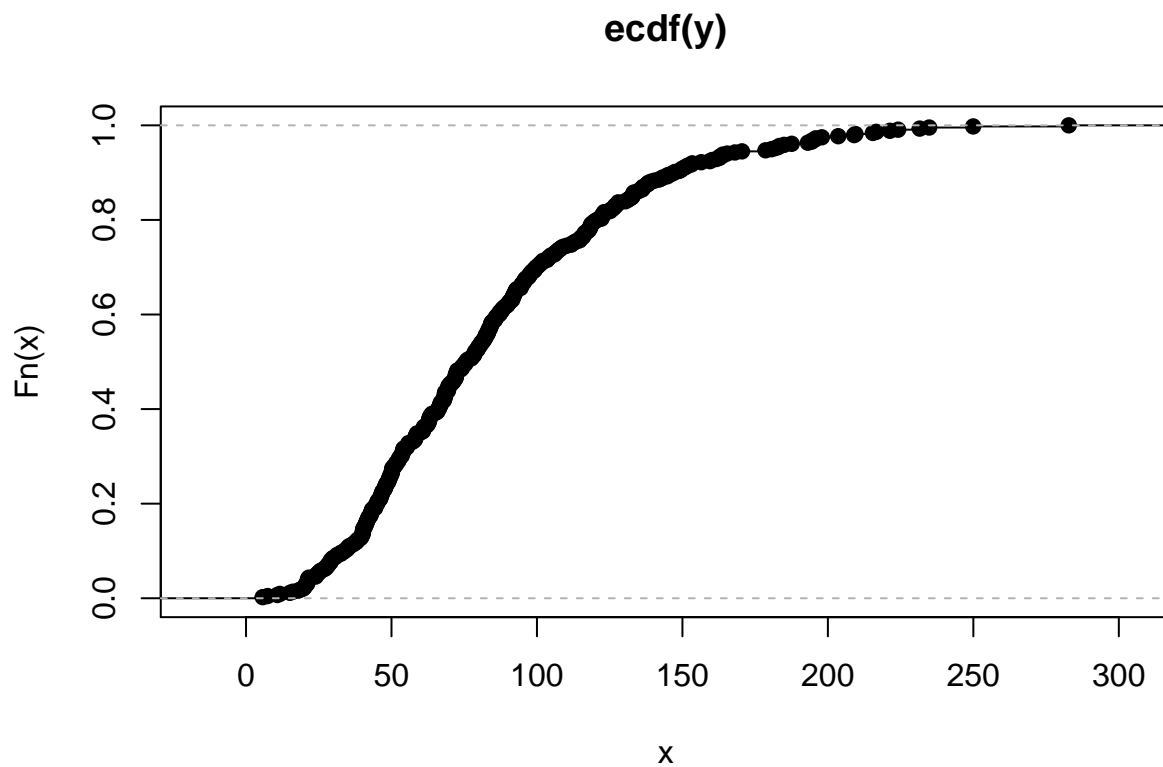


```
plot(ecdf(y))

alpha <- 0.05

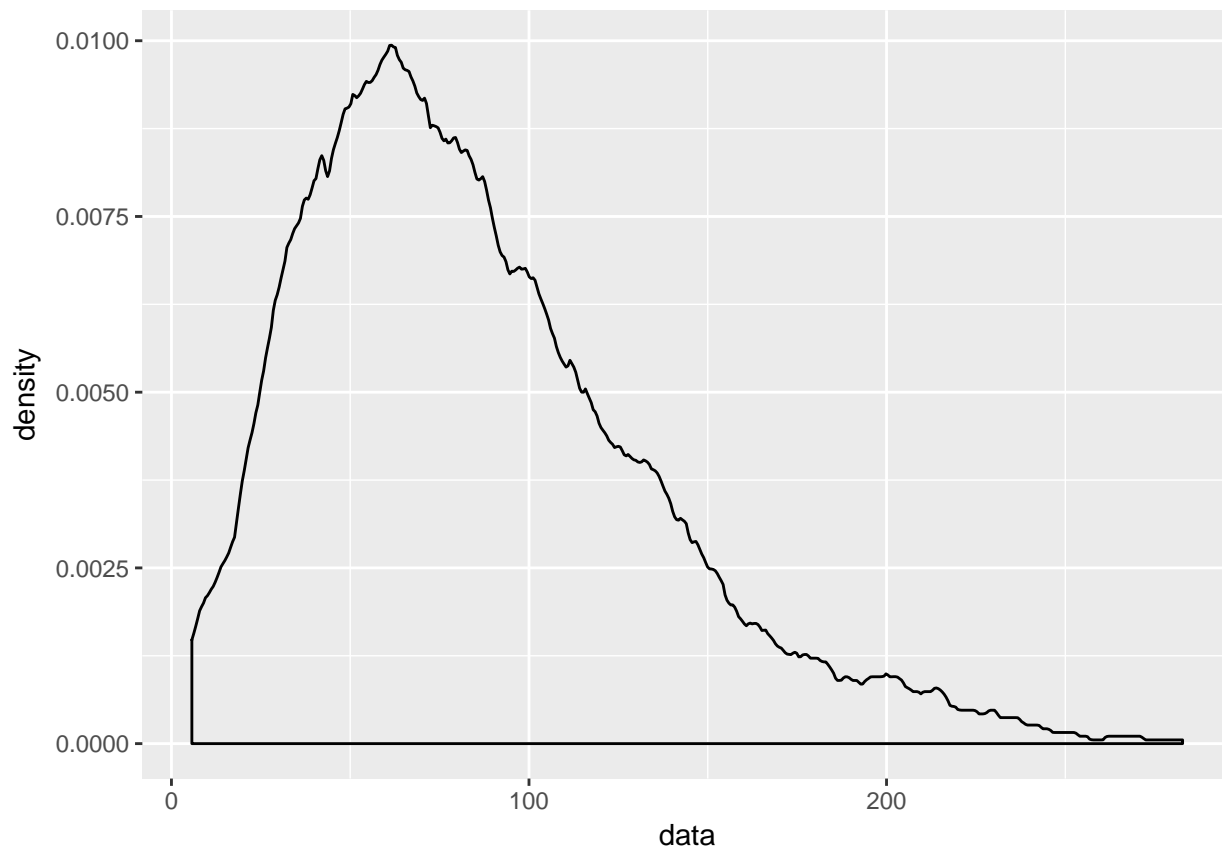
theta.hat <- mean(y)
sigma.hat <- var(y)

library(ggplot2)
```



```
library(data.table)

p <- ggplot(data.table(data=y), aes(x=data))
p + geom_density(kernel = "rectangular")
```



(c) Proponete un intervallo di confidenza per la spesa media.

```
k <- 0.8
```

```
lim.inf <-
```

(d) Quali potrebbero essere gli elementi di debolezza della vostra scelta ?

(e) Trasformate ora i dati secondo queste tre opzioni: a)  $\log(y_i)$ , b)  $\sqrt{y_i}$ , c)  $y_i^{1/3}$ .

```
y.mod.a <- log(y)
y.mod.b <- sqrt(y)
y.mod.c <- y^(1/3)
```

(f) A vostro parere quale di queste trasformazioni rende la distribuzione dei dati trasformati più simile a quella di una v.c. normale ?

```
p <- ggplot(data.table(data=y), aes(x=data))
p <- p + geom_density(aes(log(y.mod.a)), kernel = "rectangular", colour="#CC0000")
p <- p + geom_density(aes(log(y.mod.b)), kernel = "rectangular", colour="#00CC00")
```

```
p <- p + geom_density(aes(log(y.mod.c)), kernel = "rectangular", colour="#0000CC")  
p
```

