Il centro di una distribuzione.

Quale valore di c rende minima la quantità

$$\sum_{i=1}^{n}(x-c)^2\frac{1}{n}$$

1- *Proviamo ad inserire il termine x medio, sviluppando l'equazione*:

$$\sum_{i=1}^{n}(x_i-c)^2\frac{1}{n}=\sum_{i=1}^{n}[(x_i-\bar{x})-(c-\bar{x})]^2\frac{1}{n}$$

$$=\sum_{i=1}^{n}[(x_i-\bar{x})^2+(c-\bar{x})^2-2(c-\bar{x})(x_i-\bar{x})]\ \frac{1}{n}$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})^2\frac{1}{n}+\sum_{i=1}^{n}(c-\bar{x})^2\frac{1}{n}-\ 2(c-\bar{x})\sum_{i=1}^{n}(x_i-\bar{x})\frac{1}{n}$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})^2\frac{1}{n}+\ n(c-\bar{x})^2\frac{1}{n}-\ 2(c-\bar{x})\sum_{i=1}^{n}(x_i-\bar{x})\frac{1}{n}$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})^2\frac{1}{n}+\ (c-\bar{x})^2-\ 2(c-\bar{x})\left[\left(\sum_{i=1}^{n}x_i\frac{1}{n}\right)-\bar{x}\right]$$

dato che $-2(c-\bar{x})[\bar{x}-\bar{x}]=0$, si avrà che la quantità

$$\sum_{i=1}^{n}(x_i-c)^2\frac{1}{n}\ =\sum_{i=1}^{n}(x_i-\bar{x})^2\frac{1}{n}+\ (c-\bar{x})^2$$

è minima quando $c=\bar{x}$

2- Disegniamo la funzione

$$y=\sum_{i=1}^{n}(x-\mathbf{c})^2\frac{1}{n}$$

nel campione $x=\{1,2,3,4,\ldots,50\}$, utilizzando di volta in volta uno dei cinquanta valori campionari come possibile valore per **c**.

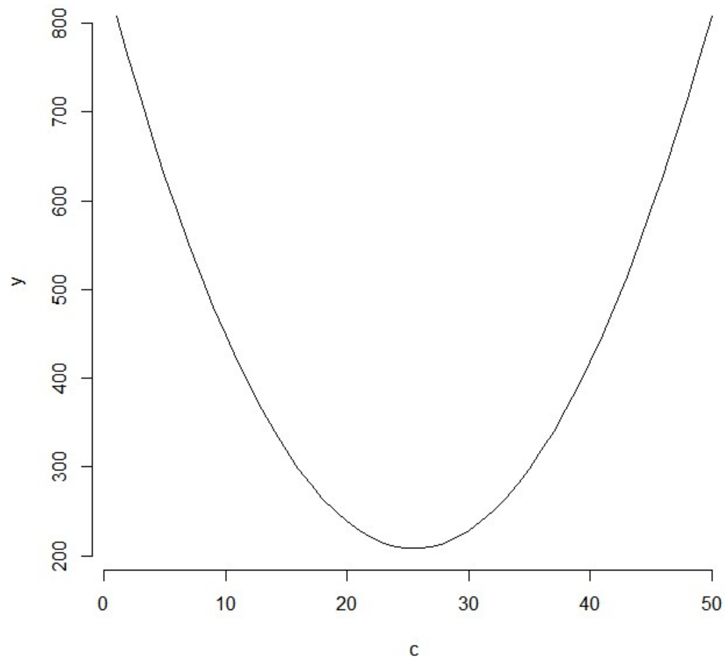
Ad esempio, per $c=1$;

$$y=\sum_{i=1}^{n}(x-\mathbf{1})^2\frac{1}{n}=808.5$$

Ad esempio, per $c=50$;

$$y=\sum_{i=1}^{n}(x-\mathbf{50})^2\frac{1}{n}=808.5$$

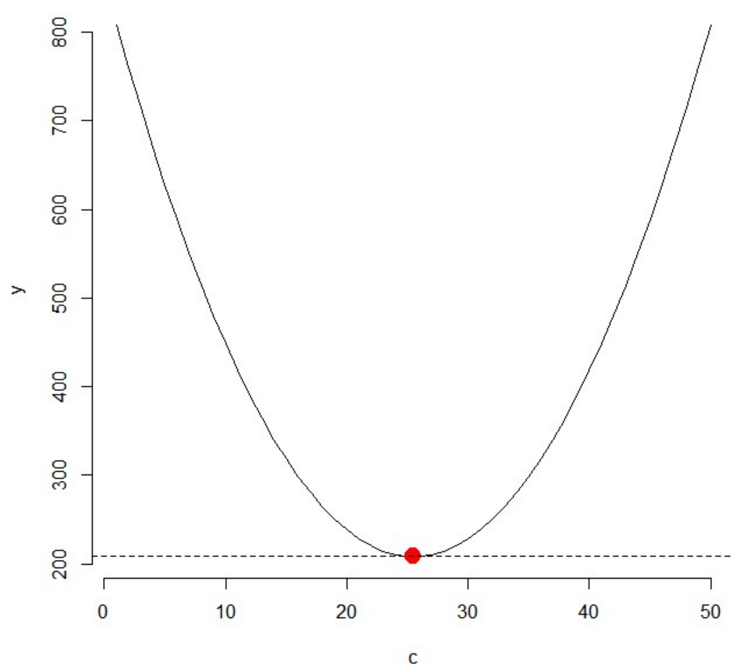
Proviamo a tracciare in un grafico tutti i valori di *y* in funzione dei possibili valori incogniti di *c.*



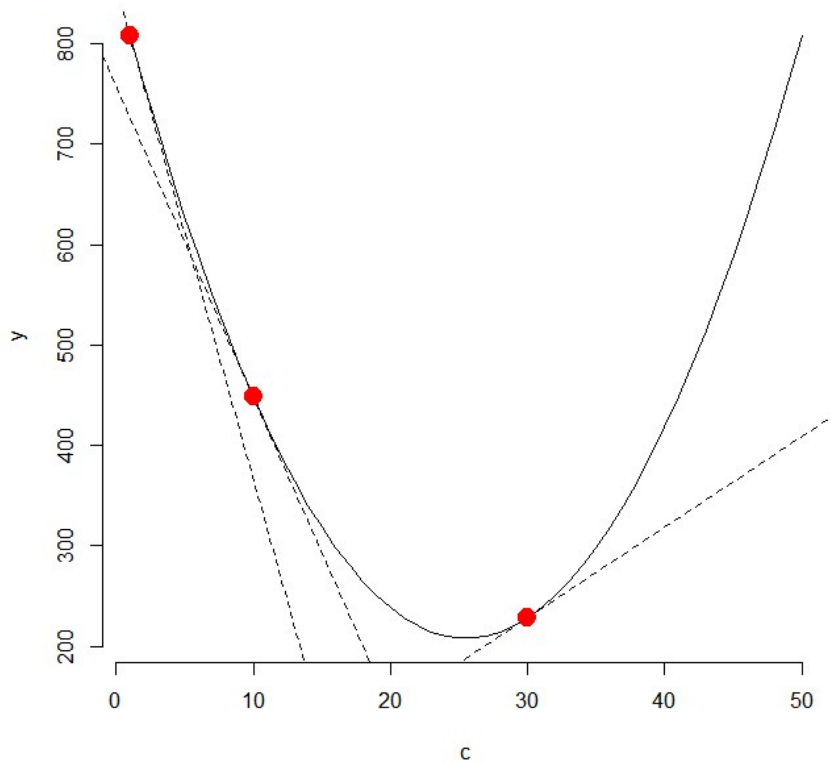
Esiste un valore di c che rende minima la funzione e si tratta della media campionaria:

$$\sum_{i=1}^{n} x_i \frac{1}{n} = 25{,}5.$$

In questo punto la *retta tangente* ha inclinazione zero,



In altri punti, diversi dal minimo valore di *y*, la *retta tangente* ha inclinazione diversa da zero.



L'inclinazione della retta tangente è (un valore dato da) una nuova *funzione,* detta *derivata di y nell'incognita c*. Per trovare questa nuova funzione y' dovremo applicare le regole di derivazione:

$$y' = \left(\sum_{i=1}^{n}(x_i - c)^2 \frac{1}{n}\right)'$$

Applichiamo la regola della somma: $(f + g + \ldots)' = (f)' + (g)'$+…

$$y' = \sum_{i=1}^{n}\left((x_i - c)^2 \frac{1}{n}\right)'$$

Applichiamo la regola della potenza: $(f^n)' = nf^{n-1}(f)'$

e della moltiplicazione per costante: $(kf)' = k(f)'$

$$y' = \sum_{i=1}^{n}\left((x_i - c)^2 \frac{\mathbf{1}}{\boldsymbol{n}}\right)'$$

$$y' = \sum_{i=1}^{n}\frac{\mathbf{1}}{\boldsymbol{n}}((x_i - c)^2)'$$

$$y' = \sum_{i=1}^{n} \frac{1}{n} (2(x_i - c)^{2-1}(x_i - c)')$$

Applichiamo le prime due regole facendo attenzione che per noi l'argomento è c, mentre la costante (non incognita) è x_i,

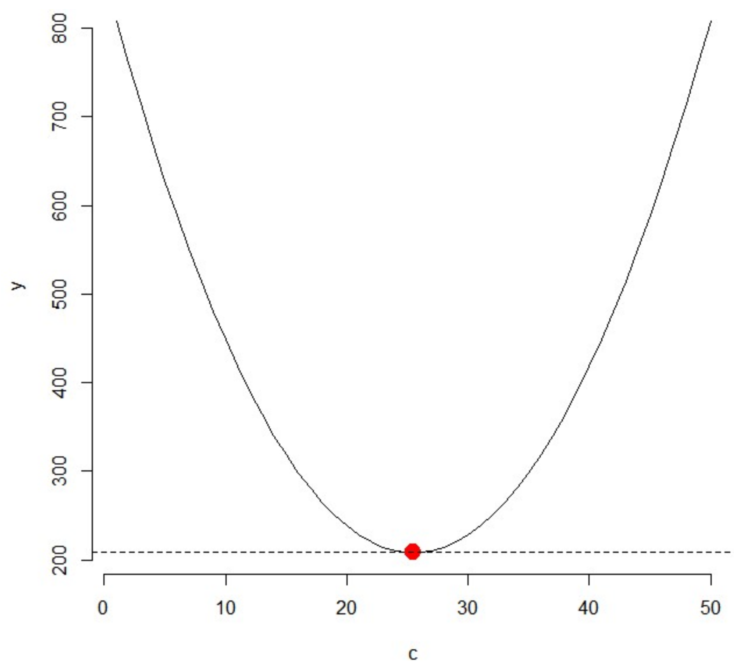
$$y' = \sum_{i=1}^{n} \frac{1}{n} (2(x_i - c)^{2-1}(0 - 1))$$

$$y' = \sum_{i=1}^{n} \frac{1}{n} (-2x_i + 2c)$$

$$y' = 2\left[-\sum_{i=1}^{n} x_i \frac{1}{n} + nc\frac{1}{n}\right]$$

$$y' = 2(-\bar{x} + c)$$

Con soluzione $y' = 0$ (retta tangente non inclinata) quando $\bar{x} = c$.



In generale, la retta ha funzione $y = a + bc$ e nel caso specifico della retta tangente al punto

$$y' = 2(-\bar{x} + c) = b$$

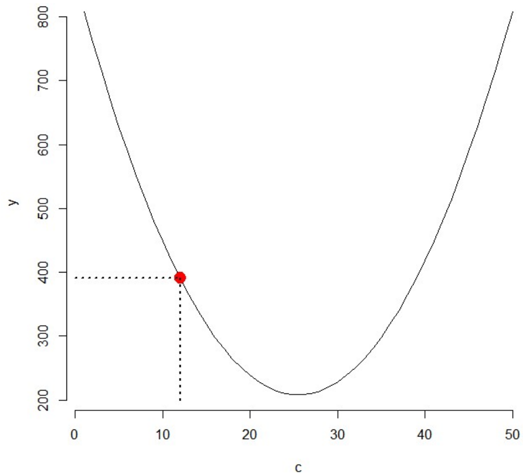
e per $\bar{x} = c$ (pendenza nulla) avremo che

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 \frac{1}{n} = a + 0\bar{x} = a$$

dove a indica il punto di passaggio della retta per l'asse y.

Per $c = 12$ avremo che

$$y = \sum_{i=1}^{n} (x_i - 12)^2 \frac{1}{n} = 390{,}5$$



ed essendo naturalmente un punto della retta tangente possiamo scrivere l'equazione

$$390{,}5 = a + b12$$

$$390{,}5 = a + 2(-\bar{x} + 12)12$$

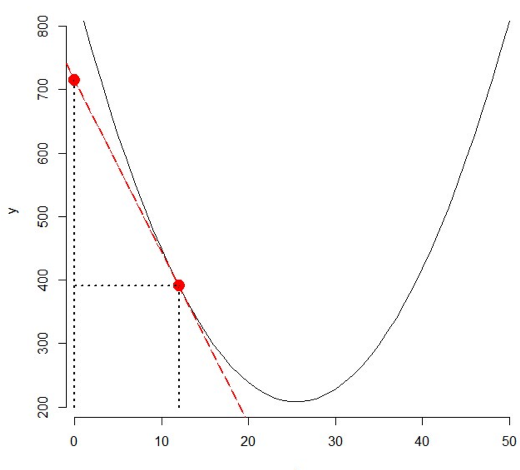
$$390{,}5 = a - 27 \times 12$$

dove $b = -27$ (retta con pendenza negativa) e dove *a* può essere ricavata semplicemente

$$390{,}5 + 27 \times 12 = 714{,}5 = a$$

La retta tangente può dunque essere tracciata come

$$y = 714{,}5 - 27c$$



Programmino in R per calcolare i valori di y per tutti i valori scelti di c, affiancando la stima della pendenza $\text{beta} = 2[-\bar{x} + c]$ e dell'intercetta $\text{alpha} = y - 2[-\bar{x} + c]c$ delle rette tangenti ad ogni coppia di valori (c, y).

```
x<- c(1:50)
Media<-mean(x)
Var<-sum((x-Media)^2)*1/50
#--- var(x)*49/50 equivalente

#
# Derivata prima della funzione Varianza
#
dx1x <-D(expression((x-c)^2*1/50),"c")
dx1x<- function(c){-2*(sum(x-c)/50)}

# Stima della (Somma dei quadrati degli scarti da c)/n
# c scelto di volta in volta come uno dei valori del campione (c[i], dove
i da 1 a 50)

y<-rep(0,50) #Somma dei quadrati degli scarti da c
alpha<-rep(0,50) #
beta<-rep(0,50) # inclinazione della retta tangente nel punto c[i]
c<-x # tutti i possibili valori di c scelti nel campione

# ciclo di stima della somma dei quadrati per diversi valori di c[i]
# stima della retta tangente punto per punto
for(i in 1 :50){
y[i]<- sum((x-c[i])^2)*1/50
beta[i]<- dx1x(c[i])
alpha[i]<- y[i] - beta[i]*c[i]

}

somma_quad<-y
plot(c,y,type="l",bty="n")
points(Media,Var,col="red",pch=19,cex=2)
abline(a=Var,b=0,lwd=1.5,lty="dashed")
# .....................................
abline(alpha[1],beta[1],lwd=1.5,lty="dashed")
points(c[1],y[1],col="red",pch=19,cex=2)
# .....................................
abline(alpha[10],beta[10],lwd=1.5,lty="dashed")
points(c[10],y[10],col="red",pch=19,cex=2)
# .....................................
abline(alpha[25],beta[25],lwd=1.5,lty="dashed")
points(c[25],y[25],col="red",pch=19,cex=2)
```