

HoloTC - Soluzioni

per il modulo SM3 della SSSMT di Locarno

Dr Giorgio Pioda

Ultima compilazione il 12 settembre, 2022

This work is licensed under a [Creative Commons “Attribution-NonCommercial-ShareAlike 4.0 International”](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



Contents

1	Attività 2 - Soluzioni	2
1.1	a) Studio preliminare	2
1.2	b) Istogramma	2
1.3	c) Tabella delle frequenze e ogiva	3
1.4	d) Kernel density e ecdf	4
1.5	e) Percentili e Q-Q plot	5
1.6	f) Domande sull'uso della curva di Gauss	5
1.7	g) Quantili vs deviazione standard	6
1.8	h) MAD come indice di dispersione\$	7

1 Attività 2 - Soluzioni

1.1 a) Studio preliminare

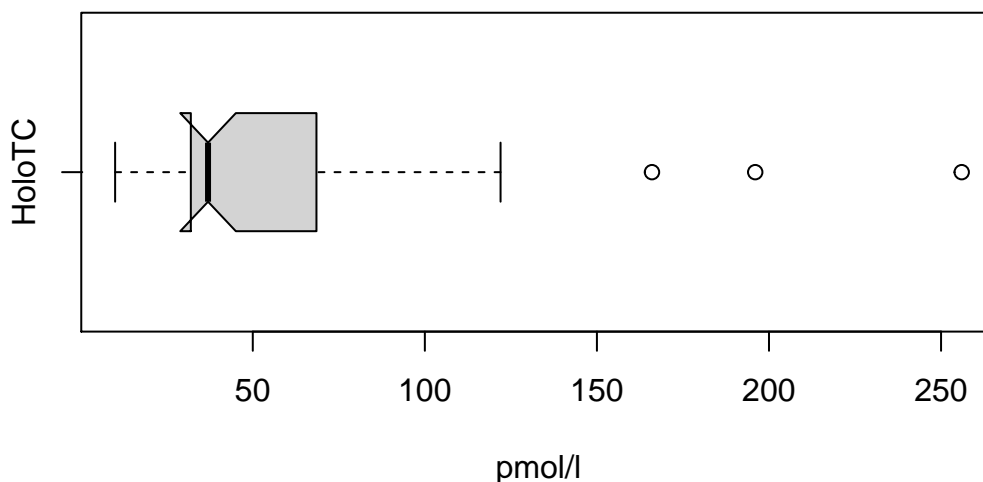
```
library("knitr")
library(kableExtra)
library("car")
holoTC <- read.table("holoTC.csv", header=T)
summary(holoTC)
```

```
##      holoTC
## Min.   : 10.0
## 1st Qu.: 32.0
## Median : 37.0
## Mean   : 57.0
## 3rd Qu.: 68.5
## Max.   :256.0
```

```
head(holoTC)
```

```
##      holoTC
## 1         34
## 2         92
## 3        115
## 4         64
## 5         10
## 6         22
```

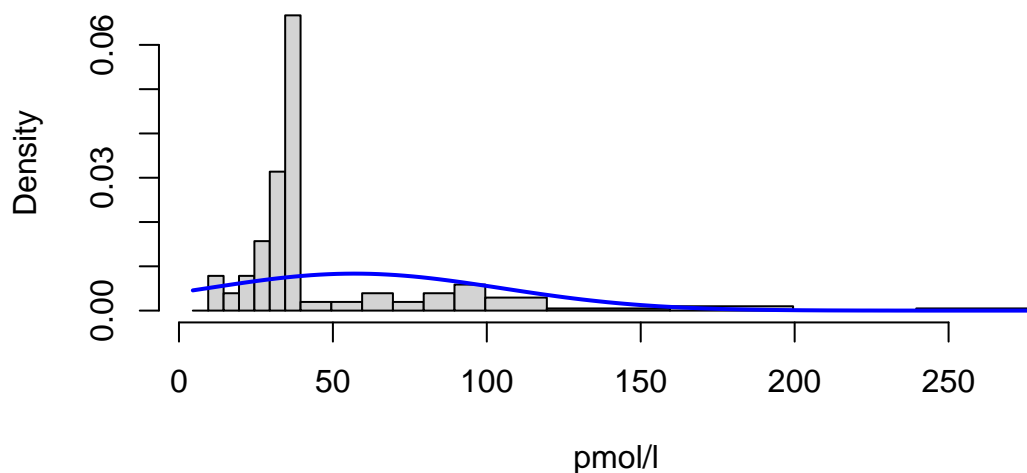
```
boxplot(holoTC$holoTC, notch=T, horizontal = T, xlab="pmol/l",
        names=c("HoloTC"), show.names=T)
```



1.2 b) Istogramma

```
holobreaks<-c(seq(4.5,40,by=5),seq(49.5,100,by=10),seq(119.5,280,by=40))
holodata<-hist(holoTC$holoTC,breaks=holobreaks,main="Holo TC",xlab="pmol/l")
holo.mean<-mean(holoTC$holoTC)
holo.sd<-sd(holoTC$holoTC)
curve(dnorm(x,mean=holo.mean,sd=holo.sd),col="blue", lwd=2,add=T,yaxt="n")
```

Holo TC



Si osservi soprattutto come la curva teorica di Gauss avrebbe una parte importante della coda di sinistra con valori negativi. Ciò chiaramente indica che i dati non sono normali. Infatti concentrazioni negative non hanno senso.

1.3 c) Tabella delle frequenze e ogiva

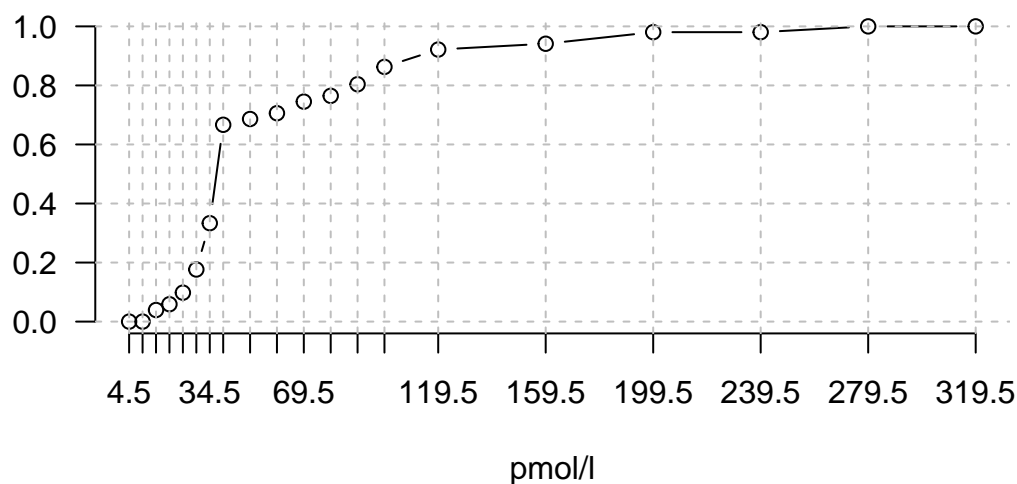
```
library("agricolae")
tabella<-table.freq(holodata)
kable(tabella,align=c('c','c','c','c','c','c'),booktabs=T)%>%
  kable_styling(position = "center",latex_options = "hold_position")
```

Lower	Upper	Main	Frequency	Percentage	CF	CPF
4.5	9.5	7.0	0	0.0	0	0.0
9.5	14.5	12.0	2	3.9	2	3.9
14.5	19.5	17.0	1	2.0	3	5.9
19.5	24.5	22.0	2	3.9	5	9.8
24.5	29.5	27.0	4	7.8	9	17.6
29.5	34.5	32.0	8	15.7	17	33.3
34.5	39.5	37.0	17	33.3	34	66.7
39.5	49.5	44.5	1	2.0	35	68.6
49.5	59.5	54.5	1	2.0	36	70.6
59.5	69.5	64.5	2	3.9	38	74.5
69.5	79.5	74.5	1	2.0	39	76.5
79.5	89.5	84.5	2	3.9	41	80.4
89.5	99.5	94.5	3	5.9	44	86.3
99.5	119.5	109.5	3	5.9	47	92.2
119.5	159.5	139.5	1	2.0	48	94.1
159.5	199.5	179.5	2	3.9	50	98.0
199.5	239.5	219.5	0	0.0	50	98.0
239.5	279.5	259.5	1	2.0	51	100.0

Per un colpo d'occhio veloce senza l'utilizzo di Rmarkdown (senza generare un .pdf) è sufficiente la funzione `View()`.

```
t.o<-ogive.freq(holodata,main="Holo TC", xlab="pmol/l")
```

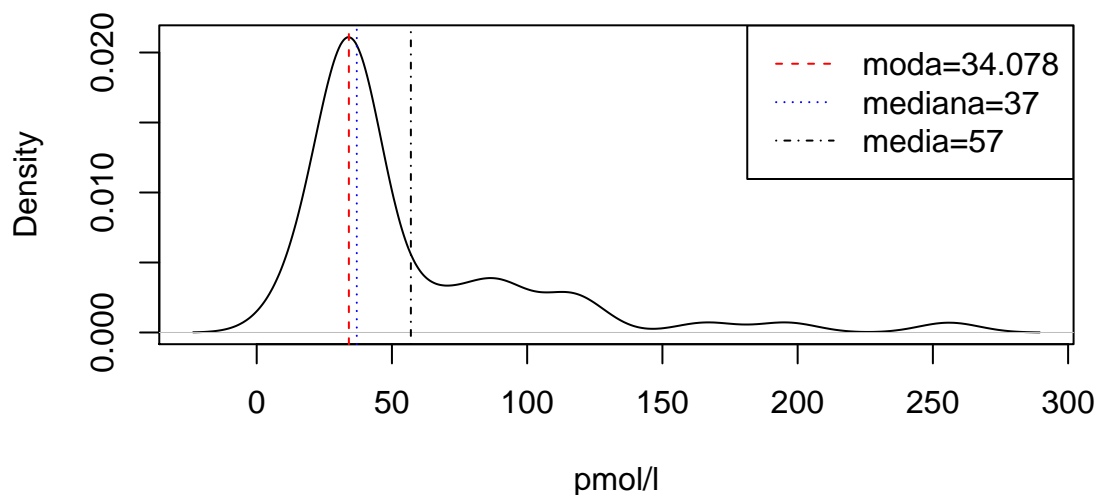
Holo TC



1.4 d) Kernel density e ecdf

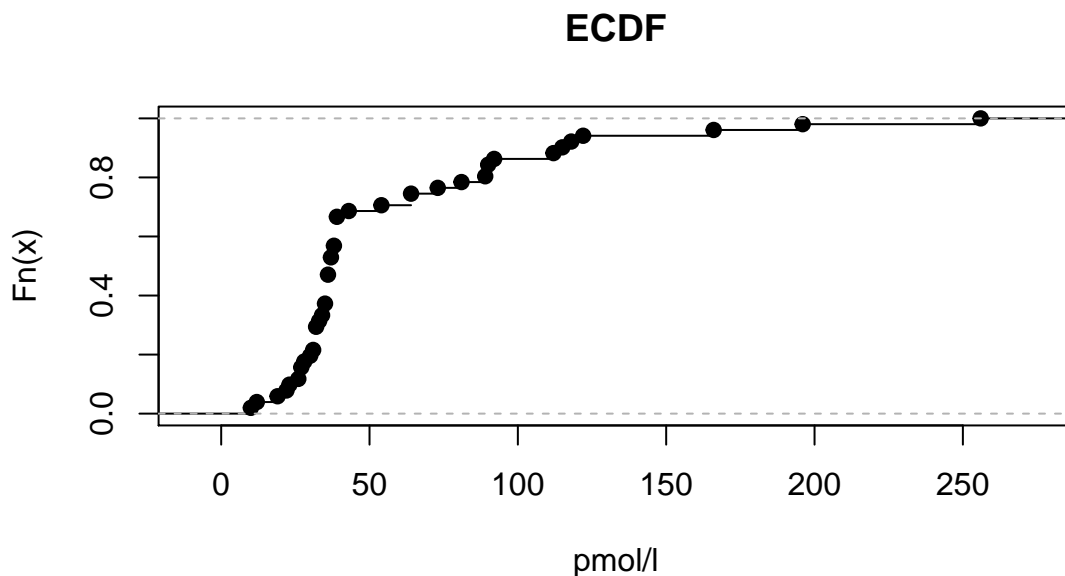
```
holo.d<-density(holoTC$holoTC)
plot(holo.d,main="Kernel density plot", xlab="pmol/l")
hol.max.i<-which.max(holo.d$y)
abline(v=holo.d$x[hol.max.i],col="red",lty=2)
abline(v=median(holoTC$holoTC),col="blue",lty=3)
abline(v=mean(holoTC$holoTC),lty=4)
legend("topright",legend=c(
  paste0("moda=",signif(holo.d$x[hol.max.i],5)),
  paste0("mediana=",signif(median(holoTC$holoTC),5)),
  paste0("media=",signif(mean(holoTC$holoTC),5))),
col=c("red","blue","black"),
lty=c(2,3,4)
)
```

Kernel density plot



Ben evidente lo scostamento della media dai valori di moda e mediana.

```
plot.ecdf(holoTC$holoTC, main="ECDF", xlab="pmol/l")
```



1.5 e) Percentili e Q-Q plot

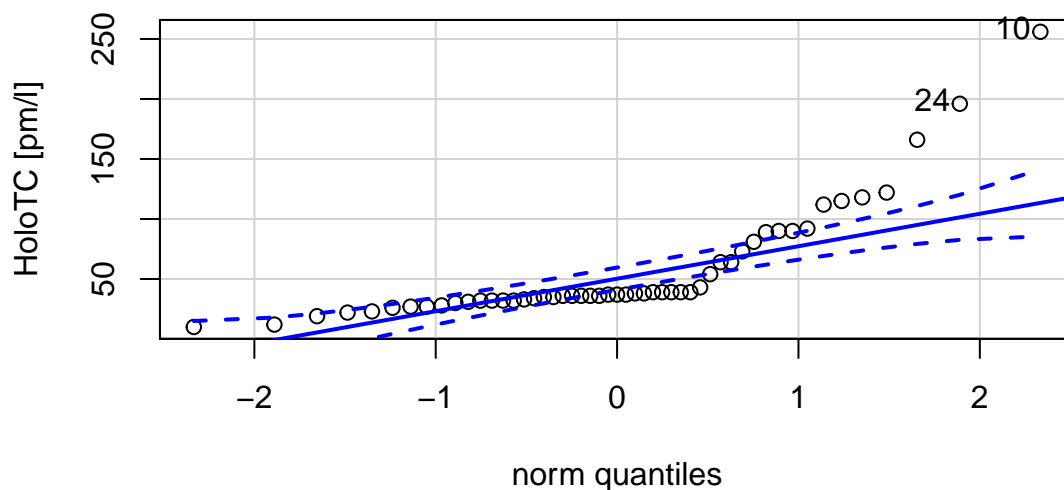
```
quant.seq<-c(seq(0,1,by=0.05))
quantile(holoTC$holoTC,quant.seq)

##      0%      5%      10%      15%      20%      25%      30%      35%      40%      45%      50%      55%      60%
## 10.0  20.5  26.0  27.5  31.0  32.0  33.0  35.0  36.0  36.0  37.0  38.0  39.0
## 65%   70%   75%   80%   85%   90%   95%  100%
## 39.0  54.0  68.5  89.0  91.0 115.0 144.0 256.0

quant.seq2<-c(0.025,0.25,0.5,0.75,0.975)
quantile(holoTC$holoTC,quant.seq2)

##      2.5%      25%      50%      75%  97.5%
## 13.75  32.00  37.00  68.50 188.50

library(car)
t.q<-qqPlot(holoTC$holoTC,distribution="norm",ylab="HoloTC [pm/l]")
```



Si nota bene un andamento non normale con una coda lunga a destra. Si osservi il QQ plot sul lato destro con i valori che divergono verso l'alto.

1.6 f) Domande sull'uso della curva di Gauss

Probabilità con holoTC di 20 e più di 110

```
(meno20piu110<-pnorm(20,mean=holo.mean,sd=holo.sd)+
(1-pnorm(110,mean=holo.mean,sd=holo.sd)))
```

```
## [1] 0.3528894
```

Valore che tiene alla destra il 15% dei dati corrisponde a P85

```
qnorm(0.85,mean=holo.mean,sd=holo.sd)
```

```
## [1] 106.5077
```

```
quantile(holoTC$holoTC,0.85)
```

```
## 85%
## 91
```

Valore che tiene alla sinistra il 5% confrontato con il P5

```
qnorm(0.05,mean=holo.mean,sd=holo.sd)
```

```
## [1] -21.5703
```

```
quantile(holoTC$holoTC,0.05)
```

```
## 5%
## 20.5
```

Qui si nota la fortissima distorsione che causa un valore insensato (negativo) sul modello di Gauss; segno che i dati non aderiscono ad una distribuzione normale.

1.7 g) Quantili vs deviazione standard

```
area<-0.6826
resto<-(1-area)
code<-resto/2
q.inf<-quantile(holoTC$holoTC,code)
q.sup<-quantile(holoTC$holoTC,(1-code))
sd.surrogato<-(q.sup-q.inf)/2
sd.surrogato
```

```
## 84.13%
## 31.0975
```

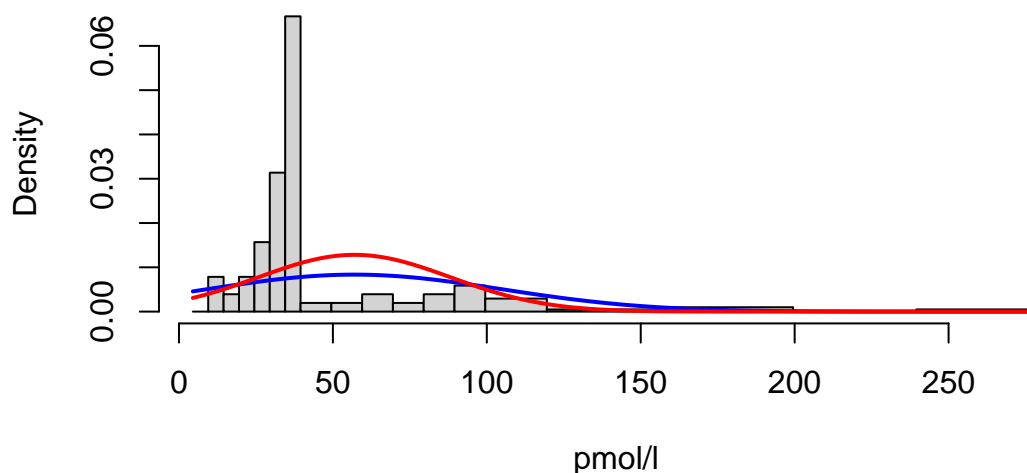
```
sd(holoTC$holoTC)
```

```
## [1] 47.76735
```

Si noti come il risultato ottenuto è parecchio più basso in quanto i valori estremi (soprattutto quelli più alti) con i percentili vengono ignorati, mentre gli stessi hanno un peso importante nella determinazione dello scarto quadratico. Usando questo valore come deviazione standard il grafico del primo punto assume il seguente aspetto (in rosso la curva di Gauss con la deviazione standard surrogata)

```
holobreaks<-c(seq(4.5,40,by=5),seq(49.5,100,by=10),seq(119.5,280,by=40))
holodata<-hist(holoTC$holoTC,breaks=holobreaks,main="Holo TC",xlab="pmol/l")
holo.mean<-mean(holoTC$holoTC)
holo.sd<-sd(holoTC$holoTC)
curve(dnorm(x,mean=holo.mean,sd=holo.sd),col="blue",lwd=2,add=T,yaxt="n")
curve(dnorm(x,mean=holo.mean,sd=sd.surrogato),col="red",lwd=2,add=T,yaxt="n")
```

Holo TC



Questo procedimento NON è nulla di più che un esperimento mentale per far riflettere sull'importanza dell'analisi del tipo di distribuzione a cui i dati appartengono.

1.8 h) MAD come indice di dispersione

```
holo.mad<-mad(holoTC$holoTC)
holo.median<-median(holoTC$holoTC)
holodata<-hist(holoTC$holoTC,breaks=holobreaks,main="Holo TC",xlab="pmol/l")
curve(dnorm(x,mean=holo.mean,sd=holo.sd,col="blue", lwd=2,add=T,yaxt="n")
curve(dnorm(x,mean=holo.median,sd=sd.surrogato,col="red", lwd=2,add=T,yaxt="n")
curve(dnorm(x,mean=holo.median,sd=holo.mad),col="purple", lwd=2,add=T,yaxt="n")
```

