

# Demo: trimmed means

*Guillaume A. Rousselet*

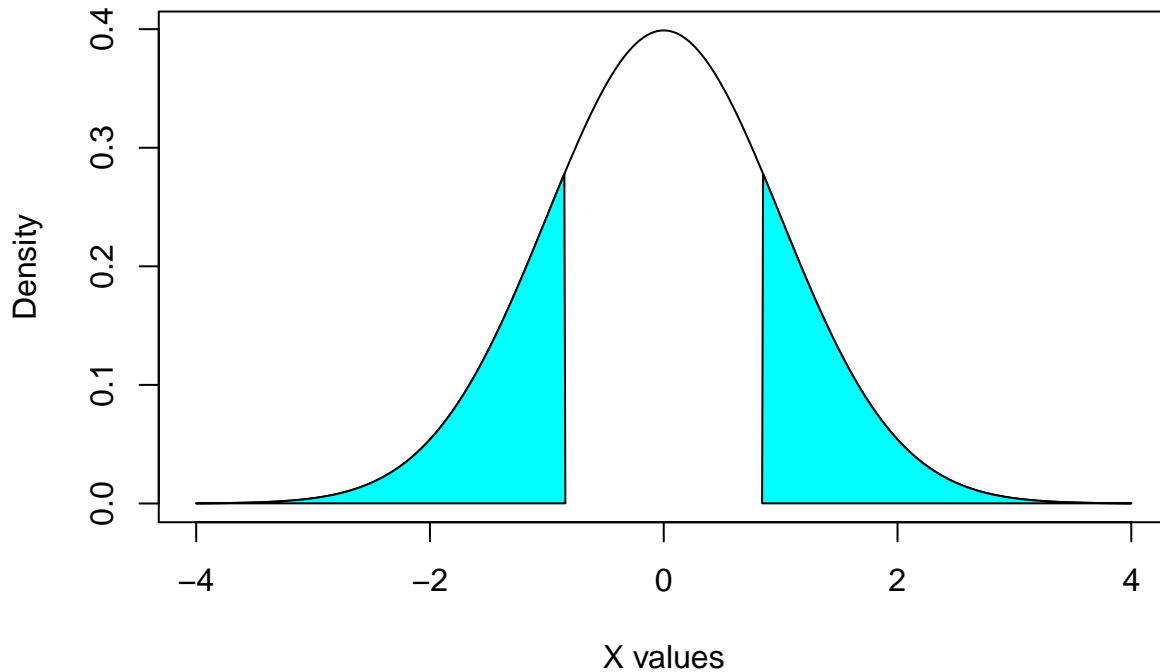
Built with R version 3.4.1

```
# dependencies
library(ggplot2)
library(tibble)
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.4.2
```

```
library(cowplot)
source("akerd.txt")
```

```
# illustration 20% trimming - Normal distribution
# jpeg('normdist.jpg')
tr<-.2
xv<-seq(-4,4,0.01)
yv<-dnorm(xv)
plot(xv,yv,type="l", xlab="X values", ylab="Density")
zval<-qnorm(tr, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
polygon(c(xv[xv<=zval],zval),c(yv[xv<=zval],yv[xv==zval]),col=5)
polygon(c(xv[xv>=zval],-zval),c(yv[xv>=zval],yv[xv==zval]),col=5)
```



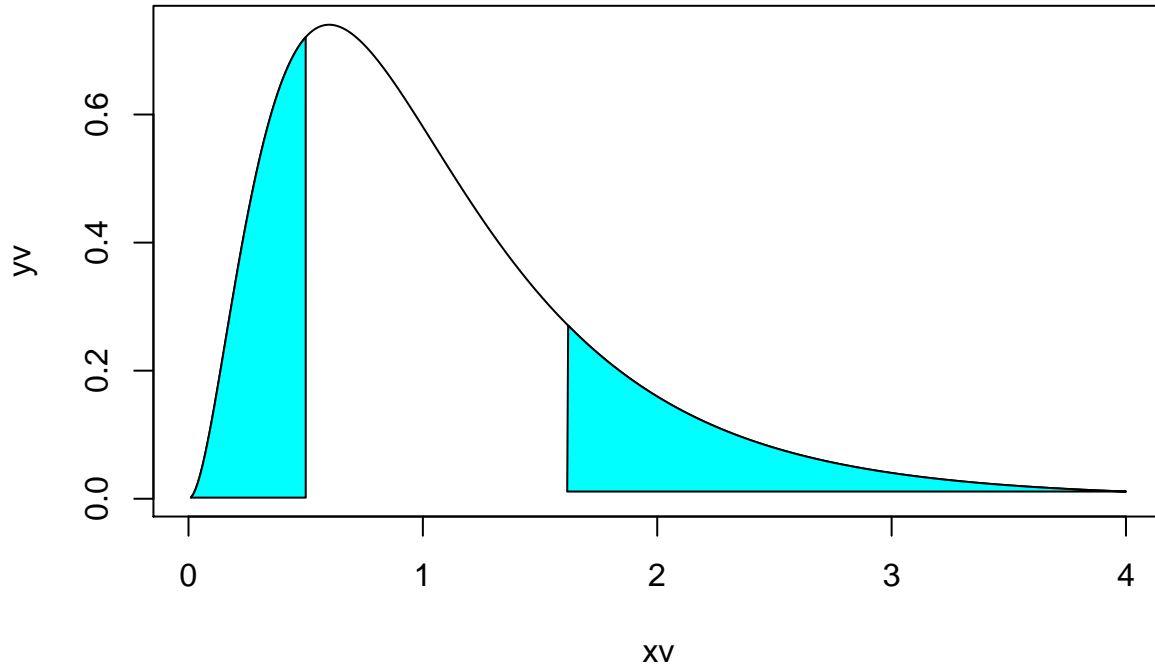
```
# dev.off()
```

```
# illustration 20% trimming - F distribution
# jpeg('fdist.jpg')
tr<-.2
xv<-seq(0.01,4,0.01)
yv<-df(xv,6,18) #fx<-dlnorm(x)
```

```

plot(xv,yv,type="l")
zval<-qf(tr,6,18)
polygon(c(xv[xv<=zval],zval),c(yv[xv<=zval],yv[xv==0.01]),col=5)
zval<-qf(1-tr,6,18)
polygon(c(xv[xv>=zval],zval),c(yv[xv>=zval],yv[xv==4]),col=5)

```



```

# dev.off()

# generate data
set.seed(4) # change the seed to generate different random samples
n <- 100 # sample size
sample.norm <- rnorm(n) # sample from normal distribution
sample.lnorm <- rlnorm(n) # sample from lognormal distribution

# save measures of central tendency
n.m <- mean(sample.norm) # mean
n.tm <- mean(sample.norm, tr = 0.2) # 20% trimmed mean
n.md <- median(sample.norm) # median
ln.m <- mean(sample.lnorm) # mean
ln.tm <- mean(sample.lnorm, tr = 0.2) # 20% trimmed mean
ln.md <- median(sample.lnorm) # median

# compute kernel density estimates
# KDE are like smooth histograms
kde.norm <- akerd(sample.norm, pyhat = TRUE, plotit = FALSE)
kde.lnorm <- akerd(sample.lnorm, pyhat = TRUE, plotit = FALSE)

# make figure ----

# save data.frames for plot
df.n <- tibble(xval=c(n.m,n.tm,n.md), MCT=c('Mean','Trimmed mean','Median'))
df.ln <- tibble(xval=c(ln.m,ln.tm,ln.md), MCT=c('Mean','Trimmed mean','Median'))
# Keep order of MCT names

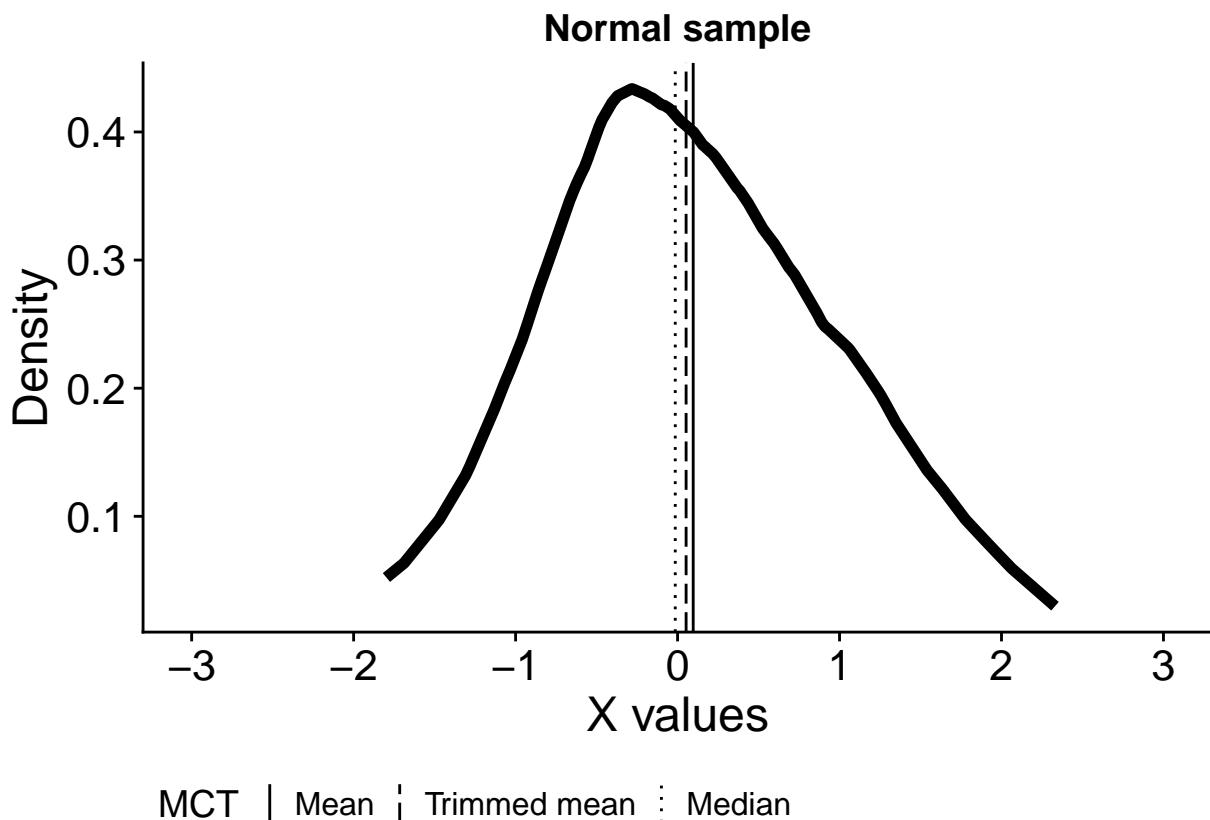
```

```

df.n$MCT <- as.character(df.n$MCT)
df.n$MCT <- factor(df.n$MCT, levels=unique(df.n$MCT))
df.ln$MCT <- as.character(df.ln$MCT)
df.ln$MCT <- factor(df.ln$MCT, levels=unique(df.ln$MCT))

# Panel A: normal distribution -----
# make data frame
df <- tibble(`x` = sort(sample.norm), `y` = kde.norm)
# ggplot
pA <- ggplot(df, aes(x,y)) + ggtitle("Normal sample") +
  geom_line(size = 2) +
  scale_x_continuous(limits=c(-3,3), breaks=seq(-3,3,1)) +
  # scale_y_continuous(limits=c(-1,8), breaks=seq(-1,8,1)) +
  theme(axis.title.x = element_text(size = 18),
        axis.text = element_text(size = 16),
        axis.title.y = element_text(size = 18),
        plot.title = element_text(face="bold"),
        legend.position = "bottom") +
  labs(x = "X values", y = "Density") +
  # geom_vline(xintercept = n.m, colour = 'black', linetype = "solid") +
  # geom_vline(xintercept = n.tm, colour = 'black', linetype = "longdash") +
  # geom_vline(xintercept = n.md, colour = 'black', linetype = "dotted")
  geom_vline(aes(xintercept=xval, linetype=MCT), data=df.n, show.legend=TRUE) +
  scale_linetype_manual(values=c("solid", "longdash", "dotted"))
pA

```

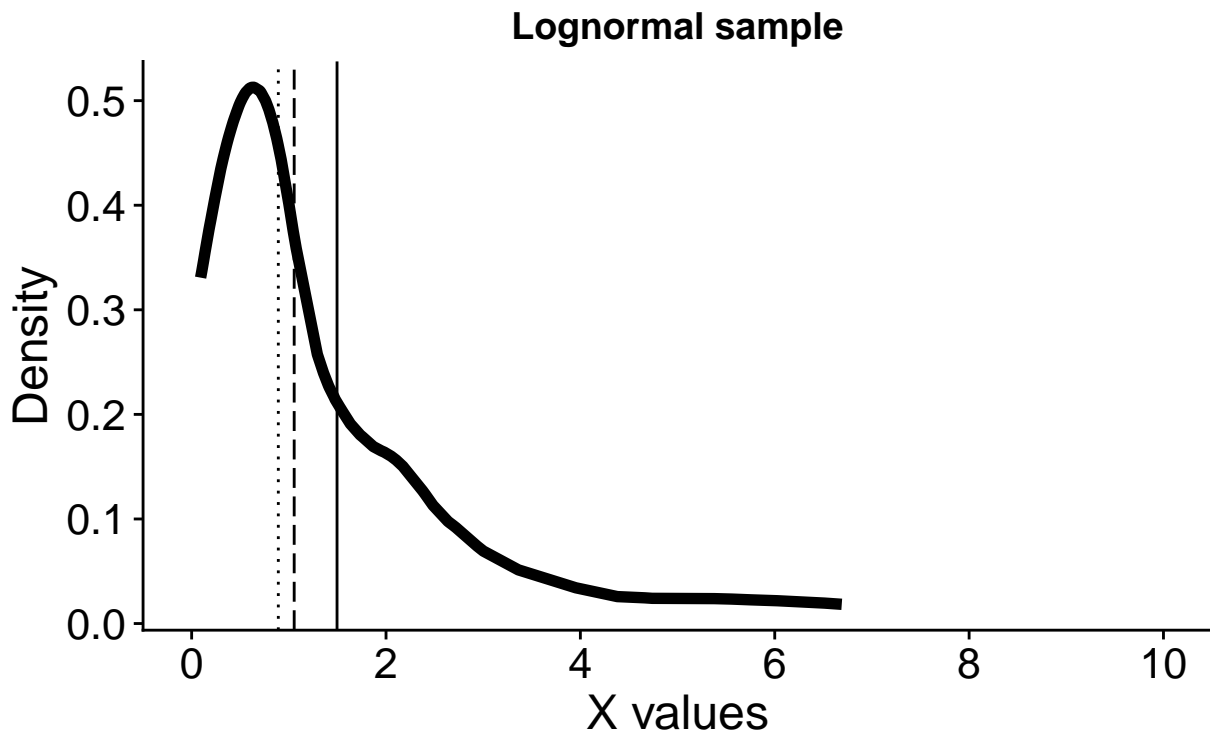


```

# Panel B: lognormal distribution -----
# make data frame

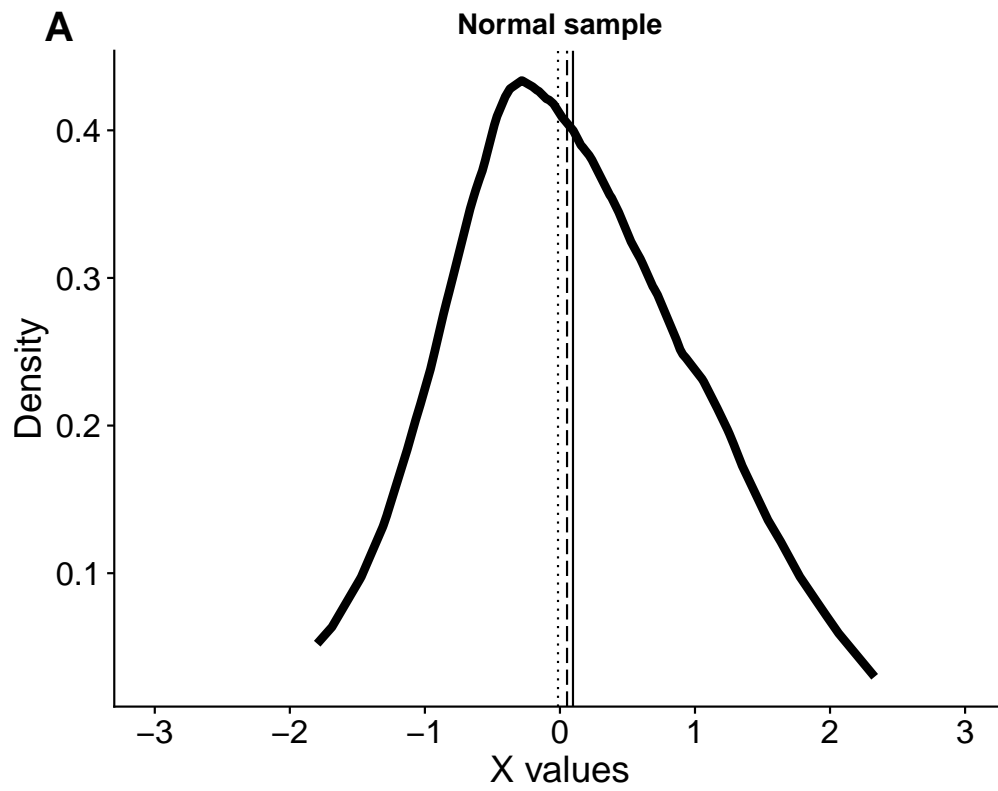
```

```
df <- tibble(`x` = sort(sample.lnorm), `y` = kde.lnorm)
# ggplot
pB <- ggplot(df, aes(x,y)) + ggtitle("Lognormal sample") +
  geom_line(size = 2) +
  scale_x_continuous(limits=c(0,10), breaks=seq(0,10,2)) +
  theme(axis.title.x = element_text(size = 18),
        axis.text = element_text(size = 16),
        axis.title.y = element_text(size = 18),
        legend.position = "bottom") +
  labs(x = "X values", y = "Density") +
  # geom_vline(xintercept = ln.m, colour = 'black', linetype = "solid") +
  # geom_vline(xintercept = ln.tm, colour = 'black', linetype = "longdash") +
  # geom_vline(xintercept = ln.md, colour = 'black', linetype = "dotted")
  geom_vline(aes(xintercept=xval, linetype=MCT), data=df.ln, show.legend=TRUE) +
  scale_linetype_manual(values=c("solid", "longdash", "dotted"))
pB
```

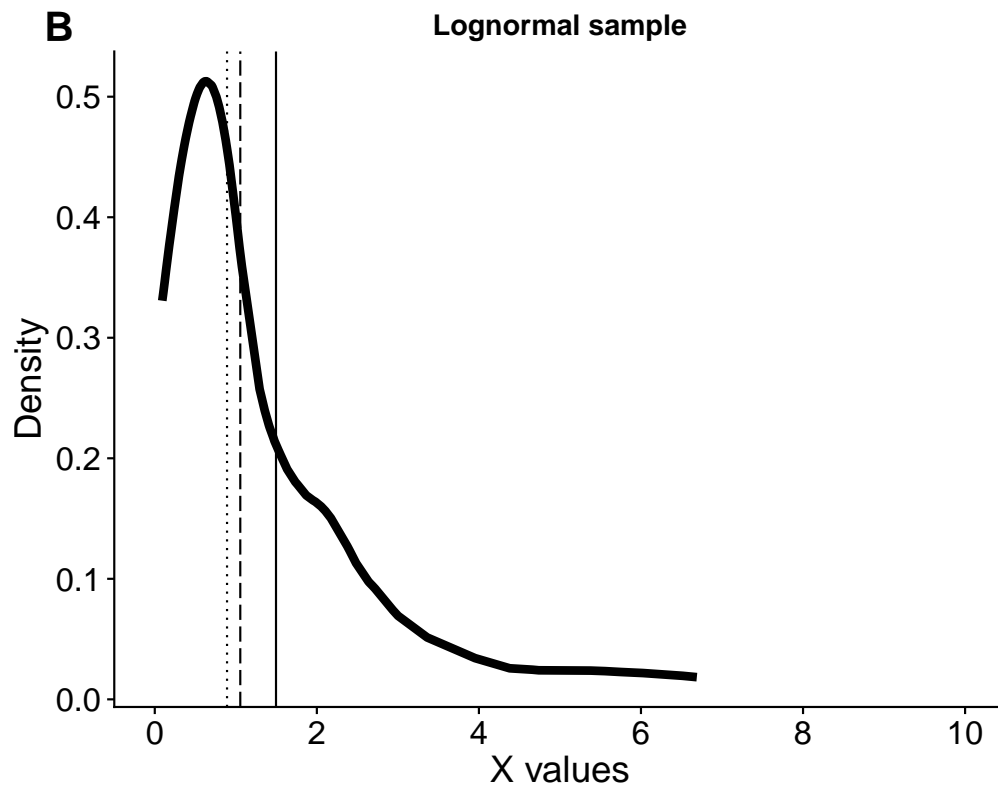


MCT | Mean | Trimmed mean | Median

```
# combine panels into one figure
cowplot::plot_grid(pA, pB,
  labels=c("A", "B"),
  ncol = 1,
  nrow = 2,
  rel_widths = c(1, 1),
  label_size = 20,
  hjust = -1.8,
  scale=1)
```



MCT | Mean | Trimmed mean | Median



MCT | Mean | Trimmed mean | Median

```
# save figure  
ggsave(filename='figure_tm_demo.jpg',width=7,height=12) #path=pathname
```