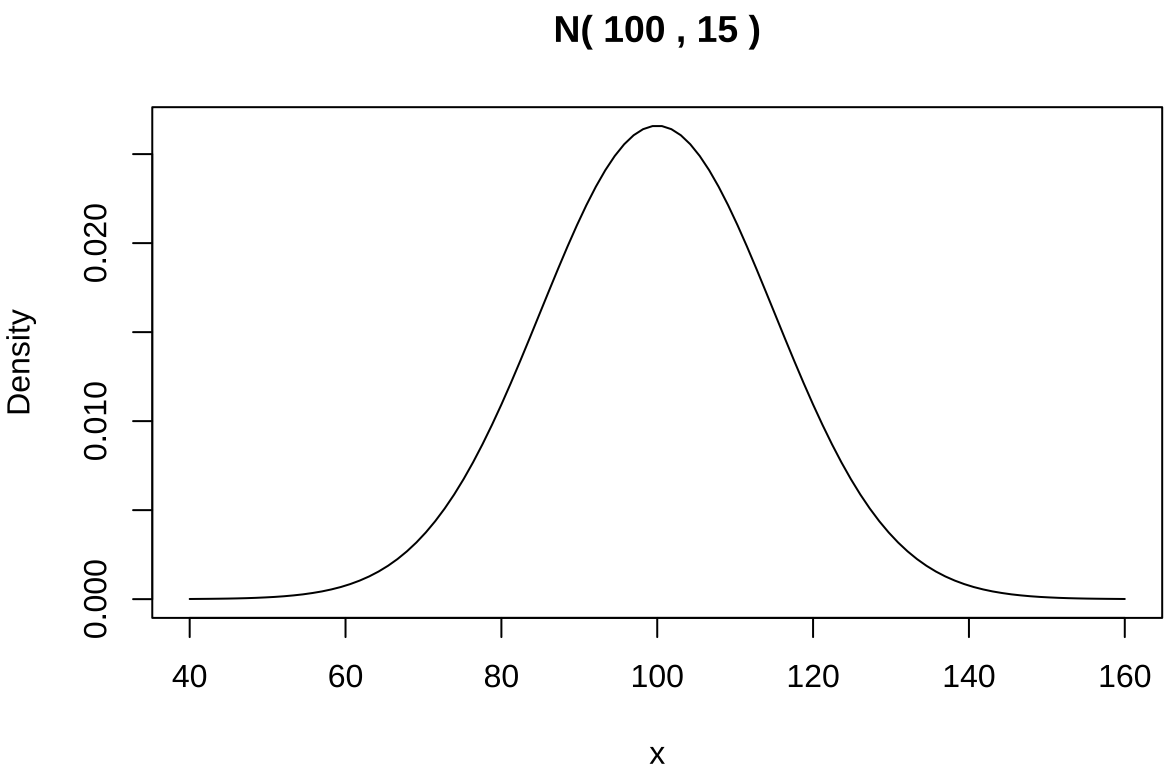
The Normal Distribution

Prof. Kapitula

9/23/2021

Write a Function to Plot a Normal Distribution

```
plotnorm<- function(mean=0,sd=1){
x <- seq(-4,4,length=100)*sd + mean
hx <- dnorm(x,mean,sd)
plot(x, hx, type="l", xlab="x", ylab="Density",main=paste("N(",mean,",",sd,")"))
}
plotnorm(mean=100,sd=15)
```

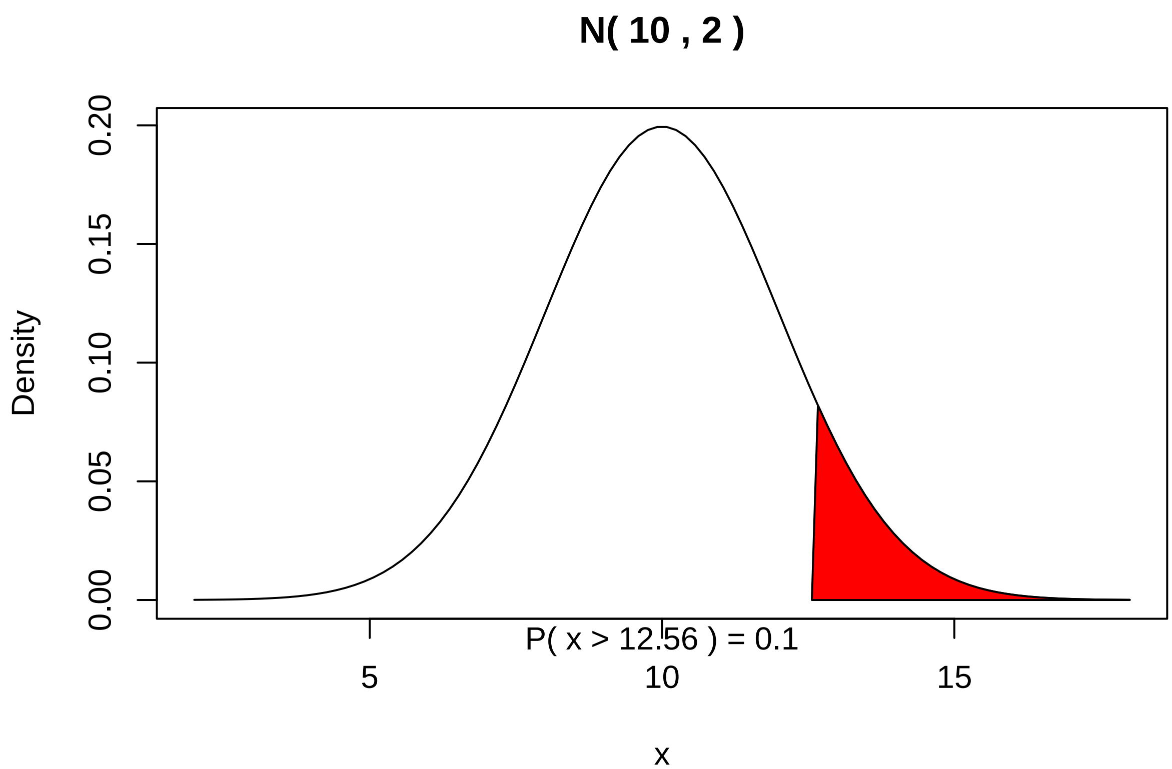



```
plotnormprob<- function(mean=0,sd=1, tail='lower', p=.5){
  x <- seq(-4,4,length=100)*sd + mean
  hx <- dnorm(x,mean,sd)
if (tail=='lower'){
lb=-4*sd+mean
ub=qnorm(p,mean,sd)
i <- x <= ub
plot(x, hx, type="l", xlab="x", ylab="Density", main=paste("N(",mean,",",sd,")"))
polygon(c(lb,x[i],ub), c(0,hx[i],0), col="red")
```

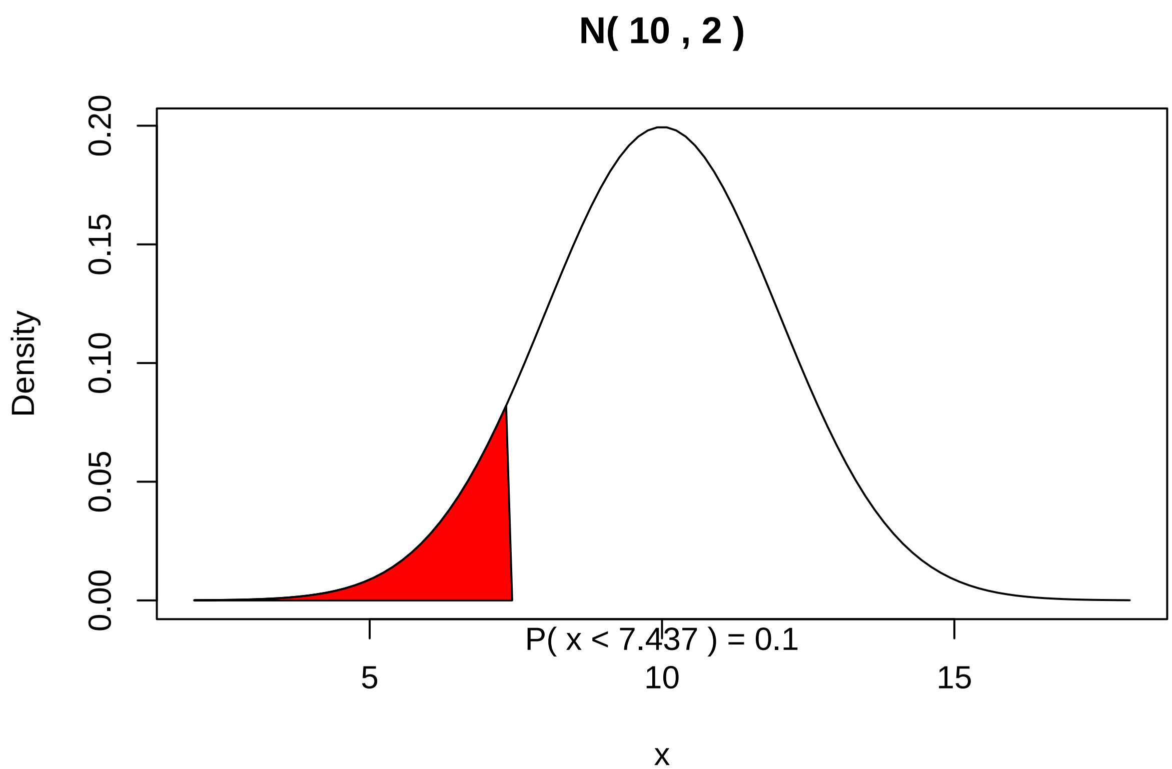
```r
#area <- pnorm(qnorm(p,mean,sd), mean, sd)
result <- paste("P( x <",signif(ub,digits=4),") =",signif(p, digits=4))
mtext(result,1)}
  else if (tail=='upper'){
p=1-p
ub=4*sd+mean
lb=qnorm(p,mean,sd)
i <- x >= lb
plot(x, hx, type="l", xlab="x", ylab="Density", main=paste("N(",mean,",",sd,")"))
polygon(c(lb,x[i],ub), c(0,hx[i],0), col="red")
result <- paste("P( x >",signif(lb,digits=4),") =",signif(1-p, digits=4))
mtext(result,1)}



  }
plotnormprob(tail='upper',mean=10,sd=2,p=.1)
```

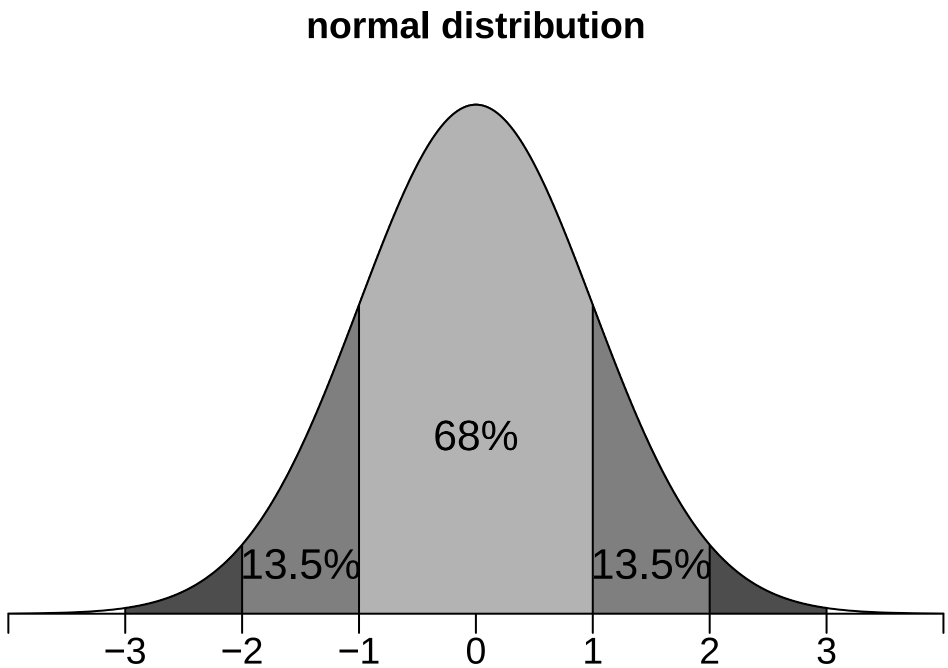



```r
plotnormprob(tail='lower',mean=10,sd=2,p=.1)
```



The Normal Distribution follows the empirical rule.

```
curve(dnorm(x), -4, 4, ylim=c(0, 0.4), xlab="", ylab="", bty="n",
      yaxs="i", main="normal distribution", xaxt="n", yaxt="n")
axis(1, c(-4, -3, -2, -1,  0,  1, 2, 3, 4),
     c("", "-3", "-2", "-1",  "0",  "1", "2", "3", ""),
     mgp=c(1.5, .5, 0), cex.axis=1.2)
colors <- c("gray70", "gray50", "gray30")
for (i in 3:1){
  grid <- seq(-i, i, .01)
  polygon(c(grid, i, -i), c(dnorm(grid), 0, 0),
          col=colors[i])
}
text(0, .35*dnorm(0), "68%", cex=1.3)
text(-1.5, .3*dnorm(1.5), "13.5%", cex=1.3)
text(1.5, .3*dnorm(1.5), "13.5%", cex=1.3)
```



In general random variables follow Chebychev's Inequality,

Let X (integrable) be a random variable with finite expected value μ and finite non-zero variance σ^2. Then for any real number $k > 0$,

$$Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Only the case $k > 1$ is useful. Why?

The Central Limit Theorem

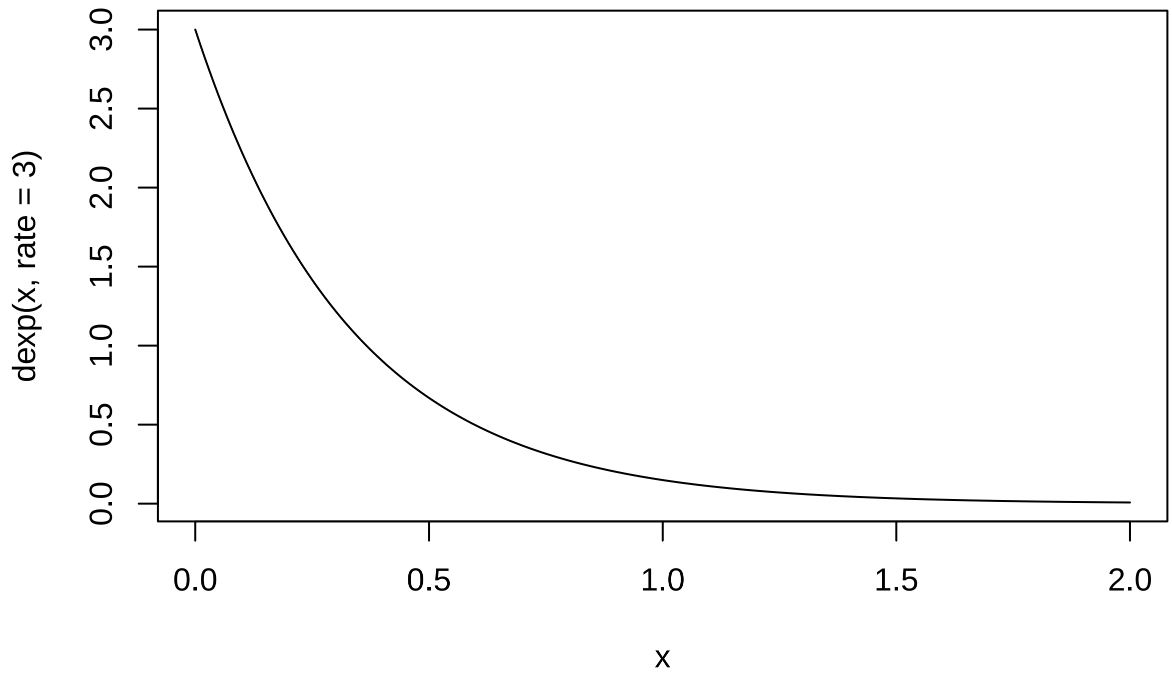
For $X_1, X_2, ..., X_n$ iid random variables. If $X_i \sim N(\mu, \sigma^2)$ or if $nislarge$ the sampling distribution of $\bar{X}$ has a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, ie.

$$\bar{X} \dot{\sim} N(\mu, \frac{\sigma^2}{n})$$

For example,

```
set.seed=522
nsims=1000
n=100
```

```
x=seq(0,2,.01)
plot(x,dexp(x,rate=3), type="l")
```



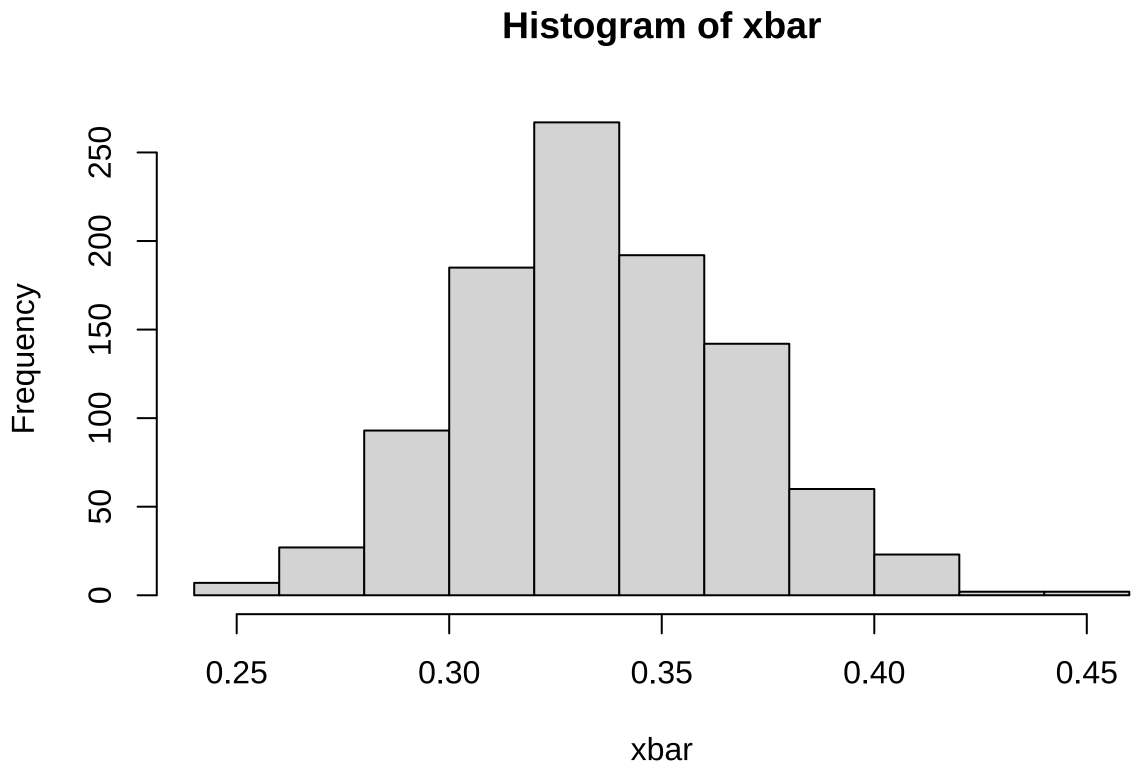

```
samples=matrix(rexp(rate=3,n=nsims*n),nrow=nsims, ncol=n)
xbar=rowMeans(samples)
mean(xbar)
```

```
## [1] 0.3357937
```

```
sd(xbar)
```

```
## [1] 0.03196083
```

```
hist(xbar)
```



Exercise, What will this look like if n is small? Make a few plots to illustrate how changing n changes the sampling distribution of the sample mean.