February 17th estimators

Population

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

fixed-size samples

Population

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

fixed-size samples

$$= \bar{X}$$

$$= \bar{X}$$

$$= \bar{X}$$

$$= \bar{X}$$

$$\vdots \qquad \vdots$$

Population

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

fixed-size samples

$= \bar{X}$

$= \bar{X}$

$= \bar{X}$

$= \bar{X}$

$\vdots$

the sampling distribution of the sample mean

Expected value of $\bar{X}$ in terms of the population

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

$$\begin{aligned} E[\bar{X}] &= E[\frac{1}{n}(X_1 + \cdots + X_n)] \\ &= \frac{1}{n}\sum_{1}^{n} E[X_i] \\ &= \frac{1}{n}\sum_{1}^{n} \mu \\ &= \frac{1}{n}(n\mu) \\ &= \mu \end{aligned}$$

Expected value of $\bar{X}$ in terms of the population

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

$$\begin{aligned}
E[\bar{X}] &= E[\frac{1}{n}(X_1 + \cdots + X_n)] \\
&= \frac{1}{n}\sum_{1}^{n} E[X_i] \\
&= \frac{1}{n}\sum_{1}^{n} \mu \\
&= \frac{1}{n}(n\mu) \\
&= \mu
\end{aligned}$$

unbiased estimator of population mean

Variance of the sample mean

$$\begin{aligned}
Var[\bar{X}] &= Var\left[\frac{1}{n}(X_1 + \ldots X_n)\right] \\
&= \frac{1}{n^2}\left(Var\left[X_1\right] + \cdots + Var\left[X_n\right]\right) \\
&= \frac{1}{n^2}\left(\sigma^2 + \cdots + \sigma^2\right) \\
&= \frac{1}{n^2}\left(n \times \sigma^2\right) \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

Un-squared: the s.d.s.m

$$Var\left[\bar{X}\right] = \frac{\sigma^2}{n}$$

$$\sqrt{Var\left[\bar{X}\right]} = \sqrt{\frac{\sigma^2}{n}}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Quantify uncertainty in estimating population mean from sample mean

$$\text{Statistic} = \text{Parameter} + error$$

or

$$\text{Parameter} = \text{Statistic} + error$$

Quantify uncertainty in estimating population mean from sample mean

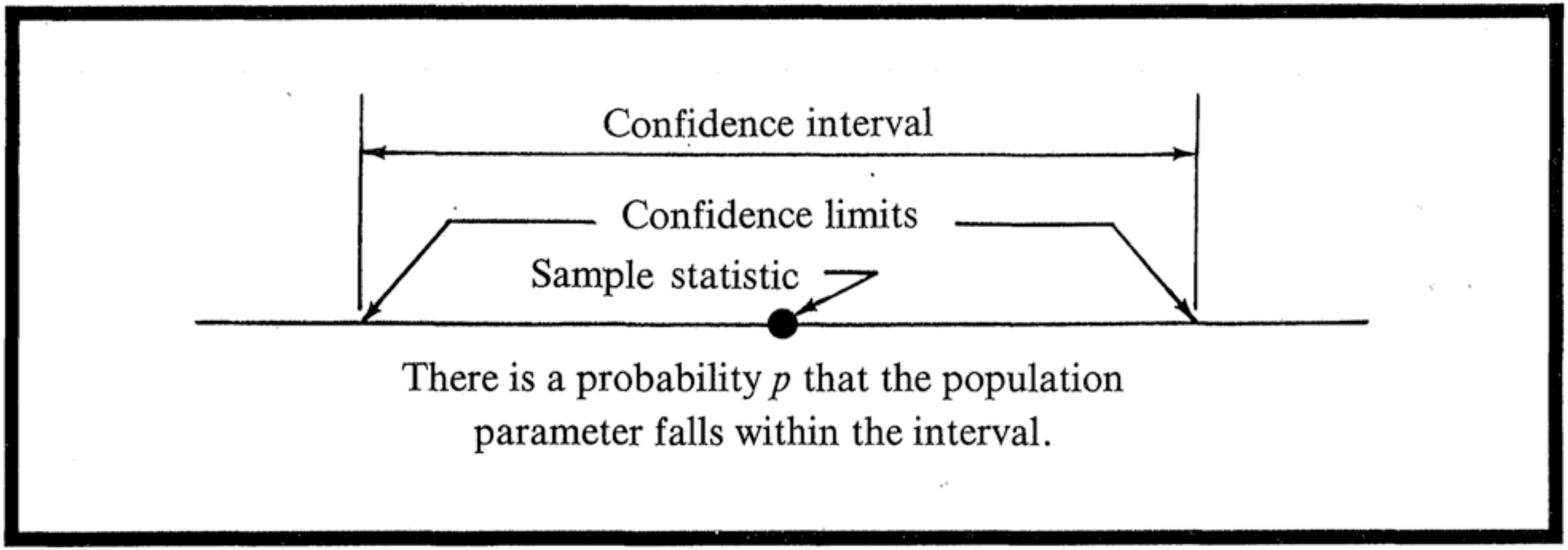


Figure 17 Schematic diagram of an interval estimate of a parameter.

Quantify uncertainty in estimating population mean from sample mean

$$P\left(\mu \,\text{is in the interval}\, \bar{X} \pm 2 \times \sigma_{\bar{X}}\right) > 0.95$$

Vasishth 3.3

Quantify uncertainty in estimating population mean from sample mean

$$P\left(\mu \text{ is in the interval } \bar{X} \pm 2 \times \sigma_{\bar{X}}\right) > 0.95$$

appropriate s.d.s.m for our sample

Vasishth 3.3

Probability that $\bar{X}$ misses μ due to sampling error that is Z standard deviations big

$$P\left(\mu \,\text{is in}\, \bar{X} \pm Z\sqrt{\frac{\sigma^2}{n}}\right) = ?$$

Claiming that μ is in $\bar{X} \pm Z\sqrt{\frac{\sigma^2}{n}}$ entails

$$\bar{X} - Z\sqrt{\frac{\sigma^2}{n}} \quad < \mu < \quad \bar{X} + Z\sqrt{\frac{\sigma^2}{n}}$$

$$\bar{X} < \mu + Z\frac{\sigma}{\sqrt{n}} \qquad \vdots$$

$$\bar{X} - \mu < Z\frac{\sigma}{\sqrt{n}} \qquad \vdots$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < Z \quad \text{and} \quad \frac{\mu - \bar{X}}{\sigma/\sqrt{n}} < Z$$

The quantity $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

Theorem 5-5: Suppose that the population from which samples are taken has a probability distribution with mean μ and variance σ^2 that is not necessarily a normal distribution. Then the standardized variable associated with $\bar{X}$, given by

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \tag{7}$$

is *asymptotically normal*, i.e.,

$$\lim_{n\to\infty} P(Z \leqq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du \tag{8}$$

follows from CLT

Lacking σ, we cannot compute $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$.

We can compute $\frac{\bar{X}-\mu}{s/\sqrt{n}}$
using s, the sample standard deviation

Getting by without sigma

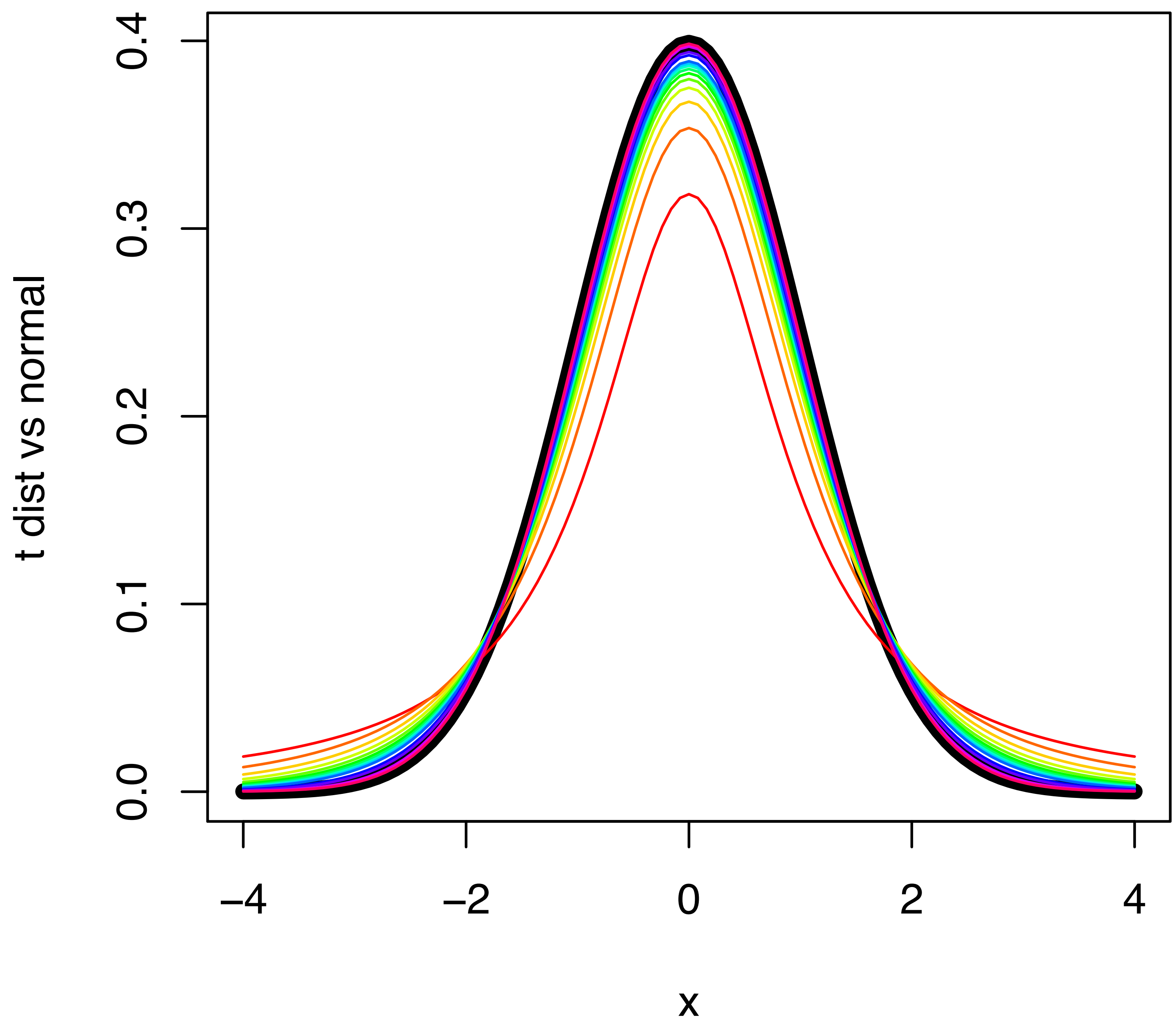
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$\text{estimate of } \sigma_{\bar{X}} = \frac{s}{\sqrt{n}}$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Standard error of the mean

$$SE_{\bar{X}} = \frac{s}{\sqrt{n}}$$



n-1 df

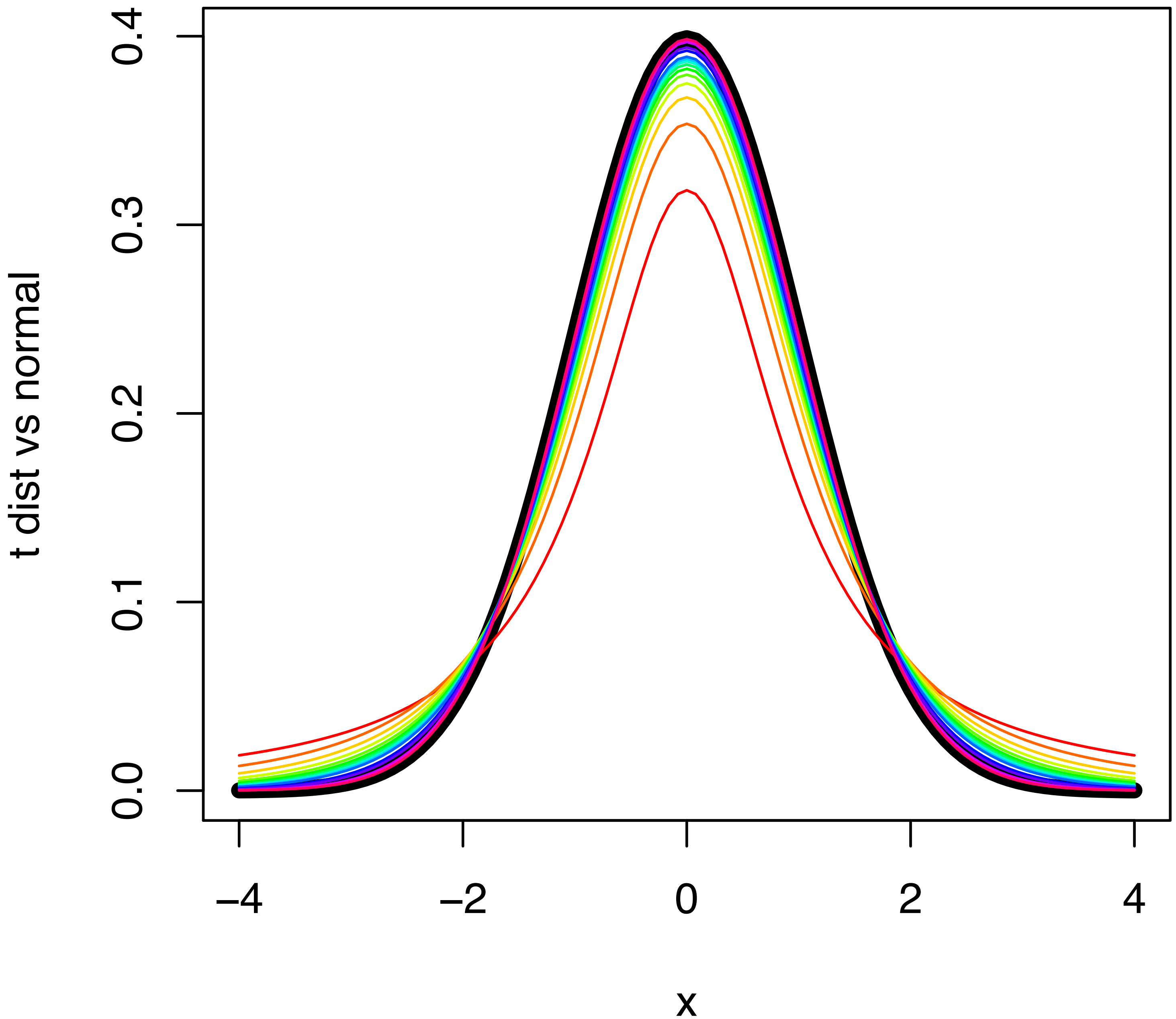
```
degsfreedom <- c(1,2,3,4,5,6,7,8,9,10,15,20,50,100,200)
tcolors <- rainbow(length(degsfreedom))

curve(dnorm(x),from=-4,to=4,col="black",lwd=6,ylab="t dist vs normal")

for (i in 1:length(degsfreedom)) {
  curve(dt(x,df=degsfreedom[i]),add=T,lwd=1,col=tcolors[i],xlab=c())
}
```

$$\frac{\bar{X}-\mu}{s/\sqrt{n}}$$

has a t-distribution of n-1 df which asymptotically approximates the Normal



n-1 df

```
degsfreedom <- c(1,2,3,4,5,6,7,8,9,10,15,20,50,100,200)
tcolors <- rainbow(length(degsfreedom))

curve(dnorm(x),from=-4,to=4,col="black",lwd=6,ylab="t dist vs normal")

for (i in 1:length(degsfreedom)) {
  curve(dt(x,df=degsfreedom[i]),add=T,lwd=1,col=tcolors[i],xlab=c())
}
```

```
se <- function(x)
      {
        y <- x[!is.na(x)] # remove the missing values, if any
        sqrt(var(as.vector(y))/length(y))
}


ci <- function (scores){
m <- mean(scores,na.rm=TRUE)
stderr <- se(scores)
len <- length(scores)
upper <- m + qt(.975, df=len-1) * stderr
lower <- m + qt(.025, df=len-1) * stderr
return(data.frame(lower=lower,upper=upper))
}
```

Vasishth 3.8

```r
se <- function(x)
      {
        y <- x[!is.na(x)] # remove the missing values, if any
        sqrt(var(as.vector(y))/length(y))
}


ci <- function (scores){
m <- mean(scores,na.rm=TRUE)
stderr <- se(scores)
len <- length(scores)
upper <- m + qt(.975, df=len-1) * stderr
lower <- m + qt(.025, df=len-1) * stderr
return(data.frame(lower=lower,upper=upper))
}
```

2.5%+2.5%=5% not in the region = 95% CI

Vasishth 3.8