March 11th: the t-test
Review of \[ \bar{X} = \frac{1}{n} (X_1 + \cdots + X_n) \]
Let $f(x)$ be the probability distribution of some given population from which we draw a sample of size $n$. Then it is natural to look for the probability distribution of the sample statistic $\bar{X}$, which is called the sampling distribution for the sample mean, or the sampling distribution of means. The following theorems are important in this connection.
Review of \( \bar{X} = \frac{1}{n} (X_1 + \cdots + X_n) \)

Let \( f(x) \) be the probability distribution of some given population from which we draw a sample of size \( n \). Then it is natural to look for the probability distribution of the sample statistic \( \bar{X} \), which is called the sampling distribution for the sample mean, or the sampling distribution of means. The following theorems are important in this connection.

**Theorem 5-1:** The mean of the sampling distribution of means, denoted by \( \mu_{\bar{X}} \), is given by

\[
E(\bar{X}) = \mu_{\bar{X}} = \mu
\]

where \( \mu \) is the mean of the population.

Theorem 5-1 states that the expected value of the sample mean is the population mean.
Review of $\bar{X} = \frac{1}{n} (X_1 + \cdots + X_n)$

Let $f(x)$ be the probability distribution of some given population from which we draw a sample of size $n$. Then it is natural to look for the probability distribution of the sample statistic $\bar{X}$, which is called the sampling distribution for the sample mean, or the sampling distribution of means. The following theorems are important in this connection.

**Theorem 5-1:** The mean of the sampling distribution of means, denoted by $\mu_{\bar{X}}$, is given by

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

where $\mu$ is the mean of the population.

Theorem 5-1 states that the expected value of the sample mean is the population mean.

**Theorem 5-2:** If a population is infinite and the sampling is random or if the population is finite and sampling is with replacement, then the variance of the sampling distribution of means, denoted by $\sigma^2_{\bar{X}}$, is given by

$$E[(\bar{X} - \mu)^2] = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

where $\sigma^2$ is the variance of the population.
Review of \( \bar{X} = \frac{1}{n} (X_1 + \cdots + X_n) \)

Let \( f(x) \) be the probability distribution of some given population from which we draw a sample of size \( n \). Then it is natural to look for the probability distribution of the sample statistic \( \bar{X} \), which is called the sampling distribution for the sample mean, or the sampling distribution of means. The following theorems are important in this connection.

**Theorem 5-1:** The mean of the sampling distribution of means, denoted by \( \mu_{\bar{X}} \), is given by

\[
E(\bar{X}) = \mu_{\bar{X}} = \mu
\]

where \( \mu \) is the mean of the population.

Theorem 5-1 states that the expected value of the sample mean is the population mean.

**Theorem 5-2:** If a population is infinite and the sampling is random or if the population is finite and sampling is with replacement, then the variance of the sampling distribution of means, denoted by \( \sigma^2_{\bar{X}} \), is given by

\[
E[(\bar{X} - \mu)^2] = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}
\]

where \( \sigma^2 \) is the variance of the population.
Review of \( \overline{X} = \frac{1}{n} (X_1 + \cdots + X_n) \)

**Theorem 5-4:** If the population from which samples are taken is normally distributed with mean \( \mu \) and variance \( \sigma^2 \), then the sample mean is normally distributed with mean \( \mu \) and variance \( \sigma^2 / n \).
Did $\bar{X}$ miss $\mu$ because of sampling error?

Recall that $\bar{X}$ is Normally distributed with variance $\frac{\sigma^2}{n}$.

So our question is

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

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

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

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

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < Z < \frac{\mu - \bar{X}}{\sigma/\sqrt{n}}$$
The quantity \( \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \)

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

\[
Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}
\]

is asymptotically normal, i.e.,

\[
\lim_{n \to \infty} P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} \, du
\]
Lacking $\sigma$, we cannot compute $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$. 
Lacking $\sigma$, we cannot compute $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$.

We can compute $\frac{\bar{X} - \mu}{s / \sqrt{n}}$ using the sample standard deviation.
Lacking $\sigma$, we cannot compute $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$.

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

$\frac{\bar{X} - \mu}{s / \sqrt{n}}$ has a t-distribution of $n-1$ df which asymptotically approximates the Normal
Asymptotically Normal

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())
}