Definitions and Key Concepts

- A sample statistic used to estimate an unknown population parameter is called an estimate.
- The discrepancy between the estimate and the true parameter value is known as sampling error.
- Sampling error is due to sampling variation.
Frequentist Approach to Statistics

• Assesses the accuracy of a sample estimate by considering how the estimate would vary around the true parameter value if repeated random samples are drawn from the same population
• A statistic is a random variable with a probability distribution - called the **sampling distribution** - which is generated by repeated sampling.
• We use the sampling distribution of a statistic to assess the sampling error of an estimate

Sample Mean

A **random sample** is a set of independently, identically distributed or i.i.d. observations \(X_1, X_2, \ldots, X_n\) (when sampling from a large population or with replacement)

Assumethatthepopulationhas mean \(\mu = E(X_i)\) and variance \(\sigma^2 = Var(X_i)\)

How doesthesamplemean \(\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i\) varyonrepeatedrandomsamplesofsize \(n\)?

This is called the sampling distribution of the sample mean.
Mean and Variance of a Die Toss

\[ \mu = \frac{1}{6} (1 + 2 + \ldots + 6) = 3.5 \]

\[ \sigma^2 = E(X^2) - \mu^2 = \frac{1}{6} (1^2 + 2^2 + \ldots + 6^2) - (3.5)^2 = 2.9166. \]

Simulating a Die Toss in JMP
Rolling Two Dice

Each one of these 36 outcomes are equally likely, i.e., each one occurs with 1/36 probability.

<table>
<thead>
<tr>
<th>(x_1, x_2)</th>
<th>μ</th>
<th>σ²</th>
<th>μ</th>
<th>σ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>1.0</td>
<td>0.0</td>
<td>(4, 1)</td>
<td>2.5</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>1.5</td>
<td>0.5</td>
<td>(4, 2)</td>
<td>3.0</td>
</tr>
<tr>
<td>(1, 3)</td>
<td>2.0</td>
<td>2.0</td>
<td>(4, 3)</td>
<td>3.5</td>
</tr>
<tr>
<td>(1, 4)</td>
<td>2.5</td>
<td>4.5</td>
<td>(4, 4)</td>
<td>4.0</td>
</tr>
<tr>
<td>(1, 5)</td>
<td>3.0</td>
<td>8.0</td>
<td>(4, 5)</td>
<td>4.5</td>
</tr>
<tr>
<td>(1, 6)</td>
<td>3.5</td>
<td>12.5</td>
<td>(4, 6)</td>
<td>5.0</td>
</tr>
<tr>
<td>(2, 1)</td>
<td>1.5</td>
<td>0.5</td>
<td>(5, 1)</td>
<td>3.0</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>2.0</td>
<td>0.0</td>
<td>(5, 2)</td>
<td>3.5</td>
</tr>
<tr>
<td>(2, 3)</td>
<td>2.5</td>
<td>0.5</td>
<td>(5, 3)</td>
<td>4.0</td>
</tr>
<tr>
<td>(2, 4)</td>
<td>3.0</td>
<td>2.0</td>
<td>(5, 4)</td>
<td>4.5</td>
</tr>
<tr>
<td>(2, 5)</td>
<td>3.5</td>
<td>4.5</td>
<td>(5, 5)</td>
<td>5.0</td>
</tr>
<tr>
<td>(2, 6)</td>
<td>4.0</td>
<td>8.0</td>
<td>(5, 6)</td>
<td>5.5</td>
</tr>
<tr>
<td>(3, 1)</td>
<td>2.0</td>
<td>2.0</td>
<td>(6, 1)</td>
<td>3.5</td>
</tr>
<tr>
<td>(3, 2)</td>
<td>2.5</td>
<td>0.5</td>
<td>(6, 2)</td>
<td>4.0</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>3.0</td>
<td>0.0</td>
<td>(6, 3)</td>
<td>4.5</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>3.5</td>
<td>0.5</td>
<td>(6, 4)</td>
<td>5.0</td>
</tr>
<tr>
<td>(3, 5)</td>
<td>4.0</td>
<td>2.0</td>
<td>(6, 5)</td>
<td>5.5</td>
</tr>
<tr>
<td>(3, 6)</td>
<td>4.5</td>
<td>4.5</td>
<td>(6, 6)</td>
<td>6.0</td>
</tr>
</tbody>
</table>

Rolling Two Dice Sampling Distribution

Table 5.2 Sampling Distribution of $\bar{X}$ for Samples of Size 2 from a Uniform Distribution on 1, 2, . . . , 6

<table>
<thead>
<tr>
<th>$\bar{X}$</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
<th>4.5</th>
<th>5.0</th>
<th>5.5</th>
<th>6.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(\bar{X})$</td>
<td>$\frac{1}{36}$</td>
<td>$\frac{2}{36}$</td>
<td>$\frac{3}{36}$</td>
<td>$\frac{4}{36}$</td>
<td>$\frac{5}{36}$</td>
<td>$\frac{6}{36}$</td>
<td>$\frac{1}{36}$</td>
<td>$\frac{2}{36}$</td>
<td>$\frac{3}{36}$</td>
<td>$\frac{4}{36}$</td>
<td>$\frac{5}{36}$</td>
</tr>
</tbody>
</table>

$E(\bar{X}) = 1.0 \times \frac{1}{36} + 2.0 \times \frac{2}{36} + \cdots + 6.0 \times \frac{1}{36} = 3.5$ and $\text{Var}(\bar{X}) = (1.0 - 3.5)^2 \times \frac{1}{36} + (2.0 - 3.5)^2 \times \frac{2}{36} + \cdots + (6.0 - 3.5)^2 \times \frac{1}{36} = 1.4583 = 2.9166/2$.

These values are consistent with the formulas $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$. 

6/14/2003  Unit 5 - Stat 571 - Ramon V. Leon
Homework To be Done Right Away

Use the Sampling Distribution simulation Java applet at the Rice Virtual Lab in Statistics to do the following.

• Draw 10,000 random samples of size N=5 from the normal distribution provided.
  – Construct the histogram of the sampling distribution of the sample mean.
  – Construct the histogram of the sampling distribution of the sample variance.
  – Turn in this output with the rest of the homework for Unit 5.

• Draw 10,000 random samples of size N=20 from the normal distribution provided.
  – Construct the histogram of the sampling distribution of the sample mean.
  – Construct the histogram of the sampling distribution of the sample variance.

• Draw 10,000 random samples of size N=5 from a uniform distribution on [0,32].
  – Construct the histogram of the sampling distribution of the sample mean.
  – Construct the histogram of the sampling distribution of the sample variance.

• Draw 10,000 random samples of size N=20 from a uniform distribution on [0,32].
  – Construct the histogram of the sampling distribution of the sample mean.
  – Construct the histogram of the sampling distribution of the sample variance.

• Draw 10,000 random samples of size N=5 from the skewed distribution provided.
  – Construct the histogram of the sampling distribution of the sample mean.
  – Construct the histogram of the sampling distribution of the sample variance.
  – Construct the histogram of the sampling distribution of the sample median.

• Draw 10,000 random samples of size N=20 from the skewed distribution provided.
  – Construct the histogram of the sampling distribution of the sample mean.
  – Construct the histogram of the sampling distribution of the sample variance.
  – Construct the histogram of the sampling distribution of the sample median.

Distribution of Sample Means

• If the i.i.d. r.v.’s are Bernoulli, Normal, or Exponential the distribution of the sample mean can be calculated exactly.

• However, in general the exact distribution of the sample mean is difficult to calculate.

• What can be said about the distribution of the sample mean when the sample is drawn from an arbitrary population?

• In many cases we can approximate the distribution of the sample mean when $n$ is large by a normal distribution.
  This result is called the Central Limit Theorem.
Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from an arbitrary distribution with a finite mean $\mu$ and variance $\sigma^2$. Then if $n$ is sufficiently large

$$\frac{\bar{X} - \mu}{\sigma \sqrt{n}} \approx N(0,1)$$

Sometimes the theorem is given in terms of the sums:

$$\frac{\sum_{j=1}^{n} X_i - n \mu}{\sigma \sqrt{n}} \approx N(0,1)$$

---

**Figure 5.1** Histograms of Means of 500 Random Samples from a $U[0, 10]$ Distribution: (a) Single observations from a uniform distribution; (b) Sample means from $n = 2$ observations; (c) Sample means from $n = 10$ observations; (d) Sample means from $n = 30$ observations.
Screen Shots of the Output of the Sampling Distribution Simulation Java Applet

Central Limit Theorem and Law of Large Numbers

- Both are asymptotic results about the sample mean
- Law of Large Numbers says that as $n$ goes to infinity the sample mean converges to the population mean, i.e.

$$\bar{X} - \mu \text{ converges to } 0 \text{ as } n \to \infty$$

- CLT says that as $n$ goes to infinity

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ converges to } N(0,1) \text{ as } n \to \infty$$
Central Limit Theorem

Let \( X_1, X_2, \ldots, X_n \) be a random sample drawn from an arbitrary distribution with a finite mean \( \mu \) and variance \( \sigma^2 \). Then if \( n \) is sufficiently large

\[
\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)
\]

Sometimes the theorem is given in terms of the sums:

\[
\frac{\sum_{j=1}^{n} X_i - n\mu}{\sigma \sqrt{n}} \approx N(0,1)
\]

Normal Approximation to the Binomial

A binomial r.v. is the sum of i.i.d. Bernoulli r.v.’s so the CLT can be used to approximate its distribution.

Suppose that \( Z \) is Bernoulli. Then the mean of \( Z \) is \( p \) and its variance is \( p(1-p) \).

By the CLT we have for the Binomial \( (n, p) \) r.v. \( X \):

\[
\frac{X - np}{\sqrt{np(1-p)}} = \frac{\sum_{i=1}^{n} Z_i - np}{\sqrt{np(1-p)}} = \frac{\sum_{i=1}^{n} Z_i - nE(Z)}{\sqrt{Var(Z) \times n}} \approx N(0,1)
\]

How large of a sample, \( n \), do we need for the approximation to be good?

Rule of Thumb: \( np \geq 10 \) and \( n(1-p) \geq 10 \)
CLT Approximation to the Binomial
When $p$ is Close to 0.5

![Histograms for $p = 0.5$, $n = 5$ (a) and $p = 0.5$, $n = 10$ (b)]

For a good approximation $np = n(1-p) = n0.5$ should be at least 10.
So, for a good approximation $n$ should be at least 20

CLT Approximation to the Binomial
When $p$ is Not Close to 0.5

$np = n(1-p)$ should be at least 10.
So $n$ should be at least 100

![Histograms for $p = 0.1$, $n = 10$ (a), $p = 0.1$, $n = 50$ (b), and $p = 0.1$, $n = 100$ (c)]
Continuity Correction

\[ P(X \leq x) \simeq \Phi \left( \frac{x + 0.5 - np}{\sqrt{np(1 - p)}} \right) \]

\[ P(X \leq 8) = \Phi \left( \frac{8.5 - np}{\sqrt{np(1 - p)}} \right) \]

Similarly:

\[ P(X \geq x) \simeq 1 - \Phi \left( \frac{x - 0.5 - np}{\sqrt{np(1 - p)}} \right) \]

\[ P(X \geq 8) = 1 - \Phi \left( \frac{7.5 - np}{\sqrt{np(1 - p)}} \right) \]

Screen Shots of the Output of the Java Applet
“Normal Approximation to the Binomial Distribution”

Homework: See the Homework Log.
Why the Normal Approximation to the Binomial Distribution Works in Pictures

Green area is approximately the same as the red area

Java Applet for \( N=100 \) and \( p=0.1 \)
Example: CLT Approximation to the Binomial

In a noisy communication channel there is a 2% chance that each transmitted bit (0 or 1) will be corrupted. If 1000 bits are transmitted, what is the probability that no more than 10 will be corrupted?

Let \( X \) = the number of corrupted bits. Then \( X \) has a binomial distribution with \( n = 1000 \) and \( p = 0.02 \). The required probability is

\[
P(X \leq 10) = \sum_{x=0}^{10} \binom{1000}{x} (0.02)^x (0.98)^{1000-x}.
\]

This probability is tedious to compute by hand. Using a computer package, we obtain the answer as 0.0102. The normal approximation can be computed more easily. We regard \( X \) as approximately normally distributed with \( \mu = np = (1000)(0.02) = 20 \) and \( \sigma = \sqrt{np(1-p)} = \sqrt{(1000)(0.02)(0.98)} = 4.427 \). Therefore

\[
P(X \leq 10) = P \left( Z = \frac{X - \mu}{\sigma} \leq \frac{10 - 20}{4.427} \right) \approx \Phi(-2.259) = 0.0118.
\]

This approximate probability is quite close to the exact binomial probability.

---

Table 5.1  All Possible Samples of Size 2 from a Uniform Distribution on 1, 2, . . . , 6 and Associated Sample Means and Variances

<table>
<thead>
<tr>
<th>((\chi_1, \chi_2))</th>
<th>(x)</th>
<th>(x^2)</th>
<th>(\overline{\chi}_1)</th>
<th>(\overline{\chi}_2)</th>
<th>(\overline{\chi})</th>
<th>(\overline{\chi^2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>1.0</td>
<td>0.0</td>
<td>(4, 1)</td>
<td>2.5</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>(1, 2)</td>
<td>1.5</td>
<td>0.5</td>
<td>(4, 2)</td>
<td>3.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>(1, 3)</td>
<td>2.0</td>
<td>2.0</td>
<td>(4, 3)</td>
<td>3.5</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(1, 4)</td>
<td>2.5</td>
<td>4.5</td>
<td>(4, 4)</td>
<td>4.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>(1, 5)</td>
<td>3.0</td>
<td>8.0</td>
<td>(4, 5)</td>
<td>4.5</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(1, 6)</td>
<td>3.5</td>
<td>12.5</td>
<td>(4, 6)</td>
<td>5.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>(2, 1)</td>
<td>1.5</td>
<td>0.5</td>
<td>(5, 1)</td>
<td>3.0</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>(2, 2)</td>
<td>2.0</td>
<td>0.0</td>
<td>(5, 2)</td>
<td>3.5</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>(2, 3)</td>
<td>2.5</td>
<td>0.5</td>
<td>(5, 3)</td>
<td>4.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>(2, 4)</td>
<td>3.0</td>
<td>2.0</td>
<td>(5, 4)</td>
<td>4.5</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(2, 5)</td>
<td>3.5</td>
<td>4.5</td>
<td>(5, 5)</td>
<td>5.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>(2, 6)</td>
<td>4.0</td>
<td>8.0</td>
<td>(5, 6)</td>
<td>5.5</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(3, 1)</td>
<td>2.0</td>
<td>2.0</td>
<td>(6, 1)</td>
<td>3.5</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>(3, 2)</td>
<td>2.5</td>
<td>0.5</td>
<td>(6, 2)</td>
<td>4.0</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>(3, 3)</td>
<td>3.0</td>
<td>0.0</td>
<td>(6, 3)</td>
<td>4.5</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>(3, 4)</td>
<td>3.5</td>
<td>0.5</td>
<td>(6, 4)</td>
<td>5.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>(3, 5)</td>
<td>4.0</td>
<td>2.0</td>
<td>(6, 5)</td>
<td>5.5</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(3, 6)</td>
<td>4.5</td>
<td>4.5</td>
<td>(6, 6)</td>
<td>6.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>
Sampling Distribution of the Sample Variance: Two Dice Example

Table 5.3 Sampling Distribution of \( s^2 \) for Samples of Size 2 from a Uniform Distribution on 1, 2, \ldots, 6

<table>
<thead>
<tr>
<th>( s^2 )</th>
<th>0.0</th>
<th>0.5</th>
<th>2.0</th>
<th>4.5</th>
<th>8.0</th>
<th>12.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(s^2) )</td>
<td>6/36</td>
<td>10/36</td>
<td>5/36</td>
<td>4/36</td>
<td>4/36</td>
<td>2/36</td>
</tr>
</tbody>
</table>

Example 5.4 (Rolling Two Dice: Sampling Distribution of \( S^2 \))

Using the values of \( s^2 \) for all possible samples in Table 5.1, we obtain the distribution given in Table 5.3. Note that the distribution of \( S^2 \) is skewed to the right, in contrast to the distribution of \( \bar{X} \) given in Table 5.2, which is symmetric. This is quite typical; in general, the distribution of \( \bar{X} \) is approximately symmetric for modestly large \( n \) even if the population distribution is not symmetric (and exactly symmetric for any \( n \) if the population distribution is symmetric), while the distribution of \( S^2 \) is always skewed to the right. The mean of the distribution of \( S^2 \) is

\[
E(S^2) = \frac{6}{36} (0) + \frac{10}{36} (0.5) + \frac{8}{36} (2.0) + \frac{6}{36} (4.5) + \frac{4}{36} (8.0) + \frac{2}{36} (12.5) = 2.9166,
\]

which verifies that \( E(S^2) = \sigma^2 \). \( \blacksquare \)

Chi-Square Distribution

Chi-Square Distribution. For \( v \geq 1 \), let \( Z_1, Z_2, \ldots, Z_v \) be i.i.d. \( \mathcal{N}(0, 1) \) r.v.'s and let

\[
X = Z_1^2 + Z_2^2 + \cdots + Z_v^2.
\]  
(53)

Then the probability density function (p.d.f) of \( X \) can be shown to be

\[
f(x) = \frac{1}{2^{v/2} \Gamma (v/2)} x^{(v/2)-1} e^{-x/2} \quad \text{for } x \geq 0.
\]  
(54)

This is known as the \( \chi^2 \)-distribution with \( v \) degrees of freedom (d.f.), abbreviated as \( X \sim \chi_v^2 \). Note that (5.4) is a special case of the gamma distribution (see Section 2.8.2) with \( \lambda = 1/2 \) and \( \epsilon = v/2 \).
Using JMP to Simulate a Chi-Square Random Sample with 5 d.f.

The number of rows is the size of the random sample.

See the JMP tutorial “Chi-Square Simulation” on the course home page.

Sample of 1000 Random Chi-Square Random Variables

Notice the right skewness.
Fitted Chi-Square Based on the Sample

Chi-Square Density Function Curves

Notice how similar is this density function to the histogram in the previous page.

Figure 5.5 Chi-Square Density Function Curves for $\nu = 5$ and 10
Critical Values for the Chi-Square

![Chi-Square Distribution Diagram]

See the JMP tutorial “Tabled Values of Common Distributions”

<table>
<thead>
<tr>
<th>( \chi^2_{v,0.05} )</th>
<th>( \chi^2_{v,0.025} )</th>
<th>( \chi^2_{v,0.01} )</th>
<th>( \chi^2_{v,0.005} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.01</td>
<td>0.0025</td>
<td>0.001</td>
</tr>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td>2</td>
<td>0.010</td>
<td>0.020</td>
<td>0.051</td>
</tr>
<tr>
<td>3</td>
<td>0.072</td>
<td>0.115</td>
<td>0.216</td>
</tr>
<tr>
<td>4</td>
<td>0.207</td>
<td>0.297</td>
<td>0.484</td>
</tr>
</tbody>
</table>

Distribution of Sample Variance

Assuming that the random sample comes from a normal distribution

It can be shown that (see the derivation given at the end of this section)

\[
\frac{(n - 1)S^2}{\sigma^2} \sim \chi^2_{n-1}
\]  

(5.7)

or equivalently

\[
S^2 \sim \frac{\sigma^2 \chi^2_{n-1}}{n-1}
\]  

(5.8)

From the expressions for the mean and variance of the gamma distribution given in Section 2.8.2, we obtain

\[
E(\chi^2_{v}) = v
\]

and

\[
\text{Var}(\chi^2_{v}) = 2v.
\]

From this it follows that

\[
E(S^2) = E \left( \frac{\sigma^2 \chi^2_{n-1}}{n-1} \right) = \frac{\sigma^2}{n-1} E(\chi^2_{n-1}) = \frac{\sigma^2}{n-1} \times (n-1) = \sigma^2
\]  

(5.9)
Application of the Distribution of Sample Variance – Measurement Precision

Introduction to the ideas of hypothesis testing

The precision of an instrument refers to its ability to make repeat measurements that closely agree. One measure of precision is the variance among the repeat measurements; the smaller the variance the greater the precision. Suppose that an advertised claim for a thermometer is $\sigma = 0.01^\circ F$, but it is suspected that the true $\sigma > 0.01^\circ F$. To verify this claim, 10 repeat measurements are taken on a controlled temperature bath using this thermometer. If the observed sample variance $s^2$ is significantly larger than the claimed variance $\sigma^2 = (0.01)^2$, this casts doubt on the advertised claim. In other words, if $s^2 > c(0.01)^2$ for sufficiently large $c > 0$, then it is unlikely that $\sigma^2 = (0.01)^2$. How large a value of $c$ should be required to confidently reject the claim that the true $\sigma^2 = (0.01)^2$?

Application of the Distribution of Sample Variance – Measurement Precision

One commonly used approach is to determine $c$ so that there is only a 5% chance that $S^2$ will be larger than $c(0.01)^2$ when in fact $\sigma^2 = (0.01)^2$. Using this approach there is only a 5% chance of incorrectly concluding that the claim is unjustified, so we can be “95% confident” (this phrase is defined rigorously in the next chapter) in our rejection of the claim. Thus we need to find a value of $c$ that satisfies

$$P\{S^2 > c(0.01)^2\} = 0.05$$

when the claim $\sigma^2 = (0.01)^2$ is true. We use the relationship

$$\frac{(n-1)S^2}{\sigma^2} = \frac{9S^2}{(0.01)^2} \sim \chi^2\cdot$$

to write the above equation as

$$P\left\{\frac{9S^2}{(0.01)^2} > \frac{9c(0.01)^2}{(0.01)^2}\right\} = P\{\chi^2_9 > 9c\} = 0.05.$$

Therefore $9c = \chi^2_{9,0.05} = 16.92$ or $c = \frac{16.92}{9} = 1.88$. Thus, if $s^2$ exceeds $1.88(0.01)^2$, we decide that the advertised claim is not supported by the data.

6/14/2003       Unit 5 - Stat 571 - Ramon V. Leon 34
Student’s \( t \)-Distribution

Consider a random sample, \( X_1, X_2, \ldots, X_n \) drawn from \( N(\mu, \sigma^2) \).

It is known that \( \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \) is exactly distributed as \( N(0,1) \) for any \( n \).

But \( T = \frac{(\bar{X} - \mu)}{S / \sqrt{n}} \) is not longer distributed as \( N(0,1) \).

The distribution of \( T \) is named Student’s \( t \)-distribution.

(A different distribution for each number \( \nu = n - 1 \) = degrees of freedom)

Play with the Java applet “Student’s \( t \) Distribution”

---

### \( t \)-Distribution Table

Table A.4 Critical Values \( t_{\nu, \alpha} \) for the \( t \)-Distribution

<table>
<thead>
<tr>
<th>( \nu )</th>
<th>( .10 )</th>
<th>( .05 )</th>
<th>( .025 )</th>
<th>( .01 )</th>
<th>( .005 )</th>
<th>( .001 )</th>
<th>( .0005 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
<td>31.821</td>
<td>63.657</td>
<td>318.31</td>
<td>636.62</td>
</tr>
<tr>
<td>2</td>
<td>1.886</td>
<td>2.920</td>
<td>4.303</td>
<td>6.634</td>
<td>9.925</td>
<td>22.326</td>
<td>31.598</td>
</tr>
<tr>
<td>3</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.541</td>
<td>5.841</td>
<td>10.213</td>
<td>12.924</td>
</tr>
<tr>
<td>4</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td>4.604</td>
<td>7.173</td>
<td>8.610</td>
</tr>
<tr>
<td>29</td>
<td>1.311</td>
<td>1.699</td>
<td>2.045</td>
<td>2.462</td>
<td>2.756</td>
<td>3.396</td>
<td>3.659</td>
</tr>
<tr>
<td>30</td>
<td>1.310</td>
<td>1.697</td>
<td>2.042</td>
<td>2.457</td>
<td>2.750</td>
<td>3.385</td>
<td>3.646</td>
</tr>
<tr>
<td>40</td>
<td>1.303</td>
<td>1.684</td>
<td>2.021</td>
<td>2.423</td>
<td>2.704</td>
<td>3.307</td>
<td>3.551</td>
</tr>
<tr>
<td>60</td>
<td>1.296</td>
<td>1.671</td>
<td>2.000</td>
<td>2.390</td>
<td>2.660</td>
<td>3.232</td>
<td>3.460</td>
</tr>
<tr>
<td>120</td>
<td>1.289</td>
<td>1.658</td>
<td>1.980</td>
<td>2.358</td>
<td>2.617</td>
<td>3.160</td>
<td>3.373</td>
</tr>
<tr>
<td>\infty</td>
<td>1.282</td>
<td>1.645</td>
<td>1.960</td>
<td>2.326</td>
<td>2.576</td>
<td>3.090</td>
<td>3.291</td>
</tr>
</tbody>
</table>

See the JMP tutorial “Tabled Values of Common Distributions”
Application of the $t$-Distribution Calculation – Process Control

Half-pint (8 oz.) milk cartons are filled at a dairy by a filling machine. To provide a check on the machine, a sample of 10 cartons is periodically measured. If the sample mean $\bar{x}$ deviates by more than a certain amount $d$ from the nominal value of 8 oz., i.e., if $|\bar{x} - 8| > d$, then the machine setting is adjusted. The chance of a false alarm indicating an unnecessary adjustment is to be limited to 1%. Find a formula for $d$.

If the true mean amount of milk, denoted by $\mu$, equals 8 oz. then no adjustment is required. Therefore the probability of a false alarm is given by

$$P \left( |\bar{x} - 8| > d | \mu = 8 \right) = 0.01.$$

---

Example: $t$-Distribution Calculation

If the true mean amount of milk, denoted by $\mu$, equals 8 oz. then no adjustment is required. Therefore the probability of a false alarm is given by

$$P \left( |\bar{x} - 8| > d | \mu = 8 \right) = 0.01.$$

Divide both sides of the inequality in the probability statement by $S/\sqrt{n}$, where $n = 10$, and use the fact that $\frac{\bar{x} - \mu}{S/\sqrt{n}}$ has a $t$-distribution with $10 - 1 = 9$ d.f. when $\mu = 8$. Therefore the above equation becomes

$$P \left( |T| > \frac{d}{S/\sqrt{10}} \right) = 0.01.$$

Let

$$c = \frac{d}{S/\sqrt{10}} \quad \text{or} \quad d = \frac{cS}{\sqrt{10}}.$$

Then $c$ must cut off an area $0.01/2 = 0.005$ in each tail of the $t$-distribution with 9 d.f., i.e., $c = t_{9,0.005} = 3.250$ from Table A.4. Thus the decision rule decides that the machine needs adjustment if

$$|\bar{x} - 8| > d = 3.250 \frac{s}{\sqrt{10}}.$$
F-Distribution

Consider two independent random samples,

\[ X_1, X_2, \ldots, X_n \] from an \( N(\mu_1, \sigma_1^2) \), \[ Y_1, Y_2, \ldots, Y_m \] from an \( N(\mu_2, \sigma_2^2) \).

Then \[ \frac{S_1^2}{\sigma_1^2} \] has an \( F \) distribution with \( \nu_1 = n_1 - 1 \) d.f.

in the numerator and \( \nu_2 = n_2 - 1 \) d.f. in the denominator.

![F-Distribution Table](image)

See the JMP tutorial “Tabled Values of Common Distributions”