Course Organization

**Software:** No specific software is required. SYSTAT, R, SOCR resource, etc.

**Test:** *Introduction to Probability and Statistics for Engineering and the Sciences* 5th edition -- Jay Devore

**Course Description, Class homepage, online supplements, VOH’s, etc.**

---

Course Organization

**Material Covered:** (Devore, Chapters 7-14)
- Review of Key Concepts (ch 01-06)
- Confidence Intervals (ch 07)
- Single Sample Hypotheses testing (ch 08)
- Inferences based on 2 samples (ch 09)
- One- Two- and Three-Factor ANOVA (ch 10)
- 2k Factorial Designs (ch 11)
- Linear Regression (ch 12)
- Multiple & Nonlinear Regression (ch 13)
- Goodness-of-Fit Testing (ch 14)

---

Overall Review

**What is a statistic?**
- Any quantity whose value can be calculated from sample data. It does not depend on any unknown parameter.
- Examples –

**What are Random Variables?**
- A function from the sample space to the real number line.

Before any data is collected, we view all observations and statistics as random variables

---

Properties of Expectation and Variance

- Let $X$ be a random variable and $a, b$ be constants. It follows that:
  
  $E[aX + b] = aE[X] + b$
  
  $Var[aX + b] = a^2Var[X]$
  
  $Var[X] = E[X^2] - (E[X])^2$
  
  $SD^2(X) = Var[X]$

---

Linear Combinations of Random Variables

Consider the collection of the independent random variables $X_1, \ldots, X_n$ where $E[X_i] = \mu_i$ and $Var[X_i] = \sigma_i^2$; and let $a_1, \ldots, a_n$ be constants. Define a random variable $Y$ by

$Y = a_1X_1 + \ldots + a_nX_n$

which is a linear combination of the $X_i$'s. It follows that

$E[a_1X_1 + \ldots + a_nX_n] = a_1E[X_1] + \ldots + a_nE[X_n] = a_1\mu_1 + \ldots + a_n\mu_n$

$Var[a_1X_1 + \ldots + a_nX_n] = a_1^2Var[X_1] + \ldots + a_n^2Var[X_n]$

$= a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2$
Random Sample

X₁, ..., Xₙ are an IID random sample of size n if:
1. The Xᵢ’s are independent random variables
2. Every Xᵢ has the same (identical) probability distribution

These conditions are equivalent to the Xᵢ’s being independent and identically distributed (iid) random variables.

Sample Mean and Total of a Random Sample

The sample mean is given by the random variable \( \bar{X} \) defined as:
\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

The sample total is given by the random variable \( T_0 \) defined as:
\[
T_0 = \sum_{i=1}^{n} X_i
\]

Mean and Variance of \( T_0 \)

For the total-sum random variable \( T_0 = X_1 + \ldots + X_n \)
\( T_0 \sim N(n\mu, n\sigma^2) \).

Mean and Variance of \( \bar{X} \)

For the total-sum random variable
\[
\bar{X} = \frac{1}{n} (X_1 + \ldots + X_n)
\]
\( \bar{X} \sim N(\mu, \sigma^2/n) \).

Linear Combinations of Normal Random Variables from a Random Sample

Let X₁, ..., Xₙ be a random sample from a normally distributed population with mean \( \mu \) and variance \( \sigma^2 \), i.e. \( X_i \sim N(\mu, \sigma^2) \). It follows that the random variable \( Y = a_1X_1 + \ldots + a_nX_n \) is normally distributed with mean \( a_1\mu + \ldots + a_n\mu \) and variance \( a_1^2\sigma^2 + \ldots + a_n^2\sigma^2 \). Hence, the sample mean and the sample total of the random sample will be normally distributed.

Central Limit Theorem

Arguably the most important theorem in Statistics (GUT theory)

The central limit theorem gives us information about the sample mean and the sample total for a "large" (n>30) random sample from a population that is not normally distributed. Specifically, it tells us that these will be approximately normally distributed. The larger n is, the better the approximation.
Example – Central Limit Theorem

When a certain type of electrical resistor is manufactured, the mean resistance is 4 ohms with a standard deviation of 1.5 ohms. If 36 batches are independently produced, what is the probability that the sample average resistance of the batch is between 3.5 and 4.5 ohms. What is the probability that the sample total resistance is greater than 140 ohms?

Do InteractiveNormalCurve & CLT_Sampling Distribution Applets from SOCR resource

Uni- vs. Multi-modal histograms

- Number of clear humps on the frequency histogram plot determines the modality of a histogram plot.

Skewness & Symmetry of histograms

- A histogram is symmetric if the bars (bins) to the left of some point (mean) are approximately mirror images of those to the right of the mean.
- Histogram is skewed if it is not symmetric, the histogram is heavy to the left or right, or non-identical on both sides of the mean.

Skewness & Kurtosis

- What do we mean by symmetry and positive and negative skewness? Kurtosis? Properties!!!
  \[ \text{Skewness} = \frac{1}{N} \sum (x_i - \bar{x})^3, \quad \text{Kurtosis} = \frac{1}{N} \sum (x_i - \bar{x})^4 \]
  - Skewness is linearly invariant: Sk(aX+b)=Sk(X)
  - Skewness is a measure of unsymmetry
  - Kurtosis is (also linearly invariant) a measure of flatness
  - Both are used to quantify departures from StdNormal
  - Skewness(StdNorm)=0; Kurtosis(StdNorm)=3

Comparing 3 plots of the same data

Stem-and-leaf of strength N = 33
Leaf Unit = 10

<table>
<thead>
<tr>
<th>Leaf Unit</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
</tr>
</tbody>
</table>

Three graphs of the breaking-strength data for gear-teeth in positions 4 & 10 (Minitab output).

Important points

1. The distinction between a randomized experiment and an observational study is made at the time of result interpretation. The very same statistical analysis is carried for the two situations.
2. We’ve already stressed the importance of plotting data prior to stat-analysis. Plots have many important roles – prevent dangerous misconceptions from arising (data overlaps, clusters, outliers, skewness, trends in the data, etc.).
Analyzing Histogram Plots

- **Modality** – uni- vs. multi-modal (Why do we care?)
- **Symmetry** – how skewed is the histogram?
- **Center of gravity** for the Histogram plot – does it make sense?
- If center of gravity exists, quantify the spread of the frequencies around this point.
- **Strange patterns** – gaps, atypical frequencies lying away from the center.

Measures of central tendency (location)

- **Mean** – sum of all observations divided by their number
- **Median** – (second quartile, Q2) is the half-way-point for the distribution, 50% of all data are greater than it and 50% are smaller than Q2.
- **Mode** – the (list of) most frequently occurring observation(s).

Measures of variability (deviation)

- **Mean Absolute Deviation (MAD)** –
  \[ MAD = \frac{1}{n} \sum_{i=1}^{n} |y_i - \bar{y}| \]
- **Variance** –
  \[ Var = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \]
- **Standard Deviation** –
  \[ SD = \sqrt{Var} = s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2} \]

Example:
- **Mean Absolute Deviation** –
  \[ MAD = \frac{1}{n} \sum_{i=1}^{n} |y_i - \bar{y}| \]
- **Variance** –
  \[ Var = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \]
- **Standard Deviation** –
  \[ SD = \sqrt{Var} = s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2} \]
- **Example:**
  \[ X = \{1, 2, 3, 4\} \]
  \[ MAD = 4/3 = 1.33 \]
  \[ Var = 5/3 = 1.67 \]
  \[ SD = 1.3 \]

Trimmed, Winsorized means and Resistancy

- A data-driven parameter estimate is said to be **resistant** if it does not greatly change in the presence of outliers.
- **K-times trimmed mean**
  \[ \bar{y}_{tk} = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} y_{(i)} \]
- **Winsorized k-times mean**
  \[ \bar{y}_{wk} = \frac{1}{n} \left[ (k+1) y_{(k+1)} + \sum_{i=k+2}^{n-k} y_{(i)} + (k+1) y_{(n-k+1)} \right] \]

Stationary or Non-Stationary Process?

- **To assess stationarity:**
  - **Rigorous assessment:** A stationary process has a constant mean, variance, and autocorrelation through time/place.
  - **Visual assessment:** (Plot the data – observed vs. time/place – the parameter we argue stationarity with respect to).

- **Order statistic**

- **Time-Series Plot of the KWH Data**
Stationary or Non-Stationary Process?

- **Visual assessment**: Plot the data – observed vs. time/place, etc., parameter we argue stationarity with respect to.

![Scatter Plot of the KWH Data](image1)

Moving Averages

- **Signal, Noise, Filtering**: Oftentimes high frequency oscillations in the data make it difficult to read/interpret the data.

![Moving Average Effects on the Raw Data](image2)

Properties of probability distributions

- A sequence of numbers \( \{p_1, p_2, p_3, \ldots, p_n\} \) is a probability distribution for a sample space \( S = \{s_1, s_2, s_3, \ldots, s_n\} \), if \( pr(s_k) = p_k \), for each \( 1 \leq k \leq n \). The two essential properties of a probability distribution \( p_1, p_2, \ldots, p_n \)?

\[
\begin{align*}
p_1 & \geq 0; \quad \sum p_k = 1 \\
\end{align*}
\]

- How do we get the probability of an event from the probabilities of outcomes that make up that event?

- If all outcomes are distinct & equally likely, how do we calculate \( pr(A) \) if \( A = \{a_1, a_2, a_3, \ldots, a_9\} \) and \( pr(a_1) = pr(a_2) = \ldots = pr(a_9) = p \); then

\[
pr(A) = 9 \times pr(a_1) = 9p.
\]

Conditional Probability

The *conditional probability* of \( A \) occurring *given* that \( B \) occurs is given by

\[
pr(A \mid B) = \frac{pr(A \text{ and } B)}{pr(B)}
\]

Suppose we select one out of the 400 patients in the study and we want to find the probability that the cancer is on the extremities given that it is of type nodular: \( P = \frac{73}{125} = P(C \text{ on Extremities} | \text{Nodular}) \)

#nodular patients with cancer on extremities
#nodular patients

Multiplication rule- what's the percentage of Israelis that are poor and Arabic?

\[
pr(A \text{ and } B) = pr(A \mid B)pr(B) = pr(B \mid A)pr(A)
\]

![Illustration of the multiplication rule](image3)
**Permutation & Combination**

**Permutation:** Number of ordered arrangements of \( r \) objects chosen from \( n \) distinctive objects

\[
P^n_r = \frac{n!}{(n-r)!} = n(n-1)(n-2)\ldots(n-r+1)
\]

**Combination:** Number of non-ordered arrangements of \( r \) objects chosen from \( n \) distinctive objects:

\[
C^n_r = \frac{P^n_r}{r!} = \frac{n!}{(n-r)!r!}
\]

Example:

1. Suppose car plates are 7-digit, like AB1234. If all the letters can be used in the first 2 places, and all numbers can be used in the last 4, how many different plates can be made? How many plates are there with no repeating digits?

Solution:

a) \( 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \)

b) \( P_{26}^2 \cdot P_{10}^3 = 26 \cdot 25 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \)

**Examples**

2. How many different letter arrangement can be made from the 11 letters of MISSISSIPPI?

Solution: There are: 1 M, 4 I, 4 S, 2 P letters.

**Method 1:** consider different permutations:

\[
\frac{11!}{(1!)(4!)(4!)(2!)} = 34650
\]

**Method 2:** consider combinations:

\[
\frac{11!}{4!4!2!} = \ldots = \frac{11}{2} \cdot \frac{9}{4} \cdot \frac{5}{4} \cdot \frac{1}{1}
\]
Examples

3. There are N telephones, and any 2 phones are connected by 1 line. Then how many lines are needed all together?

Solution: \( C^2_N = N (N - 1) / 2 \)

If, \( N=5 \), complete graph with 5 nodes has \( C^2_5=10 \) edges.

Binomial theorem & multinomial theorem

Binomial theorem \( (a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \)

Deriving from this, we can get such useful formula \( (a=b=1) \)

\( \binom{n}{0} + \binom{n}{1} + \binom{n}{2} = 2^n = (1+1)^n \)

Also from \( (1+x)^{m+n}=(1+x)^m(1+x)^n \) we obtain:

\( \binom{m+n}{k} = \sum_{i=0}^{k} \binom{m}{i} \binom{n}{k-i} \)

On the left is the coeff of \( 1^k x^{m+n-k} \). On the right is the same coeff in the product of \( (\ldots+coeff \times x^{(i)}+\ldots) \times (\ldots+coeff \times x^{(j)}+\ldots) \).

Multinomial theorem

\( (a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \)

Generalization: Divide n distinctive objects into r groups, with the size of every group \( n_1, n_2, \ldots, n_r \) and \( n_1+n_2+\ldots+n_r = n \)

\( (x_1 + x_2 + \ldots + x_r)^n = \sum \binom{n}{n_1,n_2,\ldots,n_r} x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r} \)

where \( \binom{n}{n_1,n_2,\ldots,n_r} = \frac{n!}{n_1! n_2! \ldots n_r!} \)

Sterling Formula for asymptotic behavior of n!

Sterling formula:

\( n! = \sqrt{2\pi n} \times \left( \frac{n}{e} \right)^n \)

Probability and Venn diagrams

Proposition

\[
\begin{align*}
Pr(A_1 \cup A_2 \cup \ldots \cup A_j) &= \\
&= \sum_{j=1}^{n} Pr(A_j) - \sum_{j=1 \leq i < j \leq n} Pr(A_i \cap A_2) + \ldots \\
&\quad + (-1)^{j-1} \sum_{j \leq i_1 < j_2 < \ldots < i_{j-1}} \Pr(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_{j-1}}) + \ldots \\
&\quad + (-1)^{j-1} \Pr(A_1 \cap A_2 \cap \ldots \cap A_j)
\end{align*}
\]

Discrete Variables, Probabilities
Binomial Probabilities – the moment we all have been waiting for!

Suppose $X \sim \text{Binomial}(n, p)$, then the probability

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad 0 \leq x \leq n$$

Where the binomial coefficients are defined by

$$\binom{n}{x} = \frac{n!}{(n-x)! x!}, \quad n! = 1 \times 2 \times 3 \times \ldots \times (n-1) \times n$$

Expected values

- The game of chance: cost to play: $1.50; Prices: $1, $2, $3; probabilities of winning each price are: 0.6, 0.3, 0.1, respectively.
- Should we play the game? What are our chances of winning/loosing?

<table>
<thead>
<tr>
<th>Prize ($)</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.6</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>Number of games won</td>
<td>0.6 x 100</td>
<td>0.3 x 100</td>
<td>0.1 x 100</td>
</tr>
<tr>
<td>Sum</td>
<td>1.0 x 100</td>
<td>0.3 x 100</td>
<td>0.1 x 100</td>
</tr>
<tr>
<td>Total prize money = Sum; Average prize money = $1.50/100 = 1.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Theoretically Fair Game: price to play $E=\text{the expected return}$

For the Binomial distribution . . . mean

- $E(X) = np$
- $\text{sd}(X) = \sqrt{np(1-p)}$

$X \sim \text{Binomial}(n, p) \Rightarrow X=Y_1+Y_2+Y_3+\ldots+Y_n,$

where $Y_k \sim \text{Bernoulli}(p)$,

- $E(Y_k) = p$
- $E(X) = E(Y_1+Y_2+Y_3+\ldots+Y_n) = np$

Poisson Distribution – Definition

- Used to model counts – number of arrivals (k) on a given interval . . .
- The Poisson distribution is also sometimes referred to as the distribution of rare events. Examples of Poisson distributed variables are number of accidents per person, number of sweepstakes won per person, or the number of catastrophic defects found in a production process.

Functional Brain Imaging - Positron Emission Tomography (PET)

- Used to model counts – number of arrivals (k) on a given interval . . .
- $Y \sim \text{Poisson}(\lambda)$, then $P(Y=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$
- Mean of $Y$, $\mu_Y = \lambda$, since

$$E(Y) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^\lambda = \lambda$$
**Poisson Distribution - Variance**

- Y~Poisson(\(\lambda\)), then P(Y=k) = \(\frac{e^{-\lambda} \lambda^k}{k!}\), k=0, 1, 2, ...
- Variance of Y, \(\sigma^2_Y = \lambda\), since
- For example, suppose that Y denotes the number of blocked shots (arrivals) in a randomly sampled game for the UCLA Bruins men's basketball team. Then a Poisson distribution with mean=4 may be used to model Y.

**Poisson as an approximation to Binomial**

- Suppose we have a sequence of Binomial(n, \(p_n\)) models, with \(\lim(n p_n) \rightarrow \lambda\), as \(n \rightarrow \infty\).
- For each 0<=y<=n, if \(Y_n \sim\) Binomial(n, \(p_n\)), then
  \[
  P(Y_n=y) = \binom{n}{y} p_n^y (1-p_n)^{n-y}
  \]
- But this converges to:
  \[
  \lim_{n \rightarrow \infty} \binom{n}{y} p_n^y (1-p_n)^{n-y} \rightarrow \frac{\lambda^y e^{-\lambda}}{y!}
  \]
- Thus, Binomial(n, \(p_n\)) \(\rightarrow\) Poisson(\(\lambda\))

**Example using Poisson approx to Binomial**

- Suppose P(defective chip) = 0.0001=10\(^{-4}\). Find the probability that a lot of 25,000 chips has > 2 defective!
- \(Y \sim\) Binomial(25,000, 0.0001), find P(Y>2). Note that \(Z \sim\) Poisson(\(\lambda\) =n \(p = 25,000 \times 0.0001 = 2.5\))

\[
\begin{align*}
P(Z > 2) &= 1 - P(Z \leq 2) = 1 - \sum_{z=0}^{2} \frac{2.5^z e^{-2.5}}{z!} \\
&= 1 - \left( \frac{2.5^0 e^{-2.5}}{0!} + \frac{2.5^1 e^{-2.5}}{1!} + \frac{2.5^2 e^{-2.5}}{2!} \right) = 0.456
\end{align*}
\]

**Geometric, Hypergeometric, Negative Binomial**

- X ~ Geometric(p), then the probability mass function is
  \[
P(X = x) = (1-p)^{x-1} p; \quad E(X) = \frac{1-p}{p}; \quad Var(X) = \frac{1-p}{p^2}
  \]
- Ex: Stat dept purchases 40 light bulbs; 5 are defective. Select 5 components at random. Find: P(3rd bulb used is the first that does not work) = ?

**Hypergeometric – X~HyperGeom(x; N, n, M)**

Total objects: N, Successes: M, Sample-size: n (without replacement). X = number of Successes in sample

\[
E(X) = \frac{n M}{N}
\]

\[
Var(X) = \frac{N-n}{N-1} \frac{M}{N} \frac{N-M}{N} \frac{x}{n-x}
\]

Ex: 40 components in a lot; 3 components are defectives. Select 5 components at random. P(obtain one defective) = P(X=1) = ?
Hypergeometric Distribution & Binomial

- Binomial approximation to Hypergeometric
  - $\frac{n}{N}$ is small (usually < 0.1), then $\frac{M}{N} \approx p$
  - $HyperGeom(x; N,n,M) \Rightarrow Bin(x;n,p)$
  - $M / N \approx p$

Ex: 4,000 out of 10,000 residents are against a new tax. 15 residents are selected at random.

$P_{HyperGeom}(\text{at most 7 favor the new tax}) = ?$ (0.78706)

Demonstration:
- Applets.dir/ProbCalc.htm
- $P_{Bin}(Y \leq 7) = 0.7869$
- $HyperGeom(x; N=10^4, n=15, M=4 \times 10^3)$
- $Bin(x;n=15, p=0.4)$

Geometric, Hypergeometric, Negative Binomial

- Negative binomial pmf $[X \sim NegBin(r, p), \text{if } r=1 \Rightarrow Geometric(p)]$
  - $P(X = n) = \frac{(n-1)!}{(r-1)!} p^r (1-p)^{n-r}$

Number of trials until the rth success (negative, since number of successes (r) is fixed & number of trials (X) is random)

- $P(X = n) = \frac{(n-1)!}{(r-1)!} p^r (1-p)^{n-r}$

Ex: 4,000 out of 10,000 residents are against a new tax. 15 residents are selected at random.

$P_{HyperGeom}(\text{at most 7 favor the new tax}) = ?$ (0.78706)

Demonstration:
- $Applets.dir/ProbCalc.htm$
- $P_{Bin}(Y \leq 7) = 0.7869$
- $HyperGeom(x; N=10^4, n=15, M=4 \times 10^3)$
- $Bin(x;n=15, p=0.4)$

Continuous RV’s

- A RV is continuous if it can take on any real value in a non-trivial interval $(a; b)$.
- PDF, probability density function, for a cont. RV, $Y$, is a non-negative function $p_Y(y)$, for any real value $y$, such that for each interval $(a; b)$, the probability that $Y$ takes on a value in $(a; b)$, $P(a < Y < b)$, equals the area under $p_Y(y)$ over the interval $(a; b)$

Finding $E(X)$ and $Var(X)$

- $E(X) = \frac{r}{p}$
- $Var(X) = \frac{r(1-p)}{p^2}$

Convergence of density histograms to the PDF

- For a continuous RV the density histograms converge to the PDF as the size of the bins goes to zero.

Measures of central tendency/variability for Continuous RVs

- Mean
  - $\mu_Y = \int_{-\infty}^{\infty} y \times p_Y(y) dy$
- Variance
  - $\sigma_Y^2 = \int_{-\infty}^{\infty} (y - \mu_Y)^2 \times p_Y(y) dy$
- SD
  - $\sigma_Y = \sqrt{\int_{-\infty}^{\infty} (y - \mu_Y)^2 \times p_Y(y) dy}$

Facts about PDF’s of continuous RVs

- Non-negative $p_Y(y) \geq 0, \forall y$
- Completeness $\int_{-\infty}^{\infty} p_Y(y) dy = 1$
- Probability $P(a < Y < b) = \int_{a}^{b} p_Y(y) dy$
Continuous Distributions

- Uniform distribution
- Normal distribution
- Student’s T distribution
- F-distribution
- Chi-squared ($\chi^2$)
- Cauchy’s distribution
- Exponential distribution
- Poisson distribution, ...

(Continuous) Uniform Distribution

- $X$ – Uniform Distribution with parameters $\alpha$ and $\beta$
  $$f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \alpha < x < \beta \\ 0, & \text{otherwise} \end{cases}$$
- $E(X) = \frac{\alpha + \beta}{2}$
- $Var(X) = \frac{(\beta - \alpha)^2}{12}$

Ex) Uniform, $\alpha = 2$, $\beta = 7$

(a) $P(X \geq 4) = \frac{7 - 2}{7 - 2} = 5.53$
(b) $P(3 < X < 5.5) = \frac{5.5 - 2}{7 - 2} = \frac{3.5}{5} = 0.7$

(General) Normal Distribution

- Normal Distribution PDF: $Y \sim \text{Normal}(\mu, \sigma^2)$
  $$p_Y(y) = \frac{e^{-(y-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}, \forall -\infty < y < \infty$$
  $$F_Y(y) = \int_{-\infty}^{y} p_Y(x)dx = \int_{-\infty}^{y} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

Continuous Distributions – Student’s T

- Student’s T distribution [approx. of Normal(0,1)]
- IID from a Normal($\mu, \sigma$)
- Variance $\sigma^2$ is unknown
- In 1908, William Gosset (pseudonym Student) derived the exact sampling distribution of the following statistics
  $$T = \frac{\bar{Y} - \mu}{\hat{\sigma}}$$
  $$T \sim \text{Student}(df=N-1)$$
- $E(T) = 0$
- $Var(T) = \frac{\sigma^2}{(N-1)\hat{\sigma}^2}$

Continuous Distributions – $\chi^2$ [Chi-Square]

- $\chi^2$ [Chi-Square] goodness of fit test:
  - Let $\{X_1, X_2, ..., X_N\}$ are IID $N(0, 1)$
  - $W = X_1^2 + X_2^2 + X_3^2 + ... + X_N^2$
  - $W \sim \chi^2(\text{df}=N)$
- Note: If $\{Y_1, Y_2, ..., Y_N\}$ are IID $N(\mu, \sigma^2)$, then
  $$SD^2(Y) = \frac{1}{N-1} \sum_{k=1}^{N} (Y_k - \bar{Y})^2$$
  $$W = \frac{N-1}{\sigma^2} SD^2(Y)$$
- And the Statistics $W \sim \chi^2(\text{df}=N-1)$
- $E(W) = N$; $Var(W) = 2N$
The practical meaning of this is that collecting 1,000 approximations of electronic devices, or arrivals of customers at the check-out counter in a grocery store.

Thus, the exponential distribution is frequently used to model the time interval between successive random events. Examples of variables distributed in this manner would be the gap length between cars crossing an intersection, life-times of electronic devices, or arrivals of customers at the check-out counter in a grocery store.

Continuous Distributions – Exponential

- Exponential distribution, X~Exponential(λ)
- The exponential model, with only one unknown parameter, is the simplest of all life distribution models.

\[
 f(x) = \lambda e^{-\lambda x}, \quad x \geq 0
\]

E(X) = 1/λ; Var(X) = 1/λ².

Another name for the exponential mean is the most time to fail or MTTF and we have MTTF = 1/λ.

If X is the time between occurrences of rare events that happen on the average with a rate 1 per unit of time, then X is distributed exponentially with parameter λ. Thus, the exponential distribution is frequently used to model the time interval between successive random events. Examples of variables distributed in this manner would be the gap length between cars crossing an intersection, life-times of electronic devices, or arrivals of customers at the check-out counter in a grocery store.

Continuous Distributions – F-distribution

- F-distribution is the ratio of two independent chi-squares divided by their respective degrees of freedom.

\[
 F = \frac{X_1}{\sigma_1^2} / \frac{X_2}{\sigma_2^2}
\]

where X₁ and X₂ are independent chi-squares.

The Cauchy distribution is (theoretically) important as an example of a pathological case. Cauchy distributions look similar to a normal distribution. However, they have much heavier tails. When studying hypothesis tests that assume normality, seeing how the tests perform on data from a Cauchy distribution is a good indicator of how sensitive the tests are to heavy-tail departures from normality. The mean and standard deviation of the Cauchy distribution are undefined!! The practical meaning of this is that collecting 1,000 data points gives no more accurate of an estimate of the mean and standard deviation than does a single point (Cauchy=Tdf=0 Normal).

Normal approximation to Binomial – Example

- Roulette wheel investigation:
- Compute P(Y>=58), where Y~Binomial(100, 0.47) –
  - The proportion of the Binomial(100, 0.47) population having more than 58 reds (successes) out of 100 roulette spins (trials).
  - Since np=47>=10 & n(1-p)=53>10 Normal approx is justified.

\[
 Z = \frac{Y - np}{\sqrt{np(1-p)}} = \frac{58 - 100*0.47}{\sqrt{100*0.47*0.53}} = 2.2
\]

- Roulette has 38 slots [18red 18black 2neutral]
- P(Y>=58) ≈ P(Z>=2.2) = 0.0139
- True P(Y>=58) = 0.177, using SOCR (demo!)
- Binomial approx useful when no access to SOCR avail.
Normal approximation to Poisson

- Let $X_i \sim \text{Poisson}(\lambda)$ and $X_i \sim \text{Poisson}(\mu)$, then $X_i + X_j \sim \text{Poisson}(\lambda + \mu)$
- Let $X_1, X_2, X_3, ..., X_n \sim \text{Poisson}(\lambda)$ and independent,
- $Y_k = X_1 + X_2 + ... + X_k \sim \text{Poisson}(k\lambda)$, and
- $E(Y_k) = Var(Y_k) = k\lambda$.
- The random variables in the sum on the right are independent and each has the Poisson distribution with parameter $\lambda$.
- By CLT the distribution of the standardized variable $(Y_k - k\lambda) / (k\lambda)^{1/2}$ approaches $N(0,1)$, as $k$ increases to infinity.
- So, for $k\lambda \geq 100$, $Z_k = (Y_k - k\lambda) / (k\lambda)^{1/2}$ follows $N(0,1)$.
- $Y_k \sim N(k\lambda, (k\lambda)^{1/2})$.

Poisson or Normal approximation to Binomial?

- Poisson Approximation: $\text{Binomial}(n, p) \Rightarrow \text{Poisson}(\lambda)$
  \[ n p \to \lambda, \quad \frac{(n-p)}{\sqrt{n p (1-p)}} \to \frac{\lambda}{\sqrt{\lambda}}, \quad y! \to e^{-\lambda} \lambda^y \]
- Normal Approximation:
  \[ \text{Binomial}(n, p) \Rightarrow N\left( \frac{np(1-p)}{\lambda}, \frac{np(1-p)}{\lambda} \right) \]
- $n \approx 100$ and $p \approx 0.01$ and $\lambda = np \approx 20$
- Normal Approximation:
  \[ \text{Binomial}(n, p) \Rightarrow N\left( \frac{np(1-p)}{\lambda}, \frac{np(1-p)}{\lambda} \right) \]
- $np \geq 10$ and $n(1-p) > 10$

Areas under Normal Curve - Example

- Many histograms are similar in shape to the standard normal curves. For example, persons height. The height of all incoming female army recruits is measured for custom training and assignment purposes (e.g., very tall people are inappropriate for constricted space positions, and very short people may be disadvantages in certain other situations). The mean height is computed to be 64 in and the standard deviation is 2 in. Only recruits shorter than 65.5 in will be trained for tank operation and recruits within 1/2 standard deviations of the mean will have no restrictions on duties.
- What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?
- About what percentage of the recruits will have no restrictions on duties?
- The mean height is 64 in and the standard deviation is 2 in.
- Only recruits shorter than 65.5 in will be trained for tank operation.
- What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?
- Recruits within 1/2 standard deviations of the mean will have no restrictions on duties. About what percentage of the recruits will have no restrictions on training/duties?
- $65.5 \to 65.5-64/2 = 1/4$ Percentage is 33.35%
- $60 \to (60-64)/2 = -2$ Percentage is 25%
- $66 \to (66-64)/2 = 1$ Percentage is 83.35%
- $X \sim \text{Gamma}(\alpha, \beta)$
- $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ for $\alpha > 0$
- Properties:
  \[ \Gamma(\alpha + n) = (\alpha - 1) \Gamma(\alpha) \]
- $\Gamma(1) = 1$ for positive integer $n$
- $\Gamma(0.5) = \sqrt{\pi}$
- $X \sim \text{Gamma}(\alpha, \beta)$
- $f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$
- where $\alpha > 0$, $\beta > 0$
Gamma and Exponential Distributions

- **Exponential Distribution (cont’d)**
  - CDF: \( F(x) = P(X \leq x) = \int_0^x \frac{1}{\beta} e^{-\frac{x}{\beta}} \, dx = 1 - e^{-\frac{x}{\beta}}, \ x > 0 \)
  - Tail probability
    \[ P(X > x) = 1 - F(x) = e^{-\frac{x}{\beta}}, \ x > 0 \]

Ex 1) \( X \) = response time at a certain on-line computer terminal
- \( X \) is exponential with \( \text{E}(X) = 5 \) (sec).

- (a) \( P(X \leq 10) = \)
- (b) \( P(5 \leq X \leq 10) = \)

Lognormal Distribution

- \( X \) ~ lognormal with parameters \( \mu \) and \( \sigma \), if
  \[ \ln(X) \sim N(\mu, \sigma^2) \]
  \[ f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x) - \mu}{\sigma}\right)^2}, \ x \geq 0 \]
  \[ 0, \ \text{otherwise} \]

- \( E(X) = \exp(\mu + \sigma^2/2) \)
  \( \text{Var}(X) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1) \)

Ex) Let \( X \) ~ lognormal with parameter \( \mu = 3.2 \) and \( \sigma = 1 \)
- \( P(X > 8) = \)

Beta Distribution

- Provides positive density only in an interval of finite length
  \( X \) ~ Beta Distribution with parameters \( \alpha \) and \( \beta \) if
  \[ f(x) = \begin{cases} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1 \ (\alpha, \beta > 0) \\ 0, & \text{otherwise} \end{cases} \]

- \( E(X) = \frac{\alpha}{\alpha + \beta} \)
  \( \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \)

Ex)
- \( X \) = proportion of TV sets requiring service during the first year
  \( \sim \) beta, \( \alpha = 3, \beta = 2 \).
  \( P(\text{at least 80% of the model sold this year will require service in 1 year}) \)

Marginal & Joint PDF’s

Central Limit Theorem (CLT)

Beta Distribution

- Provides positive density only in an interval of finite length
  \( X \) ~ Beta Distribution with parameters \( \alpha \) and \( \beta \) if
  \[ f(x) = \begin{cases} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1 \ (\alpha, \beta > 0) \\ 0, & \text{otherwise} \end{cases} \]

- \( E(X) = \frac{\alpha}{\alpha + \beta} \)
  \( \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \)

Ex)
- \( X \) = proportion of TV sets requiring service during the first year
  \( \sim \) beta, \( \alpha = 3, \beta = 2 \).
  \( P(\text{at least 80% of the model sold this year will require service in 1 year}) \)
Joint probability mass function

- The joint probability mass function of the discrete random variables X and Y, denoted as \( f_{XY}(x,y) \), satisfies:
  1. \( f_{XY}(x,y) \geq 0 \)
  2. \( \sum_{x} \sum_{y} f_{XY}(x,y) = 1 \)
  3. \( f_{XY}(x,y) = P(X = x, Y = y) \)

Marginal probability distributions

- Individual probability distribution of a random variable is referred to as its **Marginal Probability Distribution**.
- Marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables.
- Example: **Marginal probability distribution of X is found by summing the probabilities in each column, for Y, summation is done in each row.**

Mean and Variance

- If the marginal probability distribution of X has the probability function \( f(x) \), then
  \[
  E(X) = \mu_x = \sum_{x} x f(x) = \sum_{x} x f_{xy}(x,y) = \sum_{x} \sum_{y} x f_{xy}(x,y)
  \]
  \[
  V(X) = \sigma^2_x = \sum_{x} (x - \mu_x)^2 f(x) = \sum_{x} (x - \mu_x)^2 f_{xy}(x,y)
  \]
- R = Set of all points in the range of \((X,Y)\).
- Example.

Central Limit Theorem – heuristic formulation

**Central Limit Theorem:**

- When sampling from almost any distribution, \( \overline{X} \) is approximately Normally distributed in large samples.

Show Sampling Distribution Simulation Applet:
file:///C:/Ivo.dir/UCLA_Classes/Winter2002/AdditionalInstructorAids/SamplingDistributionApplet.html
Let \( \{X_1, X_2, \ldots, X_n\} \) be a sequence of independent observations from one specific random process. Let and both be finite \((0 < \sigma < \infty; \mu < \infty)\). If \( X \sim \frac{1}{n} \sum_{k=1}^{n} X_k \) sample-avg, then \( X \) has a distribution which approaches \( N(\mu, \sigma^2/n) \), as \( n \to \infty \).

Central Limit Theorem – theoretical formulation

The standard error of the mean

The standard error of the sample mean is an estimate of the SD of the sample mean

\[ \text{SE}(\bar{x}) = \frac{s}{\sqrt{n}} \]

\[ \text{SD}(\bar{x}) = \frac{\sigma}{\sqrt{n}} \]

\[ \text{SE}(\bar{x}) = \frac{s}{\sqrt{n}} \]

\[ \text{Note similarity with} \]

\[ \text{SD}(\bar{x}) = \frac{\sigma}{\sqrt{n}} \]

Cavendish’s 1798 data on mean density of the Earth, g/cm\(^3\) relative to that of H\(_2\)O

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.50</td>
<td>5.61</td>
<td>4.88</td>
<td>5.07</td>
<td>5.26</td>
</tr>
<tr>
<td>5.55</td>
<td>5.36</td>
<td>5.29</td>
<td>5.58</td>
<td>5.65</td>
</tr>
<tr>
<td>5.57</td>
<td>5.53</td>
<td>5.62</td>
<td>5.44</td>
<td>5.34</td>
</tr>
<tr>
<td>5.42</td>
<td>5.47</td>
<td>5.63</td>
<td>5.34</td>
<td>5.46</td>
</tr>
</tbody>
</table>

Source: Cavendish [1798]

Sample mean \( \bar{x} = 5.447931 \) g/cm\(^3\)

and sample SD = \( S_X = 0.2209457 \) g/cm\(^3\)

Then the standard error for these data is:

\[ \text{SE}(\bar{x}) = \frac{S_X}{\sqrt{n}} = \frac{0.2209457}{\sqrt{29}} = 0.04102858 \]

Student’s t-distribution

For random samples from a Normal distribution,

\[ T = \frac{(\bar{X} - \mu)}{SE(\bar{X})} \]

Recall that for samples from \( N(\mu, \sigma) \):

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

is exactly distributed as Student(\( df = n - 1 \)) \( \approx \) Approx Distributions

- but methods we shall base upon this distribution for \( T \) work well even for small samples sampled from distributions which are quite non-Normal
- \( df \) is number of observations – 1, degrees of freedom.

Inference & Estimation

Parameters, Estimators, Estimates …

E.g., We are interested in the population mean diameter (parameter) of washers the sample-average formula represents an estimator we can use, where as the value of the sample average for a particular dataset is the estimate (for the mean parameter).

\[ \text{parameter} = \mu; \quad \text{estimator} = \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i \]

Data: \( Y = \{0.1896, 0.1913, 0.1900\} \)

\[ \text{estimate} = \bar{y} = \frac{1}{3}(0.1896 + 0.1913 + 0.1900) \]

\[ \bar{y} = 0.1903 \]

How about \( \overline{\mu} = \frac{2}{3}(0.1896 + 0.1913 + 0.1900) \)
A 95% confidence interval

- A type of interval that contains the true value of a parameter for 95% of samples taken is called a 95% confidence interval for that parameter; the ends of the CI are called confidence limits.
- (For the situations we deal with) a confidence interval (CI) for the true value of a parameter, the ends of the CI are called confidence limits.

<table>
<thead>
<tr>
<th>Value of the Multiplier, $t$, for a 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$df$</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
</tbody>
</table>

Most of the previous samples contained the true mean?

Effect of increasing the confidence level

- 99% CI, $\bar{x} \pm 2.576$ se($\bar{x}$)
- 95% CI, $\bar{x} \pm 1.960$ se($\bar{x}$)
- 90% CI, $\bar{x} \pm 1.645$ se($\bar{x}$)
- 80% CI, $\bar{x} \pm 1.282$ se($\bar{x}$)

Why?

The greater the confidence level, the wider the interval

(General) Confidence Interval (CI)

- A level $L$ confidence interval for a parameter ($\theta$), is an interval ($\theta_1^L$, $\theta_2^L$), where $\theta_1^L$ & $\theta_2^L$ are estimators of $\theta$, such that $P(\theta_1^L < \theta < \theta_2^L) = L$.
- E.g., $C+E$ model: $Y = \mu + \varepsilon$. Where $\varepsilon \sim N(0, \sigma^2)$, then by CLT we have $Y_{\bar{y}} \sim N(\mu, \sigma^2/n)$.

**Area??** $n^1(Y_{\bar{y}} - \mu)/\sigma \sim N(0, \sigma^2)$.

- $L = P ( Z_{(1-L)/2} < n^1(Y_{\bar{y}} - \mu)/\sigma < Z_{(1+L)/2})$, where $Z_q$ is the $q$th quartile.

- E.g., 0.95 = $P ( Z_{0.025} < n^1(Y_{\bar{y}} - \mu)/\sigma < Z_{0.975} )$.

CI for population mean

Confidence Interval for the true (population) mean $\mu$: sample mean $\pm t$ standard errors or $\bar{x} \pm t$ se($\bar{x}$), where $SE(\bar{x}) = \frac{s}{\sqrt{n}}$ and $df = n - 1$

<table>
<thead>
<tr>
<th>Value of the Multiplier, $t$, for a 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$df$</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
</tbody>
</table>

Effect of increasing the sample size

- $n = 10$ data points
- $n = 40$ data points
- $n = 90$ data points

Increase Sample Size

Three random samples from a Normal($\mu = 24.83$, $s = 0.005$) distribution and their 95% confidence intervals for $\mu$.

To double the precision we need four times as many observations.
A marine biologist wishes to use male angelfish for an experiment and hopes their weights don’t vary much. In fact, a previous random sample of \( n = 16 \) angelfish yielded the data below
\[
\{y_1; \ldots; y_{16}\} = \{5.1; 2.5; 2.8; 3.4; 6.3; 3.6; 3.9; 3.0; 2.7; 5.7; 3.5; 3.6; 5.3; 5.1; 3.5; 3.3\}
\]
Sample statistics from these data include Avg. = 3.96 lbs, \( s^2 = 1.35 \) lbs, \( n = 16 \).

**Problem**: Obtain a 100(1 - \( \alpha \))% CI(\( \sigma^2 \)).

**Point Estimator for \( \sigma^2 \)? How about sample variance, \( s^2 \)?

Sampling theory for \( s^2 \)? Not in general, but under Normal assumptions ...

If a random sample \( \{Y_1; \ldots; Y_n\} \) is taken from a normal population with mean \( \mu \) and variance \( \sigma^2 \), then standardizing, we get a sum of squared \( N(0,1) \)

\[
\chi^2_{n-1} = \sum_{k=1}^{n} (y_k - \bar{Y})^2 / \sigma^2 \leq \chi^2_{n-1, 1 - \alpha} / \sigma^2
\]

This yields the CI, the sample variance is \( s^2 = 1.35 \). Note the CI is NOT symmetric (0.74 ; 3.24)

**Prediction vs. Confidence intervals – Differences?**

**Confidence Intervals** (for the population mean \( \mu \)):
\[
\left( \bar{Y} - \hat{\sigma} \times t_{n-1, 1-\alpha/2} \sqrt{n} ; \bar{Y} + \hat{\sigma} \times t_{n-1, 1-\alpha/2} \sqrt{n} \right)
\]

**Prediction Intervals**: L-level prediction interval (PI) for a new value of the process \( Y \) is defined by:
\[
(\hat{Y}_{\text{new}} - \hat{\sigma} \times t_{n-1, 1-\alpha/2} ; \hat{Y}_{\text{new}} + \hat{\sigma} \times t_{n-1, 1-\alpha/2})
\]
where the predicted value \( \hat{Y}_{\text{new}} = \bar{Y} \), is obtained as an estimator of the unknown process mean \( \mu \).
Example – Carbon content in Steel

Percentage of $C$ (Carbon) in 2 random samples taken from 2 steel shipments are measured and summarized below. The question is to determine if there are statistically significant differences between the shipments.

<table>
<thead>
<tr>
<th>#</th>
<th>N</th>
<th>$Y_-$</th>
<th>$S^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>3.62</td>
<td>0.086</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>3.18</td>
<td>0.082</td>
</tr>
</tbody>
</table>

Measuring the distance between the true-value and the estimate in terms of the SE’s

- Intuitive criterion: Estimate is credible if it’s not far-away from its hypothesized true-value!
- But how far is far-away?
- Compute the distance in standard-terms: $T = \frac{\text{Estimator} – \text{TrueParameterValue}}{\text{SE}}$
- Reason is that the distribution of $T$ is known in some cases (Student’s $t$, or $N(0,1)$).
- The estimator (obs-value) is typical/atypical if it is close to the center/tail of the distribution.

Comparing CI’s and significance tests

- These are different methods for coping with the uncertainty about the true value of a parameter caused by the sampling variation in estimates.
- **Confidence interval**: A fixed level of confidence is chosen. We determine a range of possible values for the parameter that are consistent with the data (at the chosen confidence level).
- **Significance test**: Only one possible value for the parameter, called the hypothesized value, is tested against the data. We determine the strength of the evidence (confidence) provided by the data against the proposition that the hypothesized value is the true value.

Review

- Are the carbon contents in the two steel shipments any different?

$$t_0 = \frac{\text{Est}_1 - \text{Est}_2 - 0}{SE} = \frac{3.62 - 3.18}{SE(\hat{\mu}_1 - \hat{\mu}_2)} = \frac{0.44}{\sqrt{\frac{0.086}{10} + \frac{0.082}{8}}} = 3.12$$

$$df = 7, \alpha = 0.025$$

Hypotheses

**Guiding principles**

We cannot rule in a hypothesized value for a parameter, we can only determine whether there is evidence, provided by the data, to rule out a hypothesized value.

The *null hypothesis* tested is typically a skeptical reaction to a *research hypothesis*

The t-test

Using $\hat{\theta}$ to test $H_0: \theta = \theta_0$ versus some alternative $H_1$.

**STEP 1** Calculate the test statistic:

$$t_0 = \frac{\hat{\theta} - \theta_0}{ SE(\hat{\theta}) } = \frac{\text{estimate} - \text{hypothesized value}}{\text{standard error}}$$

[This tells us how many standard errors the estimate is above the hypothesized value ($t_0$ positive) or below the hypothesized value ($t_0$ negative).]

**STEP 2** Calculate the $P$-value using the following table.

**STEP 3** Interpret the $P$-value in the context of the data.
Alternative Evidence against $H_0$: $\theta > \theta_0$

- $H_1$: $\theta$ too much bigger than $\theta_0$
  - $P = \text{pr}(T \geq t_0)$

- $H_1$: $\theta$ too much smaller than $\theta_0$
  - $P = \text{pr}(T \leq t_0)$

- $H_1$: $\theta$ too far from $\theta_0$
  - $P = 2 \text{pr}(|t| \geq |t_0|)$ where $T \sim \text{Student}(df)$

Is a second child gender influenced by the gender of the first child, in families with >1 kid?

<table>
<thead>
<tr>
<th>1st Child</th>
<th>Second Child Gender</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>3,202</td>
<td>5,978</td>
</tr>
<tr>
<td>Female</td>
<td>2,776</td>
<td>5,412</td>
</tr>
<tr>
<td>Total</td>
<td>5,978</td>
<td>5,412</td>
</tr>
</tbody>
</table>

 Mothers whose 1st child is a girl are more likely to have a girl, as a second child, compared to mothers with boys as 1st children.

Data: 20 yrs of birth records of 1 Hospital in Auckland, NZ.

Analysis of the birth-gender data

Samples are large enough to use Normal-approx.

Since the two proportions come from totally different mothers they are independent.

$P = \text{pr}(T \geq t_0) = 5.49986 = \frac{\hat{p}_1 - \hat{p}_2 - 0}{SE} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{SE(\hat{\beta}_1 - \hat{\beta}_2)} = \frac{\hat{\beta}_1 (1 - \hat{p}_1) + \hat{\beta}_2 (1 - \hat{p}_2)}{\hat{\beta}_1 + \hat{\beta}_2}$

$P = \text{pr}(T \geq t_0) = 1.9 \times 10^{-8}$

Analysis of the birth-gender data

- We have strong evidence to reject the $H_0$ and hence conclude mothers with first child a girl are more likely to have a girl as a second child.

- **Practical vs. Statistical significance:**
  - How much more likely?
  - 95% CI: $	ext{Cl}(\hat{p}_1; \hat{p}_2) = [0.033; 0.070]$. And computed by:
    - $\hat{p} = \hat{\beta}_1 + 1.96 \times \text{SE}(\hat{\beta}_1 - \hat{\beta}_2)$
    - $0.0515 \pm 1.96 \times 0.0093677 = [3\%; 7\%]$