There are three kinds of lies: lies, damn lies, and statistics

--Benjamin Disraeli, Former British Prime Minister
Tutorial outline

• Definitions

• Null and Alternative Hypotheses

• Central Limit Theorem (CLT)

• $P$-Value, Significance Level and Power

• Confusion Matrix

• Useful inequalities
Definitions

- **Population** ≡ all possible values
- **Sample** ≡ a portion of the population
- **Statistical inference** ≡ generalizing from a sample to a population with calculated degree of certainty
- Two forms of statistical inference
  - Hypothesis testing (Today)
  - Estimation (previous tutorial)
- **Parameter** ≡ a characteristic of population, e.g., population mean \( \mu \)
- **Statistic** ≡ calculated from data in the sample, e.g., sample mean \( \bar{X} \)
## Distinctions Between Parameters and Statistics

<table>
<thead>
<tr>
<th>Source</th>
<th>Parameters</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>Population</td>
<td>Sample</td>
</tr>
<tr>
<td>Random</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Calculated</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Hypothesis Testing Steps

The procedure is broken into four steps:

1. Null and alternative hypotheses
2. Test statistic
3. P-value and interpretation
4. Significance Level
Hypothesis Testing

- Confront two competing theories
  - **NULL** hypothesis $H_0$
    - any observed deviation from what we expect to see is due to chance variability
  - **ALTERNATIVE** hypothesis $H_a$
    - ‘claim’, or theory you wish to test

- The null hypothesis, $H_0$, is assumed true, until enough evidence goes against it
  - We then refute it and believe the alternative, $H_a$
Hypothesis Testing - cont.

- Null hypothesis $H_0: \mu = 170$ ("no difference")

- Alternative hypothesis $H_a: \mu > 170$
Test Statistic

- A measure of how far the observed data is from what is expected assuming the null hypothesis $H_0$
  - Compute the value of a test statistic (TS) from the data

- The particular TS computed depends on the tested parameter
  - For example, to test the population mean, the TS is the sample mean (or standardized sample mean)

- The null hypothesis, $H_0$, is rejected if the TS falls in a user-specified rejection region
Test Statistics - cont.

This is an example of a mean test when $\sigma$ is known. Use this statistic to test the problem:

$Z_{\text{stat}} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$

where $\mu_0 \equiv$ population mean assuming $H_0$ is true

and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
For the illustrative example, $\mu_0 = 170$

We know $\sigma = 40$

Take sample size of $n = 64$. Therefore

If we found a sample mean of 173, then

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{64}} = 5$$

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{173 - 170}{5} = 0.60$$
If we found a sample mean of 185, then

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{185 - 170}{5} = 3.00$$
The Central Limit Theorem (CLT)

If $X_1, X_2, \ldots, X_n$ is a random sample of size $n$ taken from a population with mean $\mu$ and variance $\sigma^2$, and if $\bar{X}$ is the sample mean, the limiting form of the distribution of

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

as $n \rightarrow \infty$, is the standard normal distribution.

- CLT Demo
Reasoning Behind $z_{\text{stat}}$

Sampling distribution of $\bar{x}$ under $H_0$: $\mu = 170$ for $n = 64 \Rightarrow \bar{x} \sim N(170, 5)$
**P-value**

- The *P*-value answers the question: What is the probability of the observed test statistic or one more extreme when $H_0$ is true?

- This corresponds to the AUC in the tail of the Standard Normal distribution beyond the $z_{\text{stat}}$.

- Convert $z$ statistics to *P*-value :
  
  For $H_a: \mu > \mu_0 \Rightarrow P = \Pr(Z > z_{\text{stat}}) = \text{right-tail beyond } z_{\text{stat}}$
P-value for $z_{stat}$ of 0.6

Distribution of $\bar{x}$ and $z_{stat}$ if $H_0$ were true

$P = 0.2743$

(Area under curve)
P-value for $z_{stat}$ of 3.0
Interpretation

- $P$-value answer the question: What is the probability of the observed test statistic … when $H_0$ is true?

- Thus, smaller and smaller $P$-values provide stronger and stronger evidence against $H_0$

- Small $P$-value $\Rightarrow$ strong evidence
Significance level ($\alpha$) and $p$-value

- **Significance level ($\alpha$)**
  The degree of certainty required in order to *reject* the null hypothesis

- A TS with a $p$-value less than some pre-specified false positive (or size) is said to be statistically significant at that level

<table>
<thead>
<tr>
<th>P-value</th>
<th>Wording</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;0.05</td>
<td>Not significant</td>
</tr>
<tr>
<td>0.01 to 0.05</td>
<td>Significant</td>
</tr>
<tr>
<td>0.001 to 0.01</td>
<td>Very significant</td>
</tr>
<tr>
<td>&lt; 0.001</td>
<td>Extremely significant</td>
</tr>
</tbody>
</table>

*commonly used p-values*
Error types

Type I Error

You’re pregnant!

False Positive

Type II Error

You’re not pregnant!

False Negative
Confusion Matrix

- **Type I error or false positive**
  - The chance of rejecting a NULL which is true is \( \alpha \)

- **Type II error or false negative**
  - The chance of not rejecting a NULL which is false is \( \beta \)

- “rightfully” accept NULL
  - this is just \( 1 - \alpha \)

- “rightfully” reject the NULL
  - this is just \( 1 - \beta \), also called power
Confusion Matrix – cont.

<table>
<thead>
<tr>
<th>Decision</th>
<th>H₀ is in fact</th>
<th>False</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject H₀</td>
<td>Power, 1 − 𝛽</td>
<td>😊</td>
<td>Type I error, 𝛼</td>
</tr>
<tr>
<td>Accept H₀</td>
<td>Type II error, 𝛽</td>
<td>😊</td>
<td>1 − 𝛼</td>
</tr>
</tbody>
</table>
Useful Inequalities

- **Markov Inequality**
  For a non-negative R.V. $X \geq 0$ and for any positive number $\lambda > 0$
  \[
  \Pr(X \geq \lambda) \leq \frac{E[X]}{\lambda}
  \]

- **Chebyshev Inequality**
  For any R.V. $X \geq 0$ and for any positive number $\lambda > 0$
  \[
  \Pr(|X - E[X]| \geq \lambda) \leq \frac{Var(X)}{\lambda^2}
  \]

- **Chernoff Inequality**
  Let $X_1, \ldots, X_m$ be a sequence $m$ i.i.d Bernoulli trials
  \[
  \Pr\left(\left|\frac{1}{m}\sum X_i - E[X]\right| \geq \lambda\right) \leq e^{-\frac{\lambda^2 m}{2}}
  \]