BASICS OF STATISTICS
Outline

- Sampling
- Estimators, bias, consistency, and mean squared error
- Law of Large Numbers
- Central Limit Theorem
- Hypothesis testing
What is statistics about?

- Statistics is concerned with data that are subject to random variations
- Collecting data through sampling,
- Summarizing and analyzing data by estimating the parameters of the underlying distribution(s)

**Example:**

- **Population:** the heights of a person
- **Sample:** a set of measured heights

<table>
<thead>
<tr>
<th>Height of a Person (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.75</td>
</tr>
<tr>
<td>1.92</td>
</tr>
<tr>
<td>1.69</td>
</tr>
<tr>
<td>1.80</td>
</tr>
</tbody>
</table>

**Estimate mean:** \( \hat{\mu} \approx E(X) \)

**Estimate variance:** \( \hat{\sigma}^2 \approx Var(X) \)
Sampling

Population

- **Convenience sample**
  - Only data that is easy to accessible is sampled
  - Biased

- **Systematic sample**
  - Ordered sampling: randomly select first element and then select every $k$’th element
  - Biased

- **Random sample**
  - Every element has equal probability of being selected
  - Unbiased

- **Stratified sample**
  - When there is variance in subpopulations, it is better to sample each population independently (sample size should reflect the proportion of the subpopulation)
  - Unbiased
Estimators

- **Definition**: An estimator is a function that uses input from the sample space to estimate a parameter of the underlying data distribution.

- **Examples**: Let \( x_1, \ldots, x_n \) be the values of i.i.d. random variables \( X_i \).

  - **Empirical mean** and the **sample mean**: \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \)

  - **Empirical variance**: \( S_{em}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2 \)

  - **Sample variance**: \( S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2 \)
Covariance estimators and correlation

- Let \((x_1, y_1), \ldots, (x_n, y_n)\) be samples of i.i.d. random variables \(X_i, Y_i\)

- **Empirical covariance:** 
  \[
  \hat{C}_{em} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y})
  \]

- **Sample covariance:** 
  \[
  \hat{C} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y})
  \]

- **Correlation:** 
  \[
  r = \frac{\hat{C}}{S_X S_Y}
  \]

  For linear dependency between two variables, e.g., \(Y = aX + b\):

  \[
  r = \begin{cases} 
  1, & a > 0 \\
  -1, & a < 0 
  \end{cases}
  \]
Let $x_1, \ldots, x_n$ be the values of i.i.d. random variables $X_i$

**Empirical distribution function:**

$$\hat{F}_{X_{i:n}}(x) = \frac{1}{n} \sum_{i=1}^{n} [x_i \leq x]$$

where $[x_i \leq x] := \begin{cases} 1, & x_i \leq x \\ 0, & x_i > x \end{cases}$ is called the indicator function

**Empirical median** $\hat{x}_{med}$ is defined as $\hat{F}_{X_{i:n}}(\hat{x}_{med}) = \frac{1}{2}$, that is, for ordered $x_{i_1} \leq \cdots \leq x_{i_n}$:

$$\hat{x}_{med} = \begin{cases} x_{(i(n+1)/2)} & \text{for odd } n \\ \frac{x_{i_{n/2}} + x_{(i(n+2)/2)}}{2} & \text{for even } n \end{cases}$$
Example

- What is the expected life time of a specific electronic device (in months)?
- Random variable $X$: life time in # months
- Random sample:
  \[ x_1 = 38, \quad x_2 = 33, \quad x_3 = 35, \quad x_4 = 32, \quad x_5 = 9, \quad x_6 = 36, \quad x_7 = 31, \]
  \[ x_8 = 37, \quad x_9 = 22, \quad x_{10} = 40, \quad x_{11} = 30 \]

- **Empirical mean:** $\bar{X} = \frac{1}{11} \sum_{i=1}^{11} x_i \approx 31.2$

- **Empirical median:** 33

- **Empirical variance:** $S_{em}^2 = \frac{1}{11} \sum_{i=1}^{11} (x_i - \bar{X})^2 \approx 70.69$

- **Sample variance:** $S^2 = \frac{1}{10} \sum_{i=1}^{11} (x_i - \bar{X})^2 \approx 77.76$
How “good” is an estimator?

How well does it approximate the true parameter on average?

Can it yield the true parameter with more and more data?

What is the variance of the estimator?

Definition: An estimator $\hat{\gamma}$ is **unbiased** if its expected value $E(\hat{\gamma})$ is equal to the true value of the parameter $\gamma$ it estimates, i.e., $E(\hat{\gamma}) = \gamma$, otherwise $\hat{\gamma}$ is biased with squared bias $E(\hat{\gamma} - \gamma)^2$

Definition: An estimator $\hat{\gamma}$ derived from $n$ values of i.i.d. random variables $X_i$ is **consistent** if $\lim_{n \to \infty} P(|\hat{\gamma} - \gamma| > \varepsilon) = 0$ for all $\varepsilon > 0$
Mean and variance estimators

Let \( x_1, \ldots, x_n \) be the values of i.i.d. random variables \( X_i \)

Theorem

- The empirical mean \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \) is an unbiased consistent estimator of the true mean \( E(X) \)

- The empirical variance \( S_{\text{em}}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2 \) is a biased consistent estimator of the true variance \( Var(X) \), it can be shown that \( E(S_{\text{em}}^2) = \frac{n-1}{n} Var(X) \)

- The sample variance \( S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2 \) is an unbiased consistent estimator of the true variance \( Var(X) \)

- The sample covariance \( \hat{C} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y}) \) is an unbiased consistent estimator of \( Cov(X) \)

- The empirical distribution function \( \hat{F}_{X;n}(x) = \frac{1}{n} \sum_{i=1}^{n} [x_i \leq x] \) is an unbiased consistent estimator of the true cumulative distribution \( F_X \)
Law of Large Numbers

- Let $x_1, x_2, \ldots, x_n$ be a random sample from a distr. $f_X(x)$ and $\bar{X} = \frac{\sum_i X_i}{n}$

- **Weak law of large numbers (weak consistency of the empirical mean)**
  - $\lim_{n \to \infty} P(|\bar{X} - E(X)| > \varepsilon) = 0$ for all $\varepsilon > 0$
  - Sample average **converges in probability** towards the mean of the distr. of $X$

- **Strong law of large numbers (strong consistency of the empirical mean)**
  - $P(\lim_{n \to \infty} |\bar{X} - E(X)| > \varepsilon) = 0$ for all $\varepsilon > 0$
  - Sample average **converges almost surely** towards the mean of the distr. of $X$
Best estimators

- Definition: An unbiased estimator \( \hat{\gamma} \) is the best estimator of the true parameter \( \gamma \) if it has lowest variance among all other unbiased estimators, i.e., for all unbiased estimators \( \hat{\gamma}' \) of \( \gamma \): \( Var(\hat{\gamma}) \leq Var(\hat{\gamma}') \)

- The mean squared error between an estimator \( \hat{\gamma} \) and \( \gamma \) is:

\[
mse(\hat{\gamma} - \gamma) = E((\hat{\gamma} - \gamma)^2) = Var(\hat{\gamma}) + Bias(\hat{\gamma})^2
\]

Because:

\[
Var(\hat{\gamma}) = Var(\hat{\gamma} - \gamma) = E((\hat{\gamma} - \gamma)^2) - E^2(\hat{\gamma} - \gamma)
\]

\[
Bias(\hat{\gamma}) = E(\hat{\gamma}) - \gamma = E(\hat{\gamma}) - E(\gamma) = E(\hat{\gamma} - \gamma)
\]

- Notes
  - The sample mean is the best estimator of the true mean for many useful distributions
  - The sample variance is the best estimator of the true variance for normally distributed data
Usefulness of estimators

- How useful is an estimator for the understanding of the underlying distribution?

- It depends on the distribution!

- Example
  - Random variable $X :=$ yearly income in $\$1000$
  - Random sample:
    $x_1 = 58; x_2 = 74; x_3 = 69; x_4 = 81; x_5 = 64; x_6 = 120; x_7 = 55;$
    $x_8 = 71; x_9 = 77; x_{10} = 65; x_{11} = 23,000 \Rightarrow \bar{X} \approx 2,158$

- Empirical median is more insightful in this case
Other useful estimators

- Maximum Likelihood Estimator, i.e., $\arg\max_\theta P(x_1, \ldots, x_n | \theta)$
  - Consistent
  - Asymptotically normal
  - Asymptotically optimal, i.e., with smallest variance

- $\min(x_1, \ldots, x_n)$

- $\max(x_1, \ldots, x_n)$

- Empirical skewness

$$Sk = \frac{1}{n} \frac{\sum_{i=1}^{n} (x_i - \bar{X})^3}{\left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2\right)^{3/2}}$$

Recap of the normal distribution

- $X$ is normally distributed $\iff X \sim N(\mu, \sigma^2) \iff f_X(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
  
  $\mu$: mean, $\sigma$: standard deviation

- Standard normal distribution: $N(0,1)$

- Cumulative distribution of $N(0,1)$: $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}} dx$

- Theorem: If $X \sim N(\mu, \sigma^2)$ then $Y := \frac{X-\mu}{\sigma} \sim N(0,1)$
Central Limit Theorem

- **Central Limit Theorem**: Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from a distr. with mean $\mu$ and finite non-zero variance $\sigma^2$. The cdf of the random variable $Z := \sum_i X_i$ converges to the cdf of the normal distribution $N(n\mu, n\sigma^2)$. That is:

$$\lim_{n \to \infty} P\left( a \leq \frac{Z - n\mu}{\sqrt{n}\sigma} \leq b \right) = \Phi(b) - \Phi(a)$$

- **Corollary**: The cdf of $Z := \frac{1}{n} \sum_i X_i$ converges to the cdf of $N\left(\mu, \frac{\sigma^2}{n}\right)$.
Central Limit Theorem: Example

\( f_X(x) \) is uniform

Avg. of \( X_1, X_2 \) sampled repeatedly from \( f_X(x) \)

Avg. of \( X_1, X_2, X_3, X_4 \) sampled repeatedly from \( f_X(x) \)
Empirical evidence for the Central Limit Theorem (by considering sequences of i.i.d. Bernoulli variables) and for the Law of Large Numbers (by considering random samples from a Binomial distribution)
Hypothesis testing

- Example hypotheses:
  - Sample originates from normal distribution
  - Two random variables are independent
  - Sample is Bernoulli distributed with p=0.5

- Goal: Falsification of hypothesis by lack of statistical evidence

- Hypothesis to be falsified: $H_0$ (null hypothesis)
- Counter hypothesis: $H_1$
- Test region $R$ from cdf of test variable $X$
  - $X \in R \Rightarrow$ reject $H_0$
  - $X \notin R \Rightarrow$ retain $H_0$

<table>
<thead>
<tr>
<th></th>
<th>Retain $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ true</td>
<td>ok</td>
<td>Type I error</td>
</tr>
<tr>
<td>$H_1$ true</td>
<td>Type II error</td>
<td>ok</td>
</tr>
</tbody>
</table>
Hypothesis testing: Example

Assume average IQ of students is 100
\( H_0: \mu = 100 \)

Run IQ test on sample

\( \bar{IQ} = 115 \)

Is this likely given \( \mu = 100 \)?

If yes retain \( H_0 \) else reject
**How well is a parameter estimated**

- Consider estimator $\hat{\theta}$ for parameter $\theta$
- How well does $\hat{\theta}$ represent $\theta$?

\[
P(\hat{\theta} - c \leq \theta \leq \hat{\theta} + c) = 1 - \alpha
\]

**Definitions**

- The interval $[\hat{\theta} - c, \hat{\theta} + c]$ is the **confidence interval**
- The value $1 - \alpha$ is the **confidence level**
- $\alpha$ is the **significance level** (typically: 0.01, 0.05, 0.1)

![Diagram](https://via.placeholder.com/150)

- $\hat{\theta}$ here: reject
- $\hat{\theta}$ here: retain

**Critical region**
One sided and two-sided tests

- A test of the form $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$ is called a **two-sided test**.

- A test of either of these forms
  - $H_0: \theta \leq \theta_0$ vs. $H_1: \theta > \theta_0$
  - $H_0: \theta \geq \theta_0$ vs. $H_1: \theta < \theta_0$

  is called a **one-sided test**.
Consider i.i.d. random variables $X_1, \ldots, X_n$, $n \gg 1$, from a distribution with unknown non-zero mean $\mu$ and known finite variance $\sigma^2$.

- We know $\bar{X} = \frac{1}{n} \sum_i X_i$ is approximately normally distributed with $N(\mu, \frac{\sigma^2}{n})$.

- We also know that $Y = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0, 1)$.

$$P \left( -z \leq \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \leq z \right) = \Phi(z) - \Phi(-z) = P \left( \bar{X} - \frac{z\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{z\sigma}{\sqrt{n}} \right)$$

⇒ For confidence interval $[\bar{X} - c, \bar{X} + c]$ set $z := \frac{c\sqrt{n}}{\sigma}$ and look up $\Phi(z)$.

⇒ For confidence level $1 - \alpha$, and a proposed value for $\mu$, reject null hypothesis if $|Y| > \Phi^{-1}(1 - \alpha/2)$.

Definition: The p-value is minimal significance level at which $H_0$ can be rejected.
# Z-score table

This table provides the area between the mean and some Z score. For example, when Z score = 1.45 the area = 0.4265.

<table>
<thead>
<tr>
<th>Z</th>
<th>0.000</th>
<th>0.001</th>
<th>0.002</th>
<th>0.003</th>
<th>0.004</th>
<th>0.005</th>
<th>0.006</th>
<th>0.007</th>
<th>0.008</th>
<th>0.009</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.000</td>
<td>0.004</td>
<td>0.008</td>
<td>0.012</td>
<td>0.016</td>
<td>0.019</td>
<td>0.023</td>
<td>0.027</td>
<td>0.031</td>
<td>0.035</td>
</tr>
<tr>
<td>0.1</td>
<td>0.039</td>
<td>0.043</td>
<td>0.048</td>
<td>0.051</td>
<td>0.055</td>
<td>0.057</td>
<td>0.059</td>
<td>0.063</td>
<td>0.067</td>
<td>0.071</td>
</tr>
<tr>
<td>0.2</td>
<td>0.079</td>
<td>0.083</td>
<td>0.087</td>
<td>0.091</td>
<td>0.094</td>
<td>0.097</td>
<td>0.102</td>
<td>0.106</td>
<td>0.110</td>
<td>0.114</td>
</tr>
<tr>
<td>0.3</td>
<td>0.117</td>
<td>0.121</td>
<td>0.125</td>
<td>0.129</td>
<td>0.133</td>
<td>0.136</td>
<td>0.140</td>
<td>0.144</td>
<td>0.148</td>
<td>0.152</td>
</tr>
<tr>
<td>0.4</td>
<td>0.155</td>
<td>0.159</td>
<td>0.162</td>
<td>0.164</td>
<td>0.170</td>
<td>0.173</td>
<td>0.177</td>
<td>0.180</td>
<td>0.184</td>
<td>0.189</td>
</tr>
<tr>
<td>0.5</td>
<td>0.191</td>
<td>0.195</td>
<td>0.198</td>
<td>0.201</td>
<td>0.205</td>
<td>0.208</td>
<td>0.213</td>
<td>0.217</td>
<td>0.220</td>
<td>0.224</td>
</tr>
<tr>
<td>0.6</td>
<td>0.225</td>
<td>0.229</td>
<td>0.234</td>
<td>0.237</td>
<td>0.239</td>
<td>0.242</td>
<td>0.245</td>
<td>0.248</td>
<td>0.252</td>
<td>0.256</td>
</tr>
<tr>
<td>0.7</td>
<td>0.258</td>
<td>0.261</td>
<td>0.264</td>
<td>0.267</td>
<td>0.270</td>
<td>0.274</td>
<td>0.276</td>
<td>0.279</td>
<td>0.282</td>
<td>0.285</td>
</tr>
<tr>
<td>0.8</td>
<td>0.288</td>
<td>0.291</td>
<td>0.293</td>
<td>0.296</td>
<td>0.299</td>
<td>0.302</td>
<td>0.305</td>
<td>0.308</td>
<td>0.310</td>
<td>0.313</td>
</tr>
<tr>
<td>0.9</td>
<td>0.315</td>
<td>0.318</td>
<td>0.321</td>
<td>0.323</td>
<td>0.326</td>
<td>0.328</td>
<td>0.331</td>
<td>0.334</td>
<td>0.336</td>
<td>0.339</td>
</tr>
<tr>
<td>1.0</td>
<td>0.341</td>
<td>0.343</td>
<td>0.346</td>
<td>0.348</td>
<td>0.350</td>
<td>0.353</td>
<td>0.355</td>
<td>0.357</td>
<td>0.359</td>
<td>0.362</td>
</tr>
<tr>
<td>1.1</td>
<td>0.364</td>
<td>0.366</td>
<td>0.368</td>
<td>0.370</td>
<td>0.372</td>
<td>0.374</td>
<td>0.377</td>
<td>0.379</td>
<td>0.381</td>
<td>0.383</td>
</tr>
<tr>
<td>1.2</td>
<td>0.384</td>
<td>0.386</td>
<td>0.388</td>
<td>0.390</td>
<td>0.392</td>
<td>0.394</td>
<td>0.396</td>
<td>0.398</td>
<td>0.399</td>
<td>0.401</td>
</tr>
<tr>
<td>1.3</td>
<td>0.403</td>
<td>0.404</td>
<td>0.406</td>
<td>0.408</td>
<td>0.409</td>
<td>0.411</td>
<td>0.413</td>
<td>0.414</td>
<td>0.416</td>
<td>0.417</td>
</tr>
<tr>
<td>1.4</td>
<td>0.419</td>
<td>0.420</td>
<td>0.422</td>
<td>0.423</td>
<td>0.425</td>
<td>0.426</td>
<td>0.427</td>
<td>0.429</td>
<td>0.430</td>
<td>0.431</td>
</tr>
<tr>
<td>1.5</td>
<td>0.433</td>
<td>0.434</td>
<td>0.435</td>
<td>0.437</td>
<td>0.438</td>
<td>0.439</td>
<td>0.440</td>
<td>0.441</td>
<td>0.442</td>
<td>0.444</td>
</tr>
<tr>
<td>1.6</td>
<td>0.445</td>
<td>0.446</td>
<td>0.447</td>
<td>0.448</td>
<td>0.449</td>
<td>0.450</td>
<td>0.451</td>
<td>0.452</td>
<td>0.453</td>
<td>0.454</td>
</tr>
<tr>
<td>1.7</td>
<td>0.455</td>
<td>0.456</td>
<td>0.457</td>
<td>0.458</td>
<td>0.459</td>
<td>0.460</td>
<td>0.461</td>
<td>0.462</td>
<td>0.463</td>
<td>0.464</td>
</tr>
<tr>
<td>1.8</td>
<td>0.464</td>
<td>0.464</td>
<td>0.465</td>
<td>0.466</td>
<td>0.467</td>
<td>0.468</td>
<td>0.469</td>
<td>0.470</td>
<td>0.471</td>
<td>0.472</td>
</tr>
<tr>
<td>1.9</td>
<td>0.471</td>
<td>0.471</td>
<td>0.472</td>
<td>0.473</td>
<td>0.474</td>
<td>0.475</td>
<td>0.476</td>
<td>0.476</td>
<td>0.477</td>
<td>0.478</td>
</tr>
<tr>
<td>2.0</td>
<td>0.477</td>
<td>0.478</td>
<td>0.478</td>
<td>0.479</td>
<td>0.479</td>
<td>0.480</td>
<td>0.480</td>
<td>0.481</td>
<td>0.482</td>
<td>0.482</td>
</tr>
<tr>
<td>2.1</td>
<td>0.482</td>
<td>0.482</td>
<td>0.483</td>
<td>0.483</td>
<td>0.484</td>
<td>0.484</td>
<td>0.485</td>
<td>0.485</td>
<td>0.485</td>
<td>0.486</td>
</tr>
<tr>
<td>2.2</td>
<td>0.486</td>
<td>0.486</td>
<td>0.486</td>
<td>0.487</td>
<td>0.487</td>
<td>0.487</td>
<td>0.488</td>
<td>0.488</td>
<td>0.488</td>
<td>0.488</td>
</tr>
<tr>
<td>2.3</td>
<td>0.489</td>
<td>0.489</td>
<td>0.490</td>
<td>0.490</td>
<td>0.491</td>
<td>0.491</td>
<td>0.492</td>
<td>0.492</td>
<td>0.493</td>
<td>0.493</td>
</tr>
<tr>
<td>2.4</td>
<td>0.491</td>
<td>0.492</td>
<td>0.492</td>
<td>0.493</td>
<td>0.493</td>
<td>0.494</td>
<td>0.494</td>
<td>0.495</td>
<td>0.495</td>
<td>0.495</td>
</tr>
<tr>
<td>2.5</td>
<td>0.493</td>
<td>0.494</td>
<td>0.494</td>
<td>0.494</td>
<td>0.495</td>
<td>0.495</td>
<td>0.496</td>
<td>0.496</td>
<td>0.496</td>
<td>0.496</td>
</tr>
<tr>
<td>2.6</td>
<td>0.495</td>
<td>0.495</td>
<td>0.496</td>
<td>0.496</td>
<td>0.497</td>
<td>0.497</td>
<td>0.498</td>
<td>0.498</td>
<td>0.498</td>
<td>0.498</td>
</tr>
<tr>
<td>2.7</td>
<td>0.496</td>
<td>0.496</td>
<td>0.497</td>
<td>0.497</td>
<td>0.498</td>
<td>0.498</td>
<td>0.498</td>
<td>0.499</td>
<td>0.499</td>
<td>0.499</td>
</tr>
<tr>
<td>2.8</td>
<td>0.497</td>
<td>0.497</td>
<td>0.498</td>
<td>0.498</td>
<td>0.499</td>
<td>0.499</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>2.9</td>
<td>0.498</td>
<td>0.498</td>
<td>0.498</td>
<td>0.499</td>
<td>0.499</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>3.0</td>
<td>0.499</td>
<td>0.499</td>
<td>0.499</td>
<td>0.499</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>3.1</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>3.2</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
</tr>
</tbody>
</table>
Wald Test

- For a parameter $\hat{\theta}$ derived from a sample and a proposed parameter $\theta$, we can test

$$H_0: \hat{\theta} = \theta \text{ vs. } H_1: \hat{\theta} \neq \theta$$

- $s = \sqrt{Var(\hat{\theta})}$ is called the standard error and $Var(\hat{\theta})$ is the sample variance

- Test variable $W := \frac{\hat{\theta} - \theta}{s}$ is approximately $N(0,1)$-distributed (i.e., distribution of $W$ converges to $N(0,1)$ for growing sample size)

$\Rightarrow$ Reject $H_0$ at level $\alpha$ when $|W| > \Phi^{-1}(1 - \alpha/2)$

- Example
  - $\hat{\theta}$: Average increase of height of men compared to height of women
  - Proposed parameter $\theta = 0$
Wald test: Example

- What is the expected life time of a specific electronic device (in months)?
- Random variable $X$: life time in # months
- Random sample:
  - $x_1 = 38, x_2 = 33, x_3 = 35, x_4 = 32, x_5 = 9, x_6 = 36, x_7 = 31,$
  - $x_8 = 37, x_9 = 22, x_{10} = 40, x_{11} = 30$

  - Empirical mean: $\bar{X} = \frac{1}{11} \sum_{i=1}^{11} x_i \approx 31.2$
  - Sample variance: $S^2 = \frac{1}{10} \sum_{i=1}^{11} (x_i - \bar{X})^2 \approx 77.76$

- Hypothesis I: Devices have a life time of around 2 years
  \[ W := \frac{\hat{\theta} - \theta}{\sqrt{\text{var}(\hat{\theta})}} \approx 0.82 < 1.96 \text{ (for significance level 0.05)} \]

- Hypothesis II: Devices have a life time of around 1 year
  \[ W := \frac{\hat{\theta} - \theta}{\sqrt{\text{var}(\hat{\theta})}} \approx 2.177 > 1.96 \text{ (for significance level 0.05)} \]
Example: Probability of heads

- $H_0$: coin has head probability $p = p_0$
- $X$: test variable representing #heads in $n$ tosses
- We know that approximately $X \sim N(pn, p(1 - p)n)$

$Y := \frac{(X - pn)}{\sqrt{p(1-p)n}} \sim N(0,1) \Rightarrow$ reject $H_0$ at level $\alpha$ ($= 0.05$) if

$$Y > \Phi^{-1}(1 - \alpha/2) \text{ or } Y < \Phi^{-1}(\alpha/2) \iff |Y| > \Phi^{-1}(1 - \alpha/2)$$
Consider i.i.d. random variables $X_1, \ldots, X_n$, $n \gg 1$, from a distribution with unknown, non-zero mean $\mu$ and unknown variance.

Let $s^2$ be the sample variance. $Y := \frac{(\bar{X} - \mu) \sqrt{n}}{s}$ has a Student’s t distribution with $n - 1$ degrees of freedom.

With analogous derivation as before:

$$P \left( \bar{X} - \frac{t_{n-1, \frac{1-\alpha}{2}} s}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{t_{n-1, \frac{1-\alpha}{2}} s}{\sqrt{n}} \right) = 1 - \alpha$$

⇒ For proposed $\mu$ and significance level $\alpha$, reject null hypothesis if $|Y| > t_{n-1, \frac{1-\alpha}{2}}$.
t-Test in practice

- Compare two prediction algorithms $A$ and $A'$ based on performance on $k$ labeled datasets
- Let $e_1, \ldots, e_k$ and $e_1', \ldots, e_k'$ be the error values (or any performance values), respectively
- Are the error means any different?

- Fact: $\bar{e}$ and $\bar{e}'$ are approximately normally distributed, but we neither know the means nor the variances

- Since $\sigma_e$ and $\sigma_{e'}$ are unknown, we need to use $t$-distribution with $k-1$ degrees of freedom to estimate how close $\mu_e$ and $\mu_{e'}$ are ($H_0: \mu_e = \mu_{e'}$)
- $\bar{d} = \bar{e} - \bar{e}'$ is also $t$-distributed, with $k-1$ degrees of freedom

  \[ \Rightarrow H_0: \bar{d} = 0 \text{ and } Y := \frac{(\bar{d} - 0)\sqrt{k}}{s_d} \text{ is the t-statistics} \]

- Use $t$-distribution table to determine the $t_{k-1, 1-\alpha/2}$ score
- If $t_{k-1, 1-\alpha/2} < |Y|$ reject $H_0$ otherwise retain it
Chi-Square Goodness-of-Fit-Test

Given sample $x_1, \ldots, x_n$ of i.i.d. random variables $X_i$ and absolute frequencies $h_1, \ldots, h_k$ of class $c_j, 1 \leq j \leq k$, we can test

- $H_0$: $X_i$ follow a proposed discrete distribution

- $Z_k := \frac{\sum_{j=1}^{k} (h_j - E(h_j))^2}{E(h_j)}$, with $E(h_j)$ being the expected frequency of class $c_j$ according to the proposed distribution, is $\chi^2$-distributed with $k-1$ degrees of freedom

$\rightarrow$ Reject $H_0$ at test level $\alpha$ (e.g. 0.05) if $Z_k > \chi^2_{k-1,1-\alpha}$
## Chi square distribution table

<table>
<thead>
<tr>
<th>d.f.</th>
<th>$\chi^2_{.25}$</th>
<th>$\chi^2_{.10}$</th>
<th>$\chi^2_{.05}$</th>
<th>$\chi^2_{.025}$</th>
<th>$\chi^2_{.010}$</th>
<th>$\chi^2_{.005}$</th>
<th>$\chi^2_{.001}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.32</td>
<td>2.71</td>
<td>3.84</td>
<td>5.02</td>
<td>6.63</td>
<td>7.88</td>
<td>10.8</td>
</tr>
<tr>
<td>2</td>
<td>2.77</td>
<td>4.61</td>
<td>5.99</td>
<td>7.38</td>
<td>9.21</td>
<td>10.6</td>
<td>13.8</td>
</tr>
<tr>
<td>3</td>
<td>4.11</td>
<td>6.25</td>
<td>7.81</td>
<td>9.35</td>
<td>11.3</td>
<td>12.8</td>
<td>16.3</td>
</tr>
<tr>
<td>4</td>
<td>5.39</td>
<td>7.78</td>
<td>9.49</td>
<td>11.1</td>
<td>13.3</td>
<td>14.9</td>
<td>18.5</td>
</tr>
<tr>
<td>5</td>
<td>6.63</td>
<td>9.24</td>
<td>11.1</td>
<td>12.8</td>
<td>15.1</td>
<td>16.7</td>
<td>20.5</td>
</tr>
<tr>
<td>6</td>
<td>7.84</td>
<td>10.6</td>
<td>12.6</td>
<td>14.4</td>
<td>16.8</td>
<td>18.5</td>
<td>22.5</td>
</tr>
<tr>
<td>7</td>
<td>9.04</td>
<td>12.0</td>
<td>14.1</td>
<td>16.0</td>
<td>18.5</td>
<td>20.3</td>
<td>24.3</td>
</tr>
<tr>
<td>8</td>
<td>10.2</td>
<td>13.4</td>
<td>15.5</td>
<td>17.5</td>
<td>20.1</td>
<td>22.0</td>
<td>26.1</td>
</tr>
<tr>
<td>9</td>
<td>11.4</td>
<td>14.7</td>
<td>16.9</td>
<td>19.0</td>
<td>21.7</td>
<td>23.6</td>
<td>27.9</td>
</tr>
<tr>
<td>10</td>
<td>12.5</td>
<td>16.0</td>
<td>18.3</td>
<td>20.5</td>
<td>23.2</td>
<td>25.2</td>
<td>29.6</td>
</tr>
<tr>
<td>11</td>
<td>13.7</td>
<td>17.3</td>
<td>19.7</td>
<td>21.9</td>
<td>24.7</td>
<td>26.8</td>
<td>31.3</td>
</tr>
<tr>
<td>12</td>
<td>14.8</td>
<td>18.5</td>
<td>21.0</td>
<td>23.3</td>
<td>26.2</td>
<td>28.3</td>
<td>32.9</td>
</tr>
<tr>
<td>13</td>
<td>16.0</td>
<td>19.8</td>
<td>22.4</td>
<td>24.7</td>
<td>27.7</td>
<td>29.8</td>
<td>34.5</td>
</tr>
<tr>
<td>14</td>
<td>17.1</td>
<td>21.1</td>
<td>23.7</td>
<td>26.1</td>
<td>29.1</td>
<td>31.3</td>
<td>36.1</td>
</tr>
<tr>
<td>15</td>
<td>18.2</td>
<td>22.3</td>
<td>25.0</td>
<td>27.5</td>
<td>30.6</td>
<td>32.8</td>
<td>37.7</td>
</tr>
<tr>
<td>16</td>
<td>19.4</td>
<td>23.5</td>
<td>26.3</td>
<td>28.8</td>
<td>32.0</td>
<td>34.3</td>
<td>39.3</td>
</tr>
<tr>
<td>17</td>
<td>20.5</td>
<td>24.8</td>
<td>27.6</td>
<td>30.2</td>
<td>33.4</td>
<td>35.7</td>
<td>40.8</td>
</tr>
<tr>
<td>18</td>
<td>21.6</td>
<td>26.0</td>
<td>28.9</td>
<td>31.5</td>
<td>34.8</td>
<td>37.2</td>
<td>42.3</td>
</tr>
<tr>
<td>19</td>
<td>22.7</td>
<td>27.2</td>
<td>30.1</td>
<td>32.9</td>
<td>36.2</td>
<td>38.6</td>
<td>42.8</td>
</tr>
<tr>
<td>20</td>
<td>23.8</td>
<td>28.4</td>
<td>31.4</td>
<td>34.2</td>
<td>37.6</td>
<td>40.0</td>
<td>45.3</td>
</tr>
<tr>
<td>21</td>
<td>24.9</td>
<td>29.6</td>
<td>32.7</td>
<td>35.5</td>
<td>38.9</td>
<td>41.4</td>
<td>46.8</td>
</tr>
<tr>
<td>22</td>
<td>26.0</td>
<td>30.8</td>
<td>33.9</td>
<td>36.8</td>
<td>40.3</td>
<td>42.8</td>
<td>48.3</td>
</tr>
<tr>
<td>23</td>
<td>27.1</td>
<td>32.0</td>
<td>35.2</td>
<td>38.1</td>
<td>41.6</td>
<td>44.2</td>
<td>49.7</td>
</tr>
<tr>
<td>24</td>
<td>28.2</td>
<td>33.2</td>
<td>36.4</td>
<td>39.4</td>
<td>43.0</td>
<td>45.6</td>
<td>51.2</td>
</tr>
<tr>
<td>25</td>
<td>29.3</td>
<td>34.4</td>
<td>37.7</td>
<td>40.6</td>
<td>44.3</td>
<td>46.9</td>
<td>52.6</td>
</tr>
<tr>
<td>26</td>
<td>30.4</td>
<td>35.6</td>
<td>38.9</td>
<td>41.9</td>
<td>45.6</td>
<td>48.3</td>
<td>54.1</td>
</tr>
<tr>
<td>27</td>
<td>31.5</td>
<td>36.7</td>
<td>40.1</td>
<td>43.2</td>
<td>47.0</td>
<td>49.6</td>
<td>55.5</td>
</tr>
<tr>
<td>28</td>
<td>32.6</td>
<td>37.9</td>
<td>41.3</td>
<td>44.5</td>
<td>48.3</td>
<td>51.0</td>
<td>56.9</td>
</tr>
<tr>
<td>29</td>
<td>33.7</td>
<td>39.1</td>
<td>42.6</td>
<td>45.7</td>
<td>49.6</td>
<td>52.3</td>
<td>58.3</td>
</tr>
</tbody>
</table>
Chi-Square independence test

- $r = \text{number of columns}$
- $m = \text{number of rows}$
- $n_{ij} = \text{Actual number in cell}_{ij}$
- $n^*_{ij} = \text{Expected number in cell}_{ij}$
- $(r - 1)(m - 1) = \text{degrees of freedom}$

<table>
<thead>
<tr>
<th>Feature $X$</th>
<th>Feature $Y$</th>
<th>Sum $\Sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$n_{1j}$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$n_{12}$</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>$n_{1k}$</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>$n_{1r}$</td>
</tr>
<tr>
<td>2</td>
<td>$n_{21}$</td>
<td>$n_{2j}$</td>
</tr>
<tr>
<td></td>
<td>$n_{22}$</td>
<td>$n_{22}$</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>$n_{2k}$</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>$n_{2r}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$j$</td>
<td>$n_{jk}$</td>
<td>$n_{j}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$m$</td>
<td>$n_{m1}$</td>
<td>$n_{mj}$</td>
</tr>
<tr>
<td></td>
<td>$n_{m2}$</td>
<td>$n_{mk}$</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>$n_{mr}$</td>
</tr>
<tr>
<td>Sum $\Sigma$</td>
<td>$n_{1}$</td>
<td>$n_{j}$</td>
</tr>
<tr>
<td></td>
<td>$n_{2}$</td>
<td>$n_{j}$</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>$n_{k}$</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>$n_{r}$</td>
</tr>
<tr>
<td></td>
<td>$\Sigma$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

$\chi^2 = \sum_{j=1}^{m} \sum_{k=1}^{r} \frac{(n_{jk} - n^*_{jk})^2}{n^*_{jk}}$

$\Rightarrow \text{Reject } H_0 \text{ at test level } \alpha \text{ (e.g. } 0.05) \text{ if } \chi^2 > \chi^2_{(r-1)(m-1),1-\alpha}$
General recipe for hypothesis testing

- Formulate null hypothesis
- Define corresponding random variable for the test
- Turn the variable into a $N(0,1)$-distributed variable, or a t-statistics, or a $\chi^2$-statistics, ...
- Test whether the new statistics lies in the critical region of the underlying distribution