PROBABILITY & STATISTICS

TPP Math Review

Manon Revel
mrevel@mit.edu
PROBABILITY

➤ Probability Space
➤ Conditional Probability
➤ Random Variables
➤ Expectation and Variance
➤ Gaussian Distribution
PROBABILITY SPACE
SAMPLE SPACE, SIGMA-FIELD AND PROBABILITY MEASURE

\[(\Omega, \mathcal{F}, P)\]

- Sample space \(\Omega\) ↔ possible outcomes
- \(\mathcal{F}\) ↔ \(\sigma\)-field ↔ collection of events
- \(P\) ↔ probability measure ↔ measures events' mass
A COIN TOSS

\[ \Omega = \{H, T\} \leftarrow \text{two possible outcomes} \]

\[ \mathcal{F} = \{\emptyset, \{H, T\}, \{H\}, \{T\}\} \]

\[ P_1 = \mathcal{U}(\{H, T\}) \]

fair coin

\[ P_2(\{H\}) = \frac{3}{4} \]

biased coin
A COIN TOSS

\( (\Omega, \mathcal{F}, \mathbb{P}_2) \)

\[ \mathbb{P}_2(\emptyset) = 0 \]
\[ \mathbb{P}_2(\{T\}) = 1 - \frac{3}{4} = \frac{1}{4} \]
\[ \mathbb{P}_2(\{H, T\}) = \mathbb{P}(\{H\}) + \mathbb{P}(\{T\}) = 1 \]

\( \text{one gets } H \text{ or } T \)
Let $\mathcal{F}$ be a collection of events, formally, a sigma field.

- $\emptyset \in \mathcal{F}$
- $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
- $A_i \in \mathcal{F} \implies \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$
PROBABILITY MEASURE

➤ Let \(\mathcal{P}\) be a probability measure.

➤ \(\mathcal{P} : \mathcal{F} \rightarrow [0, 1]\)

➤ \(\mathcal{P}(\emptyset) = 0, \mathcal{P}(\Omega) = 1\)

➤ If the \(B_i\) are disjoint events, \(\mathcal{P}(\bigcup_{i \in \mathbb{N}} B_i) = \sum_{i \in \mathbb{N}} \mathcal{P}(B_i)\)
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]
CONDITIONAL PROBABILITY
The Monty Hall Show

➤ You have to choose one box among 9.
➤ 8 boxes hide a goat. 1 box hides one million $.

choose

The host reveals 7 boxes with goats.

should you keep your box or switch?
**CONDITIONAL PROBABILITY**

\[
P(A \mid B) = \frac{P(A \cap B)}{P(B)}
\]

\[
P(A \mid B) = \frac{P(B \cap A)}{P(B)}
\]
BAYES' RULE

\[ P(A \mid B) = P(B \mid A) \frac{P(A)}{P(B)} \]
Two events are independent if the occurrence of not does not influence the occurrence of the other one.

\[ P(A \mid B) = P(A) \]

\[ P(A \cap B) = P(A) \times P(B) \]
LAW OF TOTAL PROBABILITIES

\[ P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A | B_i) \times P(B_i) \]
Interpreting a positive result...

has $\frac{2}{10}$ chances of being sick...

The test gives $\frac{1}{10}$ false negatives and $\frac{1}{3}$ false positives.

tested+. What is the IP that is sick?
COVID TESTING

\[ P(\text{sick} \mid \oplus) = \frac{P(\oplus \mid \text{sick}) \cdot P(\text{sick})}{P(\oplus)} \]

Bayes' Rule

\[ P(\oplus \mid \text{sick}) = 0.9 \quad \text{and} \quad P(\text{sick}) = 0.2 \]

\[ P(\oplus) = P(\oplus \mid \text{sick}) \cdot P(\text{sick}) + P(\oplus \mid \text{not sick}) \cdot P(\text{not sick}) \]

\[ P(\text{sick} \mid \oplus) = \frac{0.9 \times 0.2}{0.9 \times 0.2 + 0.3 \times 0.8} = \frac{18}{42} = \frac{3}{7} = 0.43 \]
RANDOM VARIABLES
Let X be a function

\[ X : \Omega \rightarrow \mathbb{R} \]

such that

\[ \{ w \mid X(w) \leq c \} \in \mathcal{F} \]

\( a \) possible outcome
ROLL TWO DICES

\[ \Omega = \{ \{i,j\} \in \{1,6\}^2 \} \]

\[ \mathcal{F} = \{ \emptyset, \Omega, \{1,1\}, \{1,2\}, \ldots, \} \]

1 \leq J = 2^2

\[ \mathbb{P}(i,j) = \frac{1}{36} \quad \text{← uniform probability measure} \]

\[ X(i,j) = i + j \quad \text{← sum of the die rolls} \]

\[ Y(i,j) = \max(i,j) \quad \text{← maximum die roll} \]
ROLL TWO DICES

\[ P(\{w \mid X(w) \leq 3\}) = P(\{11, 12, 21\}) = P(11) + P(12) + P(21) \]

" \( P(X \leq 3) \) " uniform probability = \( \frac{3}{36} \)

\[ P(\{w \mid Y(w) = 6\}) = P(\{61, 62, 63, 64, 65, 66\}) \]

" \( P(Y = 6) \) " = \( \frac{11}{36} \) (disjoint events + UPP)
ROLL TWO DICES

\[ P(X = 1) = 0 \]
\[ P(X = 2) = \frac{1}{36} \]
\[ P(X = 3) = \frac{2}{36} \]
\[ P(X = 4) = \frac{1}{36} \]
\[ P(X = 5) = \frac{1}{36} \]
\[ P(X = 6) = \frac{1}{36} \]
\[ P(X = 7) = \frac{6}{36} \]
\[ P(X = 8) = \frac{5}{36} \]
\[ P(X = 9) = \frac{4}{36} \]
\[ P(X = 10) = \frac{3}{36} \]
\[ P(X = 11) = \frac{2}{36} \]
\[ P(X = 12) = \frac{1}{36} \]

Probability law of \( X \):

\[ P(X = k) = \frac{6 - |k - 7|}{36} = P_X(k) \]
ROLL TWO DICES

\[ P(Y = 6) = \frac{11}{36} \]
\[ P(Y = 5) = \frac{9}{36} \]
\[ P(Y = 4) = \frac{7}{36} \]
\[ P(Y = 3) = \frac{5}{36} \]
\[ P(Y = 2) = \frac{3}{36} \]
\[ P(Y = 1) = \frac{1}{36} \]

\[ P(Y = k) = \frac{2k - 1}{36} = P_Y(k) \]

Probability law of Y
\[ \mathbb{P}(\{ w \mid X(w) = c \}) = \mathbb{P}(X = c) = \mathbb{P}_X(c) \]
DISCRETE AND CONTINUOUS RANDOM VARIABLES

➤ Discrete Random Variable assumes a **countable** number of distinct values.

\[
\mathbb{P}(X = x)
\]

\[
P(X = 1) = P(X = 2) = \ldots = \frac{1}{6}
\]

➤ Continuous Random Variable assumes values within **intervals**.

\[
\mathbb{P}(X \leq x) = F_X(x) = \int_{-\infty}^{x} f_X(t)dt
\]

\[
P(a < X < b)
\]
## Discrete Random Variables

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Sample Space</th>
<th>Probability Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli</td>
<td>{0, 1}</td>
<td>(P(X = 1) = p)</td>
</tr>
<tr>
<td>Binomial</td>
<td>{1, ..., n}</td>
<td>(P(X = k) = \binom{n}{k}p^k(1-p)^{n-k})</td>
</tr>
<tr>
<td>Poisson</td>
<td>(\mathbb{N})</td>
<td>(P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!})</td>
</tr>
</tbody>
</table>
### Continuous Random Variables

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Sample Space</th>
<th>Probability Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform ( X \sim U(\alpha, \beta) )</td>
<td>([ \alpha, \beta ])</td>
<td>( P(X \leq x) = \frac{x - \alpha}{\beta - \alpha} \mathbb{1}<em>{{ x \in [\alpha, \beta] }} ) ( f_X(x) = \frac{\mathbb{1}</em>{{ x \in [\alpha, \beta] }}}{\beta - \alpha} )</td>
</tr>
<tr>
<td>Exponential ( X \sim \exp(\lambda) )</td>
<td>( \mathbb{R}^+ )</td>
<td>( P(X \leq x) = 1 - e^{-\lambda x} ) ( f_X(x) = \lambda e^{-\lambda x} )</td>
</tr>
<tr>
<td>Gaussian ( X \sim N(\mu, \sigma^2) )</td>
<td>( \mathbb{R} )</td>
<td>( f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left( -\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right) )</td>
</tr>
</tbody>
</table>
EXPECTATIONS AND VARIANCE
Probability to get 1 is $\frac{1}{6}$

Probability to get 2 is $\frac{1}{6}$

... 

Probability to get 6 is $\frac{1}{6}$

$$E[X] = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 3.5$$
EXPECTATION

➤ Discrete Random Variables

\[ E[X] = \sum_{i}^{n} x_i P(X = x_i) \]

\[ E[g(X)] = \sum_{i}^{n} g(x_i) P(X = x_i) \]

➤ Continuous Random Variables

\[ E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx \]

\[ E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx \]
LINEARITY OF EXPECTATION

\[ W = aX + bY + c \]

\[ E[W] = aE[X] + bE[Y] + c \]
\[ \text{Var}(X) = E[(X - E[X])^2] \]

\[ = E\left( X^2 - 2X E(X) + E(X)^2 \right) \]

\[ = E(X^2) - 2E(X)E(X) + E(X)^2 \]

\[ = E(X^2) - E(X)^2 \]

\[ \text{SD}(X) = \sigma_X = \sqrt{\text{Var}(X)} \]
$W = X + b$

$Var(W) = Var(X)$
VARIANCE

\[ W = X + b \]

\[ \text{Var}(W) = a^2 \text{Var}(X) \]

\[ \sigma_W = |a| \sigma_X \]
## DISCRETE RANDOM VARIABLES

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Expectation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli</td>
<td>$p$</td>
<td>$p(1-p)$</td>
</tr>
<tr>
<td>Binomial</td>
<td>$np$</td>
<td>$np(1-p)$</td>
</tr>
<tr>
<td>Poisson</td>
<td>$\lambda$</td>
<td>$\lambda$</td>
</tr>
</tbody>
</table>
# Continuous Random Variables

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Expectation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>$\frac{b+a}{2}$</td>
<td>$\frac{(b-a)^2}{12}$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$\lambda^{-1}$</td>
<td>$\lambda^{-2}$</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$\mu$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>
COVARIANCE

\[ \text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \]

\[ -1 < \text{corr}(X, Y) = \frac{\text{COV}(X, Y)}{\sigma_X \sigma_Y} < 1 \]
\[ W = aX + bY \]

\[ \text{Var}(W) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{COV}(X, Y) \]
GAUSSIAN DISTRIBUTION
NORMAL DISTRIBUTION

➤ The most frequently occurring distribution
➤ Symmetric. Bell-shaped curve.
➤ More likely to take on values close to the mean

\[ N(\mu, \sigma^2) \]
NORMAL DISTRIBUTION

➤ $Z \sim N(0,1)$
NORMAL DISTRIBUTION

\[ P(Z \leq 0.25) = 0.6 \]
NORMAL DISTRIBUTION

\[ P(Z < -2) = P(Z > 2) \]

\[ P(Z > -2) = P(Z < 2) \]
QUANTILES

\[ P(Z < q_\alpha) = 1 - \alpha \]

\[ P(|Z| > q_{\alpha/2}) = \alpha \]
NORMAL DISTRIBUTION

- $Z \sim N(0,1)$  \hspace{1cm} $X = \sigma Z + \mu$

- Find $E[X]$ and $\text{Var}(X)$

- $X = \sigma Z + \mu \sim N(\mu, \sigma^2)$

- $X \sim N(\mu, \sigma^2)$ \hspace{1cm} $Z = (X - \mu)/\sigma = (1/\sigma)X - \mu/\sigma$

- Find $E[Z]$ and $\text{Var}(Z)$

- $Z = (X - \mu)/\sigma \sim N(0,1)$
Let $X_i$ be $n$ iid random variables.

Let $\mu$ and $\sigma^2$ be the expectation and variances of $X$.

**Central Limit Theorem**

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{(d)} \mathcal{N}(0,1)$$
X \sim \mathcal{N}(\mu, \sigma^2)

\Pr(a < X < b) = \Pr\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right)

= \Pr\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right)

\mathcal{N}(0,1)
MULTIVARIATE GAUSSIAN

\[ X = (X_1, X_2, \ldots, X_n) \]

where \( X_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \)

such that \( \mathbf{a} \in \mathbb{R}^n, a^T X \) is a Gaussian.

\[
\text{cov}(X) = \begin{pmatrix}
\text{cov}(X_i, X_j)
\end{pmatrix}_{1 \leq i, j \leq n} = \\
\begin{pmatrix}
\text{Var}(X_1) & \text{cov}(X_1, X_2) & \cdots \\
\text{cov}(X_2, X_1) & \text{Var}(X_2) & \cdots \\
\vdots & \vdots & \ddots \\
\text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & \text{Var}(X_n)
\end{pmatrix}
\]

\(\uparrow\) positive semi definite!
STATISTICS

➤ Estimation and Estimators
➤ Hypotheses testing
➤ Confidence Intervals
➤ Linear Regression
➤ Examples
ESTIMATION AND ESTIMATORS
An underlying law governs a phenomenon. Experiments allow to uncover that law.
PROBABILISTIC

➤ Recover the parameters of the distribution, for example the expected value and the variance

➤ Use the estimators computed empirically

➤ Test whether the estimator is consistent with a prior hypothesis

EMPIRICAL

➤ Conduct and experiment, compute estimators of the quantities of interest.

\[ E[X] = \frac{X_1 + \ldots + X_n}{n} \]

\[ Var(X) = \frac{(X_1 - E[X])^2 + \ldots + (X_n - E[X])^2}{n - 1} \]
**PROBABILISTIC**

- X model a phenomenon
- X is characterized by a probability distribution
- E[X] is the expected value, computed from the probability

\[ E[X] = \sum_{i}^{n} p_{i}x_{i} \]

- Var(X), COV(X,Y)… are computed from the probability

**EMPIRICAL**

- X₁, X₂, …, Xₙ describe an experiment
- X₁, X₂, …, Xₙ are characterised by experimental values
- E[X] is the empirical expected value, computed from the data

\[ E[X] = \frac{X_{1} + \ldots + X_{n}}{n} \]

- Var(X), COV(X,Y)… are computed from the data
PROBABILISTIC

- Probability to get 1 is $\frac{1}{6}$
- Probability to get 2 is $\frac{1}{6}$
- ...
- Probability to get 6 is $\frac{1}{6}$

$$E[X] = \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

EMPIRICAL

- 1, 1, 2, 3, 3, 4, 4, 5, 5, 6, 6, 6

$$E[X] = \frac{1 + 1 + 2 + 3 + 3 + 4 + 4 + 5 + 5 + 6 + 6 + 6}{12}$$

$$E[X] = \frac{1}{6} + 2 \cdot \frac{1}{12} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{3}{12} = 3.8$$
Let $X_1, X_2, ..., X_n$ be data points from the experiment.

Let's define an estimator for $p$:

$$\hat{p}_n = \frac{\sum_{i=1}^{n} X_i}{n}$$

$E(\hat{p}_n) = E(X_i) = p$

$\hat{p}_n$ is unbiased

$\text{Var}(\hat{p}_n) = \frac{p(1-p)}{n}$

$(\{0,1\}, (\text{Ber}(p))_{p\in[0.2,0.4]} )$
ESTIMATORS

\[ (\mathbb{R}^+, (\cup [a, a + 1])_{a > 0}) \]

- Let \( X_1, X_2, \ldots, X_n \) be data points from the experiment.
- Let's define an estimator for \( a \):

\[
\hat{a}_n = \frac{\sum_{i=1}^{n} X_i}{n}
\]

\[
\mathbb{E}(\hat{a}_n) = \mathbb{E}(X_1) = a + \frac{1}{2}
\]

\( \hat{a}_n \) is biased

\[
\text{Var}(\hat{a}_n) = \frac{1}{12n}
\]
QUADRATIC RISK

➤ Estimator's bias:

\[ \text{bias} = E[\hat{\theta}_n] - \theta \]

➤ Estimator's variance:

\[ \text{variance} = \text{Var}[\hat{\theta}_n] \]

➤ Quadratic risk:

\[ \text{risk} = \text{variance} + \text{bias}^2 \]
EXERCISE

\[ X_1, X_2, \ldots, X_n \sim \text{iid Ber}(p) \]

\[
\hat{p}_n = \frac{\sum_{i=1}^{n} X_i}{n}
\]

\[
\mathbb{E}(\hat{p}_n) = p
\]

\[
\text{Var}(\hat{p}_n) = \frac{p(1-p)}{n}
\]

\[
\hat{p}_n = \frac{X_1 + X_2}{2}
\]

\[
\mathbb{E}(\hat{p}_n) = p
\]

\[
\text{Var}(\hat{p}_n) = \frac{p(1-p)}{2}
\]
HYPOTHESIS TESTING
HYPOTHESIS TESTING — TATUM’S 3 POINT SHOTS

➤ Tatum that he scores 80% at 3 pts. No more, no less.
➤ Brown challenges Tatum… They collect data on Tatum shooting 3 pts.
➤ \( n = 400, \ X_1, X_2, \ldots, X_n \sim^{iid} Ber(p) \)

\[
H_0 : p = 0.8 \\
H_1 : p \neq 0.8
\]
HYPOTHESIS TESTING — TATUM’S 3 POINT SHOTS

➤ Let's build an estimator for the test:

\[ \hat{p}_n = \frac{\sum_{i=1}^{n} X_i}{n} \]

This is an unbiased estimator.

➤ If \( H_0 \) is true, then, by the Central Limit Theorem:

\[ \sqrt{n} \frac{\hat{p}_n - 0.8}{\sqrt{0.8 \times 0.2}} \approx \mathcal{N}(0,1) \]
HYPOTHESIS TESTING — TATUM’S 3 POINT SHOTS

\[ \sqrt{n} \frac{\hat{p}_n - 0.8}{\sqrt{0.8 \times 0.2}} = -1.5 \]

➤ Is it a plausible realisation for a Gaussian?

\[ \mathbb{P}(|Z| > 1.96) = 0.05 \]

\[ \mathbb{P}(|Z| > 1.5) = 0.13 \]
**REJECTION SET**

**Two-sided test**

\[ H_0 : \theta = 0.8 \]

\[ H_1 : \theta \neq 0.8 \]

\[
\left| \frac{\bar{X}_n - 0.8}{\sqrt{0.8 \times 0.2}} \right| > q_{\alpha/2}
\]

If \( Z_n^* \) is not from \( H_0 's \) distribution, \( \bar{X}_n \) is likely not from \( \theta = 0.8 \). Reject \( H_0 \).

**One-sided test**

\[ H_0 : \theta \geq 0.8 \]

\[ H_1 : \theta < 0.8 \]

\[
\sqrt{n} \frac{\bar{X}_n - 0.8}{\sqrt{0.8 \times 0.2}} < q_{\alpha}
\]

If \( Z_n^* \) is unlikely from \( \theta = 0.8 \), \( Z_n^* \) is unlikely from \( H_0 's \) distribution. Reject \( H_0 \).
Hypothesis Testing

$H_0 : \theta = 0.8$

$H_1 : \theta \neq 0.8$

$H_0 : \theta = 0.8$

$H_1 : \theta < 0.8$

$P(|Z| > z^*)$

$P(Z < z^*)$
Type 1 error: Reject $H_0$ when it is true

Type 2 error: Accept $H_0$ when $H_a$ is true
STUDENT TEST

\[ X_1, X_2, \ldots, X_n \sim \text{iid } \mathcal{N}(\mu, \sigma^2) \]

\[ \mu = \bar{X}_n \]

\[ \hat{S}_n = \frac{\sum_{i=1}^{n} (X_i - \hat{X}_n)^2}{n - 1} \]

\[ \text{unbiased estimator} \]
\[ \sqrt{n} \frac{\hat{X}_n - \mu}{\hat{S}_n} \sim t_{n-1} \]

\[ \hat{S}_n = \frac{\sum_{i=1}^{n} (X_i - \hat{X}^n)^2}{n - 1} \]
CONFIDENCE INTERVAL
CONFIDENCE INTERVALS

\[ P(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} < 1.96) = 0.95 \]

\[ P(\bar{X}_n - \frac{1.96\sigma}{\sqrt{n}} < \mu < \bar{X}_n + \frac{1.96\sigma}{\sqrt{n}}) = 0.95 \]
PROBABILISTIC

➤ Probability to get 1 is 1/6
➤ Probability to get 2 is 1/6
➤ ...
➤ Probability to get 6 is 1/6

\[
E[X] = \frac{1}{6} + \frac{1}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = 3.5
\]

EMPIRICAL

➤ 1, 1, 2, 3, 3, 4, 4, 5, 5, 6, 6, 6

\[
E[X] = \frac{1 + 1 + 2 + 3 + 3 + 4 + 4 + 5 + 5 + 6 + 6 + 6}{12} = 3.8
\]

1/4 Confidence Intervals
CONFIDENCE INTERVAL

➤ Empirical result: 3.8
➤ 95% Confidence Interval

\[ CI = [\bar{X}_n - \frac{1.96s}{\sqrt{n}}, \bar{X}_n + \frac{1.96s}{\sqrt{n}}] \]

➤ CI = [2.05, 5.35]
➤ What does 95% interval mean?
CONFIDENCE INTERVAL

- I launch 100 times 12 dices
- 1. I compute CI$_1$
- 2. I compute CI$_2$
- ...
- 100. I compute CI$_{100}$
- My TRUE parameter shall be in at least 95 of these 100 CIs.
CONFIDENCE INTERVAL

3/4 Confidence Intervals
CONFIDENCE INTERVAL

➤ Remember, in real life cases, we do NOT know the TRUE parameter. In fact, the point of the CI is to allow us to estimate this TRUE parameter.

➤ If the hypothesis is that the TRUE parameter is 3.5, it can be verified looking at the CI.

➤ There is a 0.05 risk to make an error.
PROBABILISTIC

➤ I reject the null hypothesis if my observation seems implausible, meaning the probability that it is a Gaussian is less than 5%.

EMPIRICAL

➤ I reject the null hypothesis if my observation does not belong to the confidence interval.
LINEAR REGRESSION
SIMPLE REGRESSION ANALYSIS

➤ Goal: To develop a model that relates two quantities

➤ $x$: Independent (explanatory) variable; quantity sometimes under managerial control

➤ $Y$: Dependent variable; quantity to be predicted — magnitude is determined (in large part) by $x$
EMPIRICAL WORLD

\[ Y_i = \hat{b}_0 + \hat{b}_1 X_i + \hat{e}_i \]

PROBABILISTIC WORLD

\[ Y = b_0 + b_1 X \]

\emph{true equation}

\[ Y = \hat{b}_0 + \hat{b}_1 X \]

\emph{data observations}
LINEAR REGRESSION

\[ y = \hat{b}_0 + \hat{b}_1 x \]

➤ Estimated values! Hypothesis testing? Confidence interval?
PARAMETERS ESTIMATION

\[ y = \hat{b}_0 + \hat{b}_1 x \]

\[ e_i = y_i - \hat{y}_i \]

**Linear Regression**

**Data observations**
PARAMETERS ESTIMATION

➤ Residuals

\[ e_i = y_i - \hat{y}_i \]

➤ Sum squared of residuals

\[
SSE = \sum_{i}^{n} e_i^2 = \sum_{i}^{n} (y_i - \hat{y}_i)^2
\]

➤ Minimize SSE with \( b_0 \) and \( b_1 \)

➤ Wait… where are \( b_0 \) and \( b_1 \)?

\[ \hat{y}_i = b_0 + b_1 x_i \]
PARAMETERS ESTIMATION

\[ SSE = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 \]
\[ y_i = b_0 + b_1 x_i + \epsilon_i \]
What does it mean for $b_1$ to be 0?

Let: $H_0$ (null hypothesis) $b_1 = 0$

How can we reject $H_0$ based on the data?

Compute the Confidence Intervals!

**INTERPRETATION**

- **Reject $H_0$**
  
- **Fail to Reject $H_0$**

- **Reject $H_0$**
EXAMPLE
CANDY BAGS

➤ A candy manufacturer at MIT produces Bertie Bott's Every Flavour Beans. Each jelly bag contains candies that ALL have the same flavor (hence color). Further, the manufacturer claims that each bag contains at least 49 candies.

➤ Among 50 candy bags, the students found that there is between 41 and 55 candies, and 47.76 on average, with a $\sigma = 4.42$.

➤ Do you think that the manufacturer's claim is fair?
CANDY BAGS

\[ n = 50, \ X_1, \ldots, X_n \sim \text{iid} \ X \text{ with } \mathbb{E}(X) = \mu \]

\[ \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = 47.76 \quad \sigma = 4.42 \]

\[ \text{Ho: } \mu \geq 49 \]
\[ \text{H1: } \mu < 49 \]

one-sided test

alternative possibility

claim from manufacturer
CANDY BAGS

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

\[
= -1.98
\]

\[P(Z \leq -1.98) = 0.024\]

\[\hat{p}\] for a gaussian to be more extreme than the test value

Example
CANDY BAGS

The probability that the test value is from the distribution induced by $H_0$ is less than 5%. We then reject $H_0$. 