Statistical Inference and Analysis

Youssef SALMAN

Lebanese University - Faculty of Sciences
Master 1 - Computer Sciences

9 décembre 2025

Outline

Introduction to Statistical Inference

Why Statistical Inference ?

- We rarely observe the entire population in practice.
- We work with a sample and want to generalize to the population.
- Statistical inference provides :
  - **Point estimation**
  - **Confidence intervals**
  - **Hypothesis testing**
  - **Links to ML model evaluation**
- Fundamental for decision-making under uncertainty.

Point Estimation

Population Parameters vs Sample Statistics

Population parameters (unknown) :

- True mean μ
- True proportion p
- Variance σ^2

Sample statistics (computed) :

- Sample mean $\bar{x}$
- Sample proportion $\hat{p}$
- Sample variance s^2

They act as **point estimators** for their respective parameters.

Estimator

Definition

Let θ be an unknown parameter of a population (mean, variance, proportion, etc.). An **estimator** of θ is a statistic, denoted $\hat{\theta}$, defined as a function of the sample :

$$\hat{\theta} = g(X_1, X_2, \ldots, X_n).$$

The numerical value obtained after observing the data is called an **estimate**.

Biased Estimator

An estimator $\hat{\theta}$ is **biased** if

$$\mathbb{E}[\hat{\theta}] \neq \theta.$$

The bias is defined as :

$$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta.$$

Unbiased Estimator

An estimator $\hat{\theta}$ is **unbiased** if

$$\mathbb{E}[\hat{\theta}] = \theta.$$

Estimating Common Parameters

- **Mean :**

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- **Proportion :**

$$\hat{p} = \frac{\text{number of successes}}{n}$$

- **Variance :** (unbiased)

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Example : Average Order Value

- Sample of $n = 50$ order values.
- Sample mean : $\hat{\mu} = 74.29$
- Sample standard deviation : $\hat{\sigma} = 27.68$
- These are point estimates of the population mean AOV and its variability.

Confidence Intervals

Why Confidence Intervals ?

- A point estimate does not express uncertainty.
- A confidence interval gives a **range of plausible values** for the parameter.
- Example :

We are 95% confident that $\mu \in [L, U]$.

What Affects Interval Width ?

- **Confidence level** : higher $\Rightarrow$ wider interval.
- **Sample size** n : larger $n \Rightarrow$ narrower interval.
- **Variability** (standard deviation) : larger $s \Rightarrow$ wider interval.

Confidence Interval for the Mean I

Goal : Estimate the unknown population mean μ using the sample mean $\bar{x}$ and quantify the uncertainty of this estimate.

Case 1 : Population standard deviation σ known

- By the Central Limit Theorem, for large n :

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

- Rearranging this probability statement gives :

$$\mu \in \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right].$$

- Here, $z_{\alpha/2}$ is the standard normal quantile (e.g., 1.96 for 95%).

Case 2 : Population standard deviation σ unknown (typical case)

- Replace σ by the sample standard deviation s.

Confidence Interval for the Mean II

- The correct distribution becomes the t-distribution with $n-1$ degrees of freedom :

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}.$$

- Therefore, the $(1-\alpha)$ confidence interval is :

$$\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}.$$

- The t-distribution accounts for extra uncertainty when estimating σ.

Example : CI for AOV

- From $n = 50$ orders : $\bar{x} = 72.87$, $s = 27.21$
- 95% CI :

$$[65.14,\ 80.61]$$

- Interpretation : the true average order value is likely in this interval.

Hypothesis Testing

Basic Idea

- Start with a default assumption on the population (**null hypothesis** H_0).
- Use sample data to see whether there is strong evidence against H_0.
- Decision is guided by a **p-value**.

Null and Alternative Hypotheses

- **Null hypothesis** H_0 : baseline, “no effect”, often with equality. Example : $\mu = \mu_0$.
- **Alternative hypothesis** H_1 : what we hope/suspect is true. Examples : $\mu \neq \mu_0$, $\mu > \mu_0$, $\mu < \mu_0$.
- **Two-tailed test** : $H_1 : \mu \neq \mu_0$.
- **One-tailed test** : $H_1 : \mu > \mu_0$ or $H_1 : \mu < \mu_0$.

Understanding p-values

- p-value : probability of obtaining a result as extreme as our sample **if H_0 were true**.
- Decision rule with significance level α :
  - If $p \leq \alpha$: reject H_0 (evidence for H_1).
  - If $p > \alpha$: fail to reject H_0.
- Important :
  - p-value is not $\Pr(H_0 \text{ is true})$.
  - Failing to reject H_0 does not prove H_0.

Z-Test for the Mean (σ known)

When to use :

- Large sample ($n \geq 30$) or population standard deviation σ is known.

Hypotheses :

$$H_0 : \mu = \mu_0 \qquad \text{vs} \qquad H_1 : \mu \neq \mu_0 \text{ (or } >, <).$$

Test statistic :

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Decision rule (two-sided) : reject H_0 if $|Z| > z_{\alpha/2}$.

Python (direct Z-test) :

```
from statsmodels.stats.weightstats import ztest

# sample: 1D array of observations
z_stat, p_value = ztest(sample, value=mu0)  # H0: mean = mu0
```

One-Sample t-Test (Mean)

When to use :

- Population standard deviation σ unknown.
- Sample size small or moderate.

Hypotheses :

$$H_0 : \mu = \mu_0 \qquad \text{vs} \qquad H_1 : \mu \neq \mu_0 \text{ (or } >, <).$$

Test statistic :

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}, \qquad t \sim t_{n-1}.$$

Python :

```
from scipy.stats import ttest_1samp

t_stat, p_value = ttest_1samp(sample, popmean=mu0)
```

Two-Sample t-Test (Independent Groups)

When to use :

- Compare means of two **independent** groups.
- Example : control group vs treatment group.

Hypotheses :

$$H_0 : \mu_1 = \mu_2 \qquad \text{vs} \qquad H_1 : \mu_1 \neq \mu_2.$$

Welch's test statistic :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}.$$

Python (Welch's t-test) :

```
from scipy.stats import ttest_ind

t_stat, p_value = ttest_ind(group1, group2, equal_var=False)
```

Paired t-Test

When to use :

- Same subjects measured twice (before/after).
- Paired observations : $(X_{\text{before}}, X_{\text{after}})$.

Let $D = X_{\text{after}} - X_{\text{before}}$.

Hypotheses :

$$H_0 : \mu_D = 0 \qquad \text{vs} \qquad H_1 : \mu_D \neq 0.$$

Test statistic :

$$t = \frac{\bar{D}}{s_D/\sqrt{n}}, \qquad t \sim t_{n-1}.$$

Python :

```
from scipy.stats import ttest_rel

t_stat, p_value = ttest_rel(before, after)
```

Chi-Square Test for a Variance

When to use :

- Test if population variance equals a hypothesized value σ_0^2.

Hypotheses :

$$H_0 : \sigma^2 = \sigma_0^2 \qquad \text{vs} \qquad H_1 : \sigma^2 \neq \sigma_0^2.$$

Test statistic :

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}, \qquad \chi^2 \sim \chi^2_{n-1}.$$

Two-sided test : reject H_0 for extreme values in both tails.

Python (manual using χ^2 distribution) :

```
from scipy.stats import chi2
import numpy as np

chi_stat = (n - 1) * s2 / sigma0**2
p_left  = chi2.cdf(chi_stat, df=n-1)
p_right = chi2.sf(chi_stat, df=n-1)   # sf = 1 - cdf
p_value = 2 * min(p_left, p_right)
```

Chi-Square Test of Independence

When to use :

- Test whether two categorical variables are independent.
- Data in a contingency table (observed counts).

Hypotheses :

H_0 : the two variables are independent vs H_1 : the two variables are associate

Test statistic :

$$\chi^2 = \sum \frac{(O-E)^2}{E},$$

with O = observed counts, E = expected counts under H_0.

Python :

```
from scipy.stats import chi2_contingency
import numpy as np

# table: 2D array of observed counts
chi2_stat, p_value, dof, expected = chi2_contingency(table)
```

One-Way ANOVA

When to use :

- Compare means of $k \geq 3$ independent groups.

Hypotheses :

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \qquad \text{vs} \qquad H_1 : \text{at least one mean differs.}$$

Test statistic :

$$F = \frac{\text{between-group variability}}{\text{within-group variability}},$$

with $F \sim F_{k-1,\ n-k}$ under H_0.

Python :

```
from scipy.stats import f_oneway

F_stat, p_value = f_oneway(group1, group2, group3)
# add more groups as needed
```

Scenarios and ML Links

Scenario 1 : Conversion Rate

- 55 purchases out of 1000 visitors with new checkout.
- $\hat{p} = 0.055$ (5.5%).
- 95% CI : $[0.041,\ 0.069]$.
- Interpretation : plausible range for true conversion rate under new design.

Scenario 2 : Feature Impact

- Old algorithm : $\bar{x}_{old} = 8.3$ videos/week.
- New algorithm : $\bar{x}_{new} = 8.7$ videos/week.
- p-value = 0.007 (one-sided test, $\alpha = 0.05$).
- Conclusion : reject H_0 ; evidence that new algorithm increases engagement.

Scenario 3 : Comparing ML Models

- Accuracy Model A : 88%.
- Accuracy Model B : 90%.
- Hypotheses : $H_0 : \text{Acc}_A = \text{Acc}_B$ vs $H_1 : \text{Acc}_A \neq \text{Acc}_B$.
- p-value from suitable test (e.g. McNemar) = 0.21.
- Conclusion : fail to reject H_0 ; 2% difference is not statistically significant.

Inference for ML Metrics

- **Accuracy** on n examples with estimate $\hat{p}$:

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- **MSE** from errors $e_1, \ldots, e_n$:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} e_i^2, \quad \text{MSE} \pm 1.96\frac{s_e}{\sqrt{n}}$$

- Same logic for precision, recall (proportions).

Inference in Linear Regression

Simple Linear Regression Model

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- β_0 : intercept
- β_1 : slope (effect of X on Y)
- ε : noise with mean 0
- Fitted line : $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

Interpreting the Slope

- Example : apartment rent vs size.
- Model : $\hat{Y} = 320 + 18.5X$ (rent in euros, size in m^2).
- Interpretation :
  - $\hat{\beta}_0 = 320$: baseline rent (for $X = 0$).
  - $\hat{\beta}_1 = 18.5$: each extra m^2 increases expected rent by €18.50.

CI and Test for the Slope

- **Confidence interval** for β_1 :

$$\hat{\beta}_1 \pm t_{\alpha/2,df} \cdot SE(\hat{\beta}_1)$$

- If CI does not include 0 $\Rightarrow$ evidence of linear association.
- **Hypothesis test** :

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0$$

- Small p-value : reject H_0 ; X significantly influences Y.

Python Regression Example (Summary)

- Simulated data : rent vs size.
- Estimated slope : $\hat{\beta}_1 \approx 19.87$ (true slope 20).
- 95% CI : approximately $[19.34,\ 20.40]$.
- p-value for slope $\ll 0.001$.
- Strong evidence that apartment size has a positive effect on rent.