STATISTICS — Part 2

Spread, Distributions & Inference

① INTERQUARTILE RANGE (IQR)

IQR: upper quartile − lower quartile

Position formula: (n+1)/4 from each end

Example with n = 7:

(7+1)/4 = 2 → 2nd from bottom and top

★ *IQR captures the MIDDLE 50% of the data*

Not affected by extreme outliers

② STANDARD DEVIATION σ

Formula: $\sigma = \sqrt{[\Sigma(x - \bar{x})^2 / n]}$

Method:

1. Find the mean $\bar{x}$
2. Headings: x x − $\bar{x}$ $(x - \bar{x})^2$
3. Sum the squares, divide by n, square root

✓ **TIP:** Use the calculator unless asked 'without calculator'

Frequency table: $\sigma = \sqrt{[\Sigma f(x - \bar{x})^2 / \Sigma f]}$

③ DATA DISTRIBUTIONS

Normal (Bell Curve):

Mean = Median = Mode

Right (positive) skew: Mode < Median < Mean

→ large outlier pulls mean RIGHT

Left (negative) skew: Mean < Median < Mode

→ small outlier pulls mean LEFT

★ *Mean follows the tail — long tail right = right skew*

④ EMPIRICAL RULE (68-95-99.7)

For normally distributed data:

- Within 1σ of mean → **68%**
- Within 2σ of mean → **95%**
- Within 3σ of mean → **99.7%**

★ *68 − 95 − 99.7 is the magic sequence*

✓ **TIP:** Rough check: anything beyond 2σ is unusual; beyond 3σ very rare

⑤ Z-SCORES

Z-score: $z = (x - \bar{x}) / \sigma$

Tells you: how many σ's away from the mean

Tables show LEFT of value:

- $P(z \leq a)$: direct lookup
- $P(z \geq a) = 1 - P(z \leq a)$
- $P(z \leq -a) = 1 - P(z \leq a)$ (symmetry)

★ *Always SKETCH the curve and shade what you want!*

⑥ CENTRAL LIMIT THEOREM

As sample size n grows:

- $\bar{x}$ becomes normally distributed
- $\mu_{\bar{x}} = \mu$ (sample mean = population mean)
- $\sigma_{\bar{x}} = \sigma / \sqrt{n}$ (in log tables)

★ *Bigger sample → tighter spread of sample means*

⚠ **WATCH:** Use $\sigma/\sqrt{n}$ NOT just σ — sample mean spread is smaller!

⑦ CONFIDENCE INTERVAL

Population proportion p, sample proportion $\hat{p}$

Standard error: $\sigma_{\hat{p}} = \sqrt{[\hat{p}(1-\hat{p})/n]}$

Or simpler: $\sigma_{\hat{p}} = 1/\sqrt{n}$ (conservative)

Margin of error: $1.96 \times \sigma_{\hat{p}}$

95% Confidence Interval:

$\hat{p} - 1.96 \cdot \sigma_{\hat{p}} \leq p \leq \hat{p} + 1.96 \cdot \sigma_{\hat{p}}$

Or simply: $\hat{p} \pm 1.96 \cdot \sigma_{\hat{p}}$

⑧ HYPOTHESIS TEST — 5 STEPS

1. **State H_0** (null) and **H_A** (alternative)
2. **Standard error:** $\sigma_{\hat{p}} = \sqrt{[\hat{p}(1-\hat{p})/n]}$ or $1/\sqrt{n}$
3. **Margin of error:** $1.96 \times \sigma_{\hat{p}}$
4. **Build confidence interval**
5. **Is given figure inside the CI?**

- YES → fail to reject H_0
- NO → reject H_0

⚠ **WATCH:** NEVER use 'accept' — only 'fail to reject'

⑨ HYPOTHESIS TEST — Z-SCORE METHOD

Step 1: State H_0 and H_A

Step 2: Convert to Z-score (test statistic):

$z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$ *(in log tables)*

Step 3: Decision rule

- Inside $-1.96 \leq z \leq 1.96$ → fail to reject H_0
- Outside → reject H_0

Step 4: State conclusion in plain English

⑩ p-VALUE METHOD

p-value = probability of result as extreme as observed (if H_0 true)

Decision rule:

- $p \leq 0.05$ → reject H_0
- $p > 0.05$ → fail to reject H_0

★ *Two-sided test? MULTIPLY tail probability by 2*

⚠ **WATCH:** Low p = strong evidence AGAINST H_0