**Hypothesis Testing — Detailed Explanation**

**1. Overview of Hypothesis Testing**  
Hypothesis testing is a formal procedure for deciding whether data provide enough evidence to reject a stated claim about a population. The main elements are:

* **Null hypothesis (H₀):** the default claim (e.g., “μ = μ₀”, “no effect”).
* **Alternative hypothesis (H₁ or Hₐ):** what we consider if H₀ is rejected (e.g., “μ ≠ μ₀”, “μ > μ₀”, “μ < μ₀”).
* **Test statistic:** a function of sample data that has a known sampling distribution under H₀.
* **Significance level (α):** pre-chosen probability of making a Type I error (common choices: 0.05, 0.01).
* **P-value:** probability, under H₀, of observing a test statistic as extreme or more extreme than the one observed.
* **Decision rule:** either (a) compare p-value to α (reject H₀ if p ≤ α), or (b) compare test statistic to critical value(s) (reject H₀ if statistic falls in rejection region).

**2. Type I and Type II Errors, Power**

* **Type I error (α):** rejecting H₀ when it is true. Probability = α (set by researcher).
* **Type II error (β):** failing to reject H₀ when H₁ is true. Probability depends on true effect, sample size, variance, α.
* **Power:** 1 − β = probability of correctly rejecting H₀ when H₁ is true. Increasing sample size, effect size, or α raises power.

Decision-outcome table:

* True H₀ & accept → Correct.
* True H₀ & reject → Type I error (α).
* False H₀ & accept → Type II error (β).
* False H₀ & reject → Correct (power).

Sample-size relation (two-sided z test, known σ):  
n ≈ [ (z\_{1−α/2} + z\_{1−β}) \* σ / Δ ]²  
where Δ = minimum effect size you want to detect, z quantiles from standard normal.

**3. Rejection Regions (Critical Regions)**

* **Two-tailed test:** H₀: parameter = value; H₁: parameter ≠ value. Rejection if |test statistic| > critical value (e.g., z > 1.96 for α = 0.05).
* **One-tailed test (right):** H₁: parameter > value. Reject if statistic > z\_{1−α}. Example: α=0.05 ⇒ critical z ≈ 1.645.
* **One-tailed test (left):** H₁: parameter < value. Reject if statistic < z\_{α}.  
  Critical values depend on the sampling distribution (z, t, χ², F) and degrees of freedom.

**4. Z-test**  
**Purpose:** Test means or proportions when sampling distribution is approximately normal and population variance is known (or n large so CLT applies).

**One-sample mean (known σ):**  
Test statistic: Z = (x̄ − μ₀) / (σ / √n)  
Decision: Compare Z to z critical or compute p-value from standard normal.

**Two-sample mean (known σs):**  
Z = (x̄₁ − x̄₂ − Δ₀) / √(σ₁²/n₁ + σ₂²/n₂)

**Proportion test (one sample):**  
Z = (p̂ − p₀) / √[ p₀(1−p₀) / n ]  
(Use pooled proportion for two-sample proportion tests when H₀: p₁ = p₂.)

**Assumptions:** independent samples, normality (or large n), known population σ (or large n).

**5. T-test (Student’s t)**  
Used when population variance is unknown and/or sample size is small. The sampling distribution is Student’s t with specific degrees of freedom (df).

**One-sample t-test:**  
t = (x̄ − μ₀) / (s / √n)  
df = n − 1

**Paired t-test (dependent samples):**  
Compute differences d\_i, d̄ = mean difference, s\_d = sd of differences.  
t = (d̄ − μ\_d0) / (s\_d / √n)  
df = n − 1

**Independent two-sample t-tests:**

* **Pooled t (equal variances assumed):**  
  sp² = [ (n₁−1)s₁² + (n₂−1)s₂² ] / (n₁ + n₂ − 2)  
  t = (x̄₁ − x̄₂ − Δ₀) / [ sp \* √(1/n₁ + 1/n₂) ]  
  df = n₁ + n₂ − 2
* **Welch’s t (unequal variances, recommended):**  
  t = (x̄₁ − x̄₂) / √( s₁²/n₁ + s₂²/n₂ )  
  df ≈ ( s₁²/n₁ + s₂²/n₂ )² / [ (s₁⁴ / ( n₁² (n₁−1) )) + (s₂⁴ / ( n₂² (n₂−1) )) ] (Welch–Satterthwaite approx.)

**Assumptions:** samples are independent (except paired test), underlying populations approximately normal (t robust for moderate n), variance assumption differs by test.

**6. F-test**  
**Purpose (simple):** Compare two population variances. Also used in ANOVA to compare means across multiple groups.

**Two-sample variance test:**  
F = s₁² / s₂²  
df₁ = n₁ − 1, df₂ = n₂ − 1  
Reject H₀: σ₁² = σ₂² if F is too large (or too small depending which variance is numerator). Because F is positive and asymmetric, use appropriate one- or two-sided critical values.

**ANOVA (one-way) — relation to F:**

* H₀: all group means equal.
* Between-group variability: MS\_between = SS\_between / (k − 1)
* Within-group variability: MS\_within = SS\_within / (N − k)
* F = MS\_between / MS\_within  
  Reject H₀ if F > F\_{α, k−1, N−k}.

**Assumptions (ANOVA/F-test):** independent observations, normality in each group, homogeneity of variances (equal variances).

**7. Chi-Square (χ²) Tests**  
**a) Goodness-of-fit test** — checks if observed categorical frequencies follow a specified distribution.  
Statistic: χ² = Σ (Oᵢ − Eᵢ)² / Eᵢ  
df = k − 1 − m (k categories, m parameters estimated from data)  
Reject H₀ if χ² large.

**b) Test of independence (contingency table)** — checks if two categorical variables are independent.

* Build contingency table with r rows and c columns, observed counts O\_{ij}.
* Expected counts: E\_{ij} = (row\_i\_total \* col\_j\_total) / N.
* χ² = Σ\_{i=1..r} Σ\_{j=1..c} (O\_{ij} − E\_{ij})² / E\_{ij}
* df = (r − 1)(c − 1)  
  Reject H₀ if χ² large.

**Assumptions:** expected counts Eᵢ typically ≥ 5 (rule of thumb); observations independent.

**8. Bayesian Testing**  
**Philosophy:** Treat hypotheses or parameters as random and use prior beliefs + observed data to compute posterior beliefs. Decision-making uses posterior probabilities or Bayes factors instead of p-values.

**Bayes’ theorem (for hypotheses H₀ and H₁):**  
Posterior odds = Prior odds × Bayes factor (BF)  
BF\_{10} = P(data | H₁) / P(data | H₀)

**Posterior probability of H₁:**  
P(H₁ | data) = [ P(data | H₁) P(H₁) ] / [ P(data | H₀) P(H₀) + P(data | H₁) P(H₁) ]

**Bayes factor interpretation (rough):**

* BF < 1/10: strong evidence for H₀
* BF ≈ 1: data do not prefer either hypothesis
* BF > 10: strong evidence for H₁  
  (Thresholds vary—these are conventional guidance.)

**Advantages:** direct probability statements about hypotheses, incorporate prior information, naturally penalizes model complexity (in many settings).  
**Disadvantages:** requires choosing priors (can be subjective), computation can be intensive, results depend on prior choice.

**Bayesian credible interval vs frequentist confidence interval:** credible interval gives direct probability that parameter lies in interval (given prior and data); confidence interval has a different frequentist interpretation.