**Hypothesis Testing**

* Hypothesis testing
  + To conduct an experiment to determine how credible a statement is in light of observed data
* Types of errors
  + Type I error: reject H0 when it is true (when it shouldn’t be rejected)
    - E.g. convicting an innocent person of a crime
    - P(type I) = α (significance level)
  + Type II error: do not reject H0 when it is false (when it should be rejected)
    - E.g. letting a guilty person go free
    - P(type II) = 1 − β
    - β = power of the test = (correctly) reject H0 when it is false
* Null hypothesis − the default hypothesis; a statement about the parameter/population of interest
  + Choose H0 to be the more likely outcome
  + E.g. H0: p = ½
* Alternative hypothesis − the hypothesis to support if H0 is rejected
  + E.g. H1/HA: p < ½ or p > ½ or p ≠ ½
  + Use one of the alternatives, depending on the problem given
* Ex: let population ~ X ~ N(μ, σ)
  + Sample = X1 … Xn
  + Step 1 – state H0 and HA
    - Test H0: μ = 3 vs. HA: μ ≠ 3 or μ > 3 or μ < 3
  + Step 2 – choose a test statistic
    - Test statistic/discrepancy measure – function of the data X1 … Xn that measures the degree of agreement between the data and the null hypothesis
    - T = 0 → strong agreement
    - T >> → poor agreement
  + Step 3 – find observed value of T
    - T\_obs = T(x1 … xn; θ)
  + Step 4 – critical value approach/p-value approach (see below)
  + Step 5 – compare T\_obs and t\_crit
    - Reject H0 if T\_obs falls into critical region
* **Critical value approach**
  + T as a r. v. − possible values of T are divided into 2 regions:
    - Acceptance region – i.e. H0 is true
    - Rejection/critical region – i.e. H0 is false
  + If H0: θ = θ0 vs. HA: θ > θ0 → alternative is right-sided
    - i.e. critical region is right-sided → accept | reject
    - i.e. P(T > crit) = α
    - Reject H0 if observed t > crit
  + If HA: θ < θ0
    - i.e. critical region is left-sided → reject | accept
    - i.e. P(T < crit) = α
    - Reject H0 if observed t < crit
  + If HA: θ ≠ θ0
    - i.e. critical region is two-sided → reject | accept | reject
    - i.e. P(|T| > crit) = α
    - Reject H0 if observed t < −crit or t > crit
  + crit = critical values of the test = values at the boundaries between the acceptance & rejection regions (found from table)
* Ex: toss a coin 20 times
  + Test H0: p = 0.5 vs. Ha: p < 0.5 where p = P(head) (coin is symmetric)
  + Under H0, X(# of heads) ~ Bin(20, 0.5)
  + Suppose observed value = 8 (given)
  + Since Ha is left-sided, critical region is below some x\_crit
    - Area of critical region = α
  + Suppose α = 5% = 0.05 (given)
  + Need to find x\_crit:
    - Reject H0 if P(X ≤ x\_crit) ≤ α
    - → P(X ≤ x\_crit) ≤ 0.05
  + Find x\_crit from binomial distribution table
    - x-crit = 5 for n = 20, p = 0.5, and P = 0.05
  + Since x\_obs > t\_crit we do not reject H0
* Ex: same situation
  + Test H0: p = 0.5 vs. Ha: p > 0.5
  + Reject H0 if
    - P(X ≥ x\_crit) ≤ α
    - → P(X < x\_crit) ≥ 1 – α
    - → P(X < x\_crit) ≥ 0.95
  + x\_crit = 14 for n = 20, p = 0.5, and P = 0.95
* Hypothesis testing for the mean
  + Let population ~ X ~ N(μ, σ)
  + Step 1 − test H0: μ = μ0 vs. H1: μ > μ0 vs. H2: μ < μ0 vs. H3: μ ≠ μ0
  + Step 2 – test statistic:
    - T = Z = (X-bar − μ0)/(σ/√n) if σ is known
    - T = Z = (X-bar − μ0)/(S/√n) if σ is unknown and n ≥ 30
    - T = t = (X-bar − μ0)/(S/√n) ~ tn-1 if σ is unknown and n < 30
  + Step 3 – find Z\_obs or t\_obs
  + Step 4 – find critical value given α
    - H1 → right-sided
      * Z\_crit = Z\_α
    - H2 → left-sided
      * Z\_crit = Z\_α
    - H3 → two-sided
      * Total area of critical region = α → α/2 on either side
      * Z\_crit = Z\_α/2
  + Step 5 – compare observed and critical values
* **P-value approach**
  + When H0 is true, the test statistic/discrepancy measure T should be less than the observed value t
  + P-value = probability that the test statistic is greater than the observed value of the test statistic (assuming H0 is true)
    - Right-sided alternative: p-val = P(T ≥ t\_obs)
    - Left-sided alternative: p-val = P(T ≤ t\_obs)
    - Two-sided alternative: p-val = P(|T| ≥ t\_obs) = 2 \* P(T ≥ t\_obs)
  + P-value determines the strength of the evidence against H0:
    - p-val < 0.01 = strong evidence against H0
    - 0.01 < p-val < 0.05 = evidence against H0
    - 0.05 < p-val < 0.1 = some evidence against H0
    - p-val > 0.1 = no evidence against H0
  + E.g. n = 10, x-bar = 6.4, s = 0.175 (normally distributed), α = 0.05 (one-sided alternative)
    - Test H0: μ0 = 6.3
    - t = (x-bar − μ0)/(s/√n) = 1.807
    - p-val = P(T ≥ 1.807) = 0.0521 using software
    - 0.05 < 0.0521 < 0.1 → some evidence against H0
    - 0.0521 > α → favour the alternative hypothesis
  + For two-sided alternative, i.e. Ha: μ ≠ 0
    - T = |X-bar − μ0|/(S/√n)
    - Dependent samples D = X – Y: T = |D-bar − μ0|/(SD/√n)
* Hypothesis testing for μ1 − μ2
  + Consider:
    - Population 1 ~ (μ1, σ1), population 2 ~ (μ2, σ2)
    - Independent samples X1 … Xn1, Y1 … Yn2
  + Test H0: μ1 − μ2 = 0 vs. Ha: μ1 − μ2 > 0
  + Test statistic = Z = ~ N(0,1) for n1, n2 ≥ 30 (unknown σ’s)
  + Use t for small samples
* Tests concerning population proportion
  + Consider population ~ X ~ Bin(p)
    - p = unknown population proportion
  + Step 1 – formulate hypotheses
    - Two-tailed test − H0: p = p0 vs. Ha: p ≠ p0
    - Left-tailed test – H0: p ≥ p0 vs. Ha: p < p0
    - Right-tailed test – H0: p ≤ p0 vs. Ha: p > p0
  + Step 2 – calculate test statistic = where p^ = X/n and q0 = 1 – p0
  + Step 3 – find critical value or p-value
  + Step 4 – reject/don’t reject H0 based on crit/p-val