**Estimation of Parameters**

* **Estimating parameters of population**
  + Let X1, X2, X3 be a random sample taken from a population
  + X = population attribute (a characteristic of the population)
    - → X1, X2, … Xn = random variables of sample (before experiment)
    - → x1, x2, … xn = actual values (after experiment)
  + Population’s distr. is known but depends on some unknown parameters
    - Based on a sample, we can estimate these parameters
  + If population → X ~ N(μ, σ)
    - μ = (∑ Xi)/N
    - Pop. var = σ2 = ∑(Xi – μ)2/N
  + Then (X1 … Xn) = sample
    - X-bar = (X1 + … Xn)/n
    - Sample var = S2 = ∑(Xi – X-bar)2/(n – 1)
    - X-bar is an estimator (a r. v.) for μ
    - x-bar is an estimate (a value) for μ
  + For a population with X ~ N(μ, σ) (known σ)
    - Sample mean X-bar ~ N(μ, σ/√n)
  + For a population with X ~ D(μ, σ) (any distr.) (known σ)
    - X-bar ~ (approx.) N(μ, σ/√n)
  + For a population with X ~ N(μ, σ) where σ is unknown and n < 30
    - X-bar ~ t(μ, S/√n)
    - t = student’s t-distribution, with n – 1 degrees of freedom
    - S2 = ∑(xi – x-bar)2/(n – 1) aka. sample variance
  + For a population with X ~ D(μ, σ) where σ is unknown and n ≥ 30
    - X-bar ~ (approx.) N(μ, S/√n)
  + Two methods of estimation – interval estimation & point estimation
* **Interval estimation**
  + Confidence level - the probability that the interval estimate will contain the parameter
  + CI – confidence interval
    - A specific interval estimate of a parameter determined using sample data & the specific confidence level of the estimate
  + Population ~ X ~ N(μ, σ), μ = ?, σ = known, n ≥ 30
    - α = level of significance
    - Z\_α = z-score in a Z-distribution where the area to the right of Z\_α = α
    - X-bar ~ N(μ, σ/√n)
    - Z = ~ N(0, 1) = pivotal quantity
    - P(-Z\_α/2 < Z < Z\_α/2) = 1 − α
      * =
    - → confidence interval (1 − α)100% for μ is
      * σ/√n = standard error
      * E = = marginal/maximum error
  + E.g. 96% CI → (1 − α)100% = 96% → α = 0.04 → α/2 = 0.02 → Z0.02
  + For small sample with n < 30
    - X-bar ~ t(μ, S/√n)
    - Pivotal = t = ~ t w/ d. f. = n – 1
    - → (1 − α)100% CI for μ is
* **Sample size estimation**
  + Determining sample size based on max error (E)
  + Given (1 − α)100% CI for known σ
    - |μ − X-bar| < E since X-bar – E < μ < X-bar + E
    - Minimum sample size = rounded up
* **Proportion estimation**
  + Population ~ Bin(n, p) where p = ?
  + Sample – estimate p based on p^ = sample proportion = X/n where X = # of positive outcomes
    - p^ is a point estimate of p
    - p~ is an estimator for p
    - We know that X ~ (approx.) N(np, √(npq))
    - → p~ ~ N(p, √(pq/n))
    - Interval estimate of p is for np ≥ 5 and nq ≥ 5
    - Minimum sample size for proportions =
* **Point estimation**
  + θ~ is an unbiased estimator of θ if & only if E(θ~) = θ
    - i.e. if “on average” its value = the value it is estimating
* **CI for difference b/t two population means**
  + See written note
  + Independent samples
    - Large samples (n1, n2 ≥ 30)
      * Known variances – use Z, σ
      * Unknown variances – use Z, S
    - Small samples
      * Equal variances – use t, Sp
      * Unequal variances – use t, S
  + Dependent/paired samples (i.e. Di = Xi – Yi)
    - Use t, SD
    - Note:
      * D-bar = (∑ Di)/n ← sample mean
      * μD = µX − μY  ← population mean
  + If CI for μ1 − μ2 is entirely within negative, conclude μ1 < μ2
  + If CI for μ1 − μ2 is entirely within positive, conclude μ1 > μ2
  + Otherwise, inconclusive
* **Maximum likelihood estimation (MLE)**
  + Likelihood function of θ = L(x1 … xn; θ) = Π f(xi; θ)
  + θ^(X1 … Xn) is the maximum likelihood estimator of θ if θ^ maximizes the likelihood function L(x1 … xn; θ)
  + Solve ∂/∂θ (ln L(x1 … xn; θ))
  + See written note