**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
  + a) X ~ N(μ, σ) → X-bar = N(μ, σ/√n)
  + b) X ~ D(μ, σ) (any distr.) → X-bar ~ (approx.) N(μ, σ/√n) (CLT)
  + c) X ~ D(μ, σ) where σ is unknown and n ≥ 30 → X-bar ~ (approx.) N(μ, S/√n)
  + d) X ~ N(μ, σ) where σ is unknown and n < 30 → X-bar ~ t(μ, S/√n)
    - t = student’s t-distribution
  + 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
    - 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
      * A known distribution depending on the parameters & the sample distr.
    - P(-Z\_α/2 < Z < Z\_α/2) = 1 − α (coverage probability)
    - =
    - → confidence interval (1 − α)100% for μ is
      * σ/√n = standard error
      * E = Z\_α/2 ⋅ σ/√n = marginal/maximum error
  + E.g. 96% CI → (1 − α)100% = 96% → α = 0.04 → α/2 = 0.02 → Z0.02
* **Sample size estimation**
  + Determining sample size based on max error (E)
  + Given (1 − α)100% CI
    - |μ − X-bar| < E since X-bar – E < μ < X-bar + E
    - Minimum sample size = rounded up
  + Given (1 − α)100% CI for μ, unknown σ, and n < 30
    - Pivotal = t = ~ t distribution with degrees of freedom n – 1
    - (1 − α)100% CI for μ is
  + (1 − α)100% CI for population proportion (p)
    - Population ~ Bin(n, p) where p = ?
    - Sample – estimate p based on p^ = sample proportion
      * p^ = 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))
      * Pivotal = Z =
      * Interval estimate of p is for np ≥ 5 and nq ≥ 5
      * Minimum sample size for proportions =
* **CI for difference b/t two population means**
  + Population 1 = X ~ D(μ1, σ1), population 2 = Y ~ D(μ2, σ2)
  + Samples x1 … xn1, y1 … yn2 are independent
  + Parameter = μ1 − μ2
* **Maximum likelihood estimation (MLE) for parameters of normal distr.**
  + See written note