Sampling distribution of a sample mean:

- $\bar{X}$ always has mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$
- $\bar{X}$ always has a Normal distribution **IF** the population distribution is Normal.

Central limit theorem:

- $\bar{X}$ is approximately Normal when *n* is large enough.
- We assume the CLT can be applied for samples of size:

1	if the population is Normal
10	if the population is symmetric
25	for any population, unless we know the population is severely skewed
40	any population

Inference on the Population Proportion *p* **Distribution:** $\hat{p} \,\dot{\sim}\, N\left(p, \sqrt{\frac{pq}{n}}\right)$

Conditions

- **Independence**
  - **Randomization: SRS or randomized experiment**
  - **Independence: no influence between observations**
  - **10%: sample less than 10% of population**
- **Success/Failure**
  - **Expect at least 10 successes and at least 10 failures**

CI: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$ **Test Stat:** $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$ **Sample size:** $n = \left(\frac{z^*}{M}\right)^2 p^* q^*$

Inference on the Population Mean μ **Distribution:** $\bar{X} \,\dot{\sim}\, N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Conditions

- **Independence**
  - **Randomization: SRS or randomized experiment**
  - **Independence: no influence between observations**
  - **10%: sample less than 10% of population**
- **Nearly Normal: Same as for Central Limit Theorem**

CI: σ **known:** $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ σ **not known:** $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$

Test Stat: σ **known:** $z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ σ **not known:** $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

Sample size: $n = \left(\frac{z^* s}{M}\right)^2$ **Note: for very small *n*, iterate using *t****

Inference Formulas for Two Samples

Proportions: Conditions

Independence Assumption

- **Randomization Condition:** The data are drawn independently and randomly.
- **10% Condition:** If *without replacement*, the data represent less than 10% of the population.

Independent Groups Assumption

- The two groups are independent of each other.

Sample size

- Successes and Failures for both ≥ 10

CI Formula: $\hat{p}_1 - \hat{p}_2 \pm z^* \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$ Test Statistic: $Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}\hat{q}}{n_1} + \dfrac{\hat{p}\hat{q}}{n_2}}}$ where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

Means: Conditions

Independence

- Same assumptions as proportions
- Between and within groups
- Randomization is evidence of independence.

Nearly Normal Condition

- Same as with a single mean, but must check both.

CI Formula: $\bar{x} - \bar{y} \pm t^* \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$ where $\text{df} = \min(n_x - 1, n_y - 1)$

Test Statistic: $T = \dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}}$ where $\text{df} = \min(n_x - 1, n_y - 1)$

Interpretation of *P*-values

$P-$value	Strength of Evidence Against H_0
$0 \leq p$-value ≤ 0.05	very strong
$0.05 < p$-value ≤ 0.10	strong
$0.10 < p$-value ≤ 0.20	moderate
$0.20 < p$-value ≤ 0.50	weak
$0.50 < p$-value	no

If a significance level α is given for a test, we reject H_0 in favor of H_a if the *p*-value is less than α. Otherwise, we fail to reject H_0.

Chi-Squared Formulas

Expected Counts: Expected Cell Count = $\dfrac{(\text{row total})(\text{column total})}{\text{table total}}$

The test statistic: $\chi^2 = \sum_{i=1}^{k=\#\text{ cells}} \dfrac{(o_i - e_i)^2}{e_i}$

Degrees of Freedom: df = (r-1)(c-1). If $r = 1$ or $c = 1$, then df = k-1.

Conditions when the Chi-Squared Test is appropriate:

No more than 20% of the expected cell counts should be < 5.
No expected cell count should be < 1.