**Unifications of Techniques: See the Wholeness and Many-Foldness**

<table>
<thead>
<tr>
<th>Population Continuous R.V.</th>
<th>( \mu, \sigma^2 ) Expected value and Variance both unknown Must be estimated with confidence.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Discrete R.V.</td>
<td>( \mu = \Sigma (xi pi) = E(x) ; \sigma^2 = (Ex^2) - (Ex)^2 = \Sigma (xi^2 pi) - (\Sigma (xi pi))^2 ; \sigma = \sqrt{\sigma^2} )</td>
</tr>
</tbody>
</table>

| Sample (s)                  | \( \bar{x} = \frac{\Sigma x}{n} ; s^2 = \frac{\Sigma (x-\bar{x})^2}{n-1} = \frac{SS}{n-1} ; SS = \text{Sum}(x^2) - (\text{Sum }x)^2/n. \) |
| Probability:                | \( P(A \text{ given } B) = P(A|B) = P(A \text{ and } B) / P(B) = \frac{P(AB)}{P(B)} \) |
|                            | \( P(A \text{ or } B) = P(A\cup B) = P(A) + P(B) - P(A \cap B) \) |
|                            | A and B are independent if \( P(A|B) = P(A) \) |

\[ \sigma \text{ unknown, use } s \]

**CLT in Action:**

For the parent population the **expected value** is:

\[ \mu = \Sigma (xi pi) = E(x) = 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) = 21/6 \]

which is the same for the other 3 sampling means distributions.

For the parent population the **variance** is:

\[ \sigma^2 = (Ex^2) - (Ex)^2 = \Sigma (xi^2 pi) - (Ex)^2 = 1(1/6) + 4(1/6) + 9(1/6) + 16(1/6) + 25(1/6) + 36(1/6) - (21/6)^2 = 4.04 \]

The variance for the others is \( \frac{\sigma^2}{n} = 4.04/2 = 2.02, 4.04/3 = 1.37, 4.04/4 = 1.01 \), respectively, it gets smaller as sample size increases.

**Binomial Probability Distribution**

\[ x \text{ success in } n \text{ trial with } \pi \text{ is probability success } \]

\[ P(x) = \frac{n!}{x! (n-x)!} \pi^x (1-\pi)^{n-x} \]

**Central limit theorem:** Sampling distribution of mean tends to be normal density as the (fixed) sample size increases. \( \mu = \mu_\bar{x} \) and \( \sigma_\bar{x} = \frac{\sigma}{\sqrt{n}} \) if
For Population

\[ Z = \frac{x - \mu}{\sigma / \sqrt{n}} \]

To make

\[ N(0,1) \]

For sampling distribution with sample size = n

\[ Z = \frac{x - \mu}{\sigma / \sqrt{n}} = \frac{x - \mu}{s / \sqrt{n}} \] (in that \( \mu \) and \( \sigma \) are unknown and large n, invoke the CLT)

T-statistic

Normal

\[ T = \frac{\bar{x} - \mu}{s / \sqrt{n}} \]

with small n, but population is Normal

---

**Notice the difference:** The **Z-score, Z-transformation** \( \frac{x - \bar{x}}{s} \), is used to make data dimensionless, often for comparison purposes. For example, price of the houses and their sized in Washington DC., and Tokyo.

**Estimation with Confidence**

<table>
<thead>
<tr>
<th>Population mean ( \mu )</th>
<th>1 Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{x} \pm z \sigma ) ( \bar{x} \pm z \sigma / \sqrt{n} )</td>
<td>( \bar{x} \pm t \sigma ) ( \bar{x} \pm t \sigma / \sqrt{n} ) (with ( v = n-1 ))</td>
</tr>
</tbody>
</table>

Population proportion (probability) \( p = \frac{x}{n} \)

Sample are sufficiently large

\( \hat{p} \pm z \sigma \hat{p} = \hat{p} \pm z \sqrt{\frac{p(1-p)}{n}} \)

Means: 2 pops

Matched pairs

Known also as the “before-and-after” test

Large sample size (CLT) if the d is not normal

\( \mu_d = (\mu_1 - \mu_2) \)

\( \overline{d} = \frac{z \sigma_d}{\sqrt{n}} = \overline{d} = \frac{z \sigma_d}{\sqrt{n}} \) (if \( \sigma_d unknown with large sample \)

\( \overline{d} = t \sigma_d / \sqrt{n} = \overline{d} = t \sigma_d / \sqrt{n} \) (if \( \sigma_d unknown with small sample \)

Variance \( \sigma^2 \)

\( \frac{(n-1)s^2}{X^2_{a/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{X^2_{1-a/2}} \)

Determination of sample size for Continuous and Discrete R.V.

\[
\begin{align*}
\text{With } d \text{ margin of error: } n &= \frac{z_{a/2}^2 \sigma^2}{d^2} \\
\text{With } \pi \text{ margin of error for proportion: } n &= \frac{z_{a/2}^2 \pi(1-\pi)}{d^2} \\
\end{align*}
\]

\[
\begin{align*}
\frac{s_1^2}{s_2^2 F(n_2-1,n_1-1,\frac{\sigma_1^2}{\sigma_2^2})} \leq \sigma_2^2 \leq s_2^2 F(n_1-1,n_2-1,\frac{\sigma_1^2}{\sigma_2^2}) \\
\end{align*}
\]
Hypothesis Testing

<table>
<thead>
<tr>
<th>1Population</th>
<th>2Populations</th>
</tr>
</thead>
</table>
| **1tailed: Ha:** $> \text{ or } (<)$<br>Rejection region: $Z > z_\alpha$ or $(Z < - z_\alpha)$ | **2tailed: Ha:** $\neq$
Rejection region: $Z > \frac{z_\alpha}{2}$ or $Z < - \frac{z_\alpha}{2}$ |
| **Pop mean $\mu$**<br>Ho: $\mu = \mu_0$
Ha: $\mu \neq \mu_0$

For one-sided use $Z > z_\alpha$ or $Z < - z_\alpha$

$Z = \frac{\bar{x} - \mu_0}{\sigma_\bar{x}} \sim \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Large sample size to invoke CLT. | Ho: $\mu_1 - \mu_2 = 0$
Ha: $\mu_1 - \mu_2 \neq 0$

If variances are almost the same, then use

$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p(\frac{1}{n_1} + \frac{1}{n_2})}$

with table $v = n_1 + n_2 - 2$

**Population proportion (probability) $(\pi \rightarrow p = \frac{X}{n})$**

Sample are sufficiently large

Ho: $\pi = \pi_0$
Ha: $\pi \neq \pi_0$

$Z = \sqrt{\frac{p - \pi_0}{\pi(1-\pi_0)}}$

Large sample size to invoke CLT. | Ho: $\pi_1 - \pi_2 = 0$
Ha: $\pi_1 - \pi_2 \neq 0$

With $p = \frac{x_1 + x_2}{n_1 + n_2}$ and $q = 1 - p$, large sample size.

**Population variance**

Ho: $\sigma^2 = \sigma_0^2$
Ha: $\sigma^2 \neq \sigma_0^2$

2tailed: $X^2 > \chi^2_{1-\alpha/2}$ or $X^2 < \chi^2_{\alpha/2}$

1tailed: $X^2 > \chi^2_{1-\alpha}$ (or $X^2 < \chi^2_{\alpha}$)

With $v = n - 1$

$X^2 = \frac{(n-1)s^2}{\sigma_0^2}$

Population is normal | Ho: $\frac{\sigma_1^2}{\sigma_2^2} = 1$

1tailed: Ha: $\frac{\sigma_1^2}{\sigma_2^2} > 1$

$F = \frac{s_1^2}{s_2^2}$, critical value $F_\alpha$ (n1-1, n2-1)

1tailed: Ha: $\frac{\sigma_1^2}{\sigma_2^2} < 1$

$F = \frac{s_1^2}{s_2^2}$, critical value $1/F_\alpha$ (n2-1, n1-1)

2tailed: Ha: $\frac{\sigma_1^2}{\sigma_2^2} \neq 1$, $F = \frac{\text{Larger variance}}{\text{Smaller variance}}$ critical value $F_{\alpha/2}$ (n1-1, n2-1), Populations are normal

**Population mean for Pair matched**

There is dependency, known also as the “before-and-after” test. Large sample size (CLT) if the d is not normal

Ho: $\mu_1 - \mu_2 = 0$

Ha: $\mu_1 - \mu_2 \neq 0$, $Z = \frac{d \bar{d}/\sqrt{n}}{s_d/\sqrt{n}}$