Combinations

$X_i \sim n(\mu, \sigma^2)$

$\bar{X} \sim n(\mu, \frac{\sigma^2}{n})$

$\sum X_i \sim n(n \cdot \mu, n\sigma^2)$

Normal approx of binomial

$X \sim bi(n, \pi)$

$\approx n(n\pi, n\pi(1-\pi))$

Applicable if

- $0 < \pi \pm 3\sqrt{\frac{\pi(1-\pi)}{n}} < 1$
- $n\pi > 5$, $n(1-\pi) > 5$

Bias

$MSE(\hat{\theta}^*) = var(\hat{\theta}^*) + B(\hat{\theta}^*)^2$

Proportion

Confidence interval

$$\frac{Z_{\frac{\alpha}{2}}^2 + 2pn \pm \sqrt{Z_{\frac{\alpha}{2}}^2 (z^2 + 4np(1-p))}}{2(n + Z_{\frac{\alpha}{2}}^2)}$$

$p = 78\%$ $n = 1030$

$$Z_{crit} = Z_{0.05} = \frac{p - \pi_{ci}}{\sqrt{\frac{\pi_{ci}(1-\pi_{ci})}{n}}}$$

Difference of samples (proof by m.g.f.)

Variance

$$\frac{1}{\sigma^2} \times \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

TODO: Ask why not $\frac{S^2}{\sigma^2}$?

t-distribution

if $W \sim \chi^2_v$ think: sample variance(?)

$$T = \frac{Z}{\sqrt{\frac{W}{v}}} \sim t_v$$

$P\left(\frac{}{\sqrt{\frac{S^2}{n}}} < \dots\right)$

χ^2_v dist

- $Z^2 \sim \chi^2_1$ if $Z \sim n(0,1)$

F-distribution

- $F = \frac{\frac{W_1}{v_1}}{\frac{W_2}{v_2}}$ if $W_1 \sim \chi^2_{v_1}$, $W_2 \sim \chi^2_{v_2}$

$$\boxed{F_{(x,y),\frac{\alpha}{2}} = \frac{1}{F_{(y,x),\frac{\alpha}{2}}}}$$

N.B. $\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F_{n_1, n_2}$

Differences of normal vars

$X_i \sim n(\mu_x, \sigma_x^2)$

$Y_i \sim n(\mu_y, \sigma_y^2)$

σ_x, σ_y known

$$\bar{X} - \bar{Y} \sim n\left(\mu_x - \mu_y, \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right)$$

σ_x, σ_y unknown, $n \geq 30$, $m \geq 30$

$\sigma_x \approx s_x$

$\sigma_y \approx s_y$

σ_x, σ_y unknown but $\sigma_x \neq \sigma_y$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} \sim t_v$$

$\longrightarrow$ Basically same as

$$\boxed{v = \left[\frac{\left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}\right)^2}{\frac{\left(\frac{S_x^2}{n_x}\right)^2}{n_x+1} + \frac{\left(\frac{S_y^2}{n_y}\right)^2}{n_y+1}} - 2 \right]}$$

σ_x, σ_y known $\quad \sigma_x = \sigma_y$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_p^2}{n_x} + \frac{S_p^2}{n_y}}} \sim t_{n_x + n_y - 2}$$

$$\boxed{S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}}$$

$$Var(Y_3) = E[Y_3^2] - E[Y_3]^2$$

$$\underline{Var(Y_3)} + \underline{E[Y_3]^2} = E[Y_3^2]$$

Hypothesis testing

$\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid \text{true } H_0)$

$\beta = P(\text{Type II error}) = P(\text{Fail to reject } H_0 \mid \text{false } H_0)$

Sample size estimation

$$n = \left\lceil \frac{(z_\alpha + z_\beta)^2 \times \sigma^2}{(\mu_a - \mu_0)^2} \right\rceil$$

alternate