- Expected value and Variance of a linear function of RVs $X\ and\ Y$: Let the linear function $h(X) = aX + bY$.
  $E(aX + bY) = a \cdot E(X) + b \cdot E(Y)$, $\quad V(aX + bY) = a^2 \cdot V(X) + b^2 \cdot V(Y)$
- $X \sim Bin(n,p), f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x},\ for\ x = 0, 1, 2, \dots, n.\ E[X] = np, V(X) = np(1-p)$
- $X \sim Unif(a,b)\ (Discrete): f(x) = \frac{1}{b-a+1}, \mu = EX = \frac{(a+b)}{2}, V(X) = \frac{(b-a+1)^2-1}{12}$
- $X \sim Unif(a,b)\ (Continuous): f(x) = \frac{1}{b-a}\ for\ a \le x \le b, \mu = EX = \frac{(a+b)}{2}, V(X) = \frac{(b-a)^2}{12}$
- $X \sim \mathcal{N}(\mu, \sigma^2), f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty, E[X] = \mu, V(X) = \sigma^2$
- If $X_1, X_2, \dots, X_p$ are random variables, $and\ Y = c_1 X_1 + c_2 X_2 + \dots + c_p X_p$, then
  $EY = c_1 EX_1 + c_2 EX_2 + \dots + c_p EX_p,\ V(Y) = c_1^2 V(X_1) + c_2^2 V(X_2) + \dots + c_p^2 V(X_p) + 2\sum_{i<j}\sum c_i c_j \cdot Cov(X_i, X_j)$
- If $X_1, X_2, \dots, X_p$ are independent: $V(Y) = c_1^2 V(X_1) + c_2^2 V(X_2) + \dots + c_p^2 V(X_p)$
- If $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)\ for\ i = 1, \dots, p \to Y = \sum_i c_i X_i \sim \mathcal{N}(\sum_i c_i \mu_i, \sum_i c_i^2 \sigma_i^2)$.
- $\bar{x} = \frac{\sum_{i=1}^n x_i}{n},\ \ s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{(\sum_{i=1}^n x_i^2) - n\bar{x}^2}{n-1}$
- **Central Limit Theorem:** If $X_1, X_2, \dots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\bar{X}$ is the sample mean, the limiting form of the distribution of $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ as $n \to \infty$, is the standard normal distribution.
- If we have two independent populations with means μ_1 & μ_2 and variances σ_1^2 & σ_2^2 and if $\bar{X}_1$ & $\bar{X}_2$ are the sample means of two independent random samples of size n_1 & n_2 from these populations, then the sampling distribution of $Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ is approximately standard normal.
- The point estimator $\hat{\Theta}$ is an unbiased estimator of the parameter θ if $E(\hat{\Theta}) = \theta$.
- $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + Bias(\hat{\theta})^2,\ Bias(\hat{\Theta}) = E(\hat{\Theta}) - \theta$
- Relative efficiency of $\hat{\Theta}_2$ to $\hat{\Theta}_1$ is defined as $RE = \frac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_2)}$
- If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100 \cdot (1-\alpha)\%$ CI on μ is given by $\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$
- CI on μ for a large sample ($n \ge 40$)
  1. Known population variance σ^2: $\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$
  2. Unknown population variance σ^2: $\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$
- If $\bar{x}$ and s are the mean and sample standard deviation of a random sample ($n < 40$) from a normal distribution with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval on μ is given by
  $$\bar{x} - t_{\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}} \le \mu \le \bar{x} + t_{\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}}$$
- If s^2 is the sample variance from a random sample of n observations from a normal distribution with unknown variance σ^2, then a $100(1-\alpha)\%$ confidence interval on σ^2 is $\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}$
- If $\hat{p}$ is the proportion of observations in a random sample of size n that belongs to a class of interest, an approximate $100(1-\alpha)\%$ confidence interval on the proportion p of the population that belongs to this class is $\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, where $np > 5$ and $n(1-p) > 5$
- Type I error: Rejecting the null hypothesis H_0 when it is true.
- Type II error: Failing to reject the null hypothesis H_0 when it is false.
- $\alpha = P(Type\ I\ Error),\ \ \beta = P(Type\ II\ Error),\ \ Power = 1 - \beta$
- [Note] When you calculate your test statistics, always assume H_0 is true.

- Large sample ($n \geq 40$): $z_0 = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ or $z_0 = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$
- Small Sample: Normally distributed

  1. Known population variance: $z_0 = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$
  2. Unknown population variance: $t_0 = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$
- HT on variance: $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$
- HT on proportion: similar to large sample: $\hat{p} = \frac{x}{n} \sim N\left(p, \frac{p(1-p)}{n}\right)$, $z_0 = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{n\hat{p}-np}{\sqrt{np(1-p)}}$
- Large sample ($n \geq 40$): $z_0 = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ or $z_0 = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ $if\ \sigma_1^2, \sigma_2^2\ unknown.$
- Small Sample: Normally distributed

  1. Known population variances: $z_0 = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$
  2. Unknown population variance:

   I. Equal Variance Assumption: $t_0 = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$, where $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

   II. Unequal Variance Assumption: $t_0 = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \sim t_v$, where $v = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$