Formula Sheet

- Sample mean and variance:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

- Median:

$$(0.5)(n+1)$$

- Addition rule:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

- Multiplication rule:

$$P(E_1 \cap E_2) = P(E_1) \times P(E_2|E_1)$$

- Conditional probability:

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$

- Rules of total probability:

$$P(E_1) = P(E_2) \times P(E_1|E_2) + P(E_2^C) \times P(E_1|E_2^C)$$

- Expected value:

$$\mu_Y = \sum y_i P(Y = y_i)$$

- Variance:

$$\sigma_Y^2 = \sum (y_i - \mu_Y)^2 P(Y = y_i)$$

- Linear combinations of random variables:

$$\mu_{aX+b} = a\mu_X + b, \quad \sigma_{aX+b}^2 = a^2\sigma_X^2$$

- Binomial $B(n, p)$:

  *
  $$P(Y = j) = \binom{n}{j} p^j (1-p)^{n-j},$$

  where
  $$\binom{n}{j} = \frac{n!}{j!(n-j)!} \text{ and } n! = n(n-1)\cdots 1$$

  *
  $$\mu_Y = np, \quad \sigma_Y^2 = np(1-p)$$

- Standardization:

$$Z = \frac{1}{\sigma_Y}(Y - \mu_Y)$$

- Sampling distribution of the sample mean:

$$\mu_{\bar{Y}} = \mu, \sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$$

- Normal approximation: $Y \sim B(n,p)$ can be approximated by $N(np, np(1-p))$
- Continuity correction: $P(Y = n) = P(n - 0.5 < Y < n + 0.5)$
- $1 - \alpha$ confidence interval for μ:
  - ∗ Two-sided: $\bar{Y} \pm t_{n-1}(\alpha/2) \times \text{SE}_{\bar{Y}}$
  - ∗ Upper one-sided: $(-\infty, \bar{Y} + t_{n-1}(\alpha) \times \text{SE}_{\bar{Y}})$
  - ∗ Lower one-sided: $(\bar{Y} - t_{n-1}(\alpha) \times \text{SE}_{\bar{Y}}, \infty)$

  where

$$\text{SE}_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

- $1 - \alpha$ confidence interval for $\mu_1 - \mu_2$:
  - ∗ Two-sided: $(\bar{Y}_1 - \bar{Y}_2) \pm t_\nu(\alpha/2) \times \text{SE}_{\bar{Y}_1 - \bar{Y}_2}$
  - ∗ Upper one-sided: $(-\infty, (\bar{Y}_1 - \bar{Y}_2) + t_\nu(\alpha) \times \text{SE}_{\bar{Y}_1 - \bar{Y}_2})$
  - ∗ Lower one-sided: $((\bar{Y}_1 - \bar{Y}_2) - t_\nu(\alpha) \times \text{SE}_{\bar{Y}_1 - \bar{Y}_2}, \infty)$

  where the degrees of freedom ν will be given and

$$\text{SE}_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- One-sample t test:

H_0	H_A	Test statistic	Rejection region	p-value
$\mu = c$	$\mu \neq c$	$T = \frac{\bar{Y} - c}{\text{SE}_{\bar{Y}}} \overset{H_0}{\sim} t_{n-1}$	$\|T\| > t_{n-1}(\alpha/2)$	$2 \times P(t_{n-1} > \|T\|)$
$\mu \geq c$	$\mu < c$		$T < -t_{n-1}(\alpha)$	$P(t_{n-1} < T)$
$\mu \leq c$	$\mu > c$		$T > t_{n-1}(\alpha)$	$P(t_{n-1} > T)$

- Two-sample t test:

H_0	H_A	Test statistic	Rejection region	p-value
$\mu_1 - \mu_2 = c$	$\mu_1 - \mu_2 \neq c$	$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - c}{\text{SE}_{\bar{Y}_1 - \bar{Y}_2}} \overset{H_0}{\sim} t_\nu$ with	$\|T\| > t_\nu(\alpha/2)$	$2 \times P(t_\nu > \|T\|)$
$\mu_1 - \mu_2 \geq c$	$\mu_1 - \mu_2 < c$		$T < -t_\nu(\alpha)$	$P(t_\nu < T)$
$\mu_1 - \mu_2 \leq c$	$\mu_1 - \mu_2 > c$	$\nu = \frac{(\text{SE}_1^2 + \text{SE}_2^2)^2}{\text{SE}_1^4/(n_1 - 1) + \text{SE}_2^4/(n_2 - 1)}$	$T > t_\nu(\alpha)$	$P(t_\nu > T)$

- Comparison of paired samples: calculate the difference between each pair of observations and then perform inference (confidence interval and t test) on these differences as if it were a one-sample analysis.
- 95% confidence interval for p:
  - ∗ Two-sided: $\tilde{p} \pm 1.96 \times \text{SE}_{\tilde{p}}$
  - ∗ Upper one-sided: $(0, \tilde{p} + 1.645 \times \text{SE}_{\tilde{p}})$
  - ∗ Lower one-sided: $(\tilde{p} - 1.645 \times \text{SE}_{\tilde{p}}, 1)$

  where

$$\tilde{p} = \frac{Y + 2}{n + 4}, \quad \text{SE}_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}$$

- 95% confidence interval for $p_1 - p_2$:

$$(\tilde{p}_1 - \tilde{p}_2) \pm 1.96 \times \mathrm{SE}_{\tilde{p}_1 - \tilde{p}_2}$$

where

$$\tilde{p}_1 = \frac{Y_1 + 1}{n_1 + 2}, \quad \tilde{p}_2 = \frac{Y_2 + 1}{n_2 + 2}, \quad \mathrm{SE}_{\tilde{p}_1 - \tilde{p}_2} = \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1 + 2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2 + 2}}$$

- Chi-square goodness-of-fit test:

$$T = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \overset{H_0}{\sim} \chi^2_{k-1},$$

H_0 is rejected at the α level of significance if

$$p\text{-value } = P(\chi^2_{k-1} > T) < \alpha \text{ or } T > \chi^2_{k-1}(\alpha)$$

- Chi-square test of independence:

$$T = \sum_{i=1}^{r\times k} \frac{(o_i - e_i)^2}{e_i} \overset{H_0}{\sim} \chi^2_{(r-1)(k-1)},$$

H_0 is rejected at the α level of significance if

$$p\text{-value } = P(\chi^2_{(r-1)(k-1)} > T) < \alpha \text{ or } T > \chi^2_{(r-1)(k-1)}(\alpha)$$

- ANOVA:

Source	**df**	**SS** (Sum of Squares)	**MS** (Mean Square)
Between groups	$I-1$	$\sum_{i=1}^{I} n_i(\bar{Y}_i - \bar{Y})^2$	SS/df
Within groups	$n-I$	$\sum_{i=1}^{I}(n_i - 1)s_i^2$	SS/df
Total	$n-1$	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y})^2$	

where

$$\bar{Y} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i} Y_{ij}}{n} = \frac{\sum_{i=1}^{I} n_i \bar{Y}_i}{n}, \quad n = \sum_{i=1}^{I} n_i$$

- F test:

$$T = \frac{\mathrm{MSB}}{\mathrm{MSW}} \overset{H_0}{\sim} F_{I-1,n-I},$$

H_0 is rejected at the α level of significance if

$$p\text{-value } = P(F_{I-1,n-I} > T) < \alpha \text{ or } T > F_{I-1,n-I}(\alpha)$$

- The Bonferonni-adjusted $1-\alpha$ confidence interval for $\mu_a - \mu_b$ is

$$(\bar{Y}_a - \bar{Y}_b) \pm t_{n-I}(\alpha/(2k)) \times \mathrm{SE}_{\bar{Y}_a - \bar{Y}_b},$$

where

$$\mathrm{SE}_{\bar{Y}_a - \bar{Y}_b} = \sqrt{\mathrm{MSW} \times \left(\frac{1}{n_a} + \frac{1}{n_b}\right)}$$

- Correlation coefficient:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{s_X}\right)\left(\frac{Y_i - \bar{Y}}{s_Y}\right)$$

- Fitted regression line:

$$\hat{Y} = b_0 + b_1 X$$

  where

$$b_1 = r\frac{s_Y}{s_X}, \quad b_0 = \bar{Y} - b_1\bar{X}$$

  - Residuals: $e_i = Y_i - \hat{Y}_i$ where $\hat{Y}_i = b_0 + b_1 X_i$
  - Error sum of suares: $\text{SSE} = \sum_{i=1}^{n} e_i^2$
  - Residual standard deviation: $s_e = \sqrt{\frac{\text{SSE}}{n-2}}$

- Coefficient of determination:

$$r^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 - \sum_{i=1}^{n}(Y_i - \hat{Y}_i)}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{\text{SSE}}{(n-1)s_Y^2}$$

- $1 - \alpha$ confidence interval for β_1:

$$b_1 \pm t_{n-2}(\alpha/2) \times \text{SE}_{b_1},$$

  where

$$\text{SE}_{b_1} = \frac{s_e}{s_X\sqrt{n-1}}$$

- Test of $H_0 : \beta_1 = 0$ or $H_0 : \rho = 0$:

$$T = \frac{b_1 - 0}{\text{SE}_{b_1}} = r\sqrt{\frac{n-2}{1-r^2}}$$

  H_0 is rejected at the α level of significance if

$$p\text{-value } = 2 \times P(t_{n-2} > |T|) < \alpha \text{ or } |T| > t_{n-2}(\alpha/2)$$