Formula Sheet

- Sample mean and variance:

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

- 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: $\bar{Y} \pm t_{n-1}(\alpha/2) \times \mathrm{SE}_{\bar{Y}}$
  * Upper one-sided: $(-\infty, \bar{Y} + t_{n-1}(\alpha) \times \mathrm{SE}_{\bar{Y}})$
  * Lower one-sided: $(\bar{Y} - t_{n-1}(\alpha) \times \mathrm{SE}_{\bar{Y}}, \infty)$

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

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

  where the degrees of freedom ν will be given and
  $$\mathrm{SE}_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- One-sample t test:

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

- Two-sample t test:

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

- 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: $\tilde{p} \pm 1.96 \times \mathrm{SE}_{\tilde{p}}$
  * Upper one-sided: $(0, \tilde{p} + 1.645 \times \mathrm{SE}_{\tilde{p}})$
  * Lower one-sided: $(\tilde{p} - 1.645 \times \mathrm{SE}_{\tilde{p}}, 1)$

  where

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

- 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

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