Formula Sheet

Sample Statistics:

Sample mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$

Linear Transformations/Combinations:

Mean: $E(aX+b) = aE(X)+b, E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$

Variance: $V(aX+b) = a^2V(X), V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j cov(X_i, X_j)$

Discrete:

Binomial:

$$p(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x}, x \in \{0, 1, \ldots, n\}$$

$E(X) = np, V(X) = np(1-p)$

Poisson:

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x \in \{0, 1, 2, 3, \ldots\}$$

$E(X) = V(X) = \lambda$

Negative Binomial:

$$p(x;r,p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, x \in \{r, r+1, r+2, \ldots\}$$

$E(X) = \frac{r}{p}, V(X) = \frac{r(1-p)}{p^2}$

Continuous:

Uniform:

$$f(x;a,b) = \frac{1}{b-a}, x \in [a,b]$$

$E(X) = \frac{a+b}{2}, V(X) = \frac{(b-a)^2}{12}$

Normal:

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{1}{2\sigma^2}(x-\mu)^2}, x \in (-\infty, \infty)$$

$E(X) = \mu, V(X) = \sigma^2$

Central Limit Theorem: $n > 30$

$$\bar{X} \overset{approx}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

<u>One Sample Inference</u>: Use $\mathcal{N}(0,1)$ or t quantiles where necessary, as well as σ^2 or s^2 where necessary, depending on whether the variance is known or unknown.

For a random variable Y, let Y_c denote the c^{th} percentile, i.e. $P(Y \leq Y_c) = c$.

Two sided $(1-\alpha)100\%$ C.I. for μ:

$$\bar{x} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \sigma^2 \text{ known}$$

$$\bar{x} \pm t_{(1-\alpha/2,n-1)}\frac{s}{\sqrt{n}} \qquad \sigma^2 \text{ unknown}$$

Sample size for C.I. width $\leq k$:

$$n \geq \left\lceil \left(2z_{1-\alpha/2}\frac{\sigma}{k}\right)^2 \right\rceil$$

Test Statistic for μ:

$$T.S. = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \overset{H_0}{\sim} \mathcal{N}(0,1) \qquad \sigma^2 \text{ known}$$

$$T.S. = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \overset{H_0}{\sim} t_{n-1} \qquad \sigma^2 \text{ unknown}$$

Agresti-Coull CI for a proportion: if $\widetilde{n} = n + 4$, $\widetilde{p} = (x+2)/\widetilde{n}$, then

$$\widetilde{p} \pm z_{1-\alpha/2}\sqrt{\frac{\widetilde{p}(1-\widetilde{p})}{\widetilde{n}}}$$

Two sided $(1-\alpha)100\%$ C.I. (variance):

$$\left(\frac{(n-1)s^2}{\mathcal{X}^2_{(1-\alpha/2,n-1)}}, \frac{(n-1)s^2}{\mathcal{X}^2_{(\alpha/2,n-1)}}\right)$$

Test statistic (variance):

$$T.S. = \frac{(n-1)s^2}{\sigma_0^2} \overset{H_0}{\sim} \mathcal{X}^2_{n-1}$$

Two sample inference

Two sided C.I. (means):

$$v = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{(s_x^2/n_x)^2}{n_x - 1} + \frac{(s_y^2/n_y)^2}{n_y - 1}}$$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

$$\bar{x} - \bar{y} \pm z_{1-\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \qquad \sigma_x^2, \sigma_y^2 \text{ both known}$$

$$\bar{x} - \bar{y} \pm t_{1-\alpha/2, n_x+n_y-2}\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}} \qquad \sigma_x^2, \sigma_y^2 \text{ unknown, assume } \sigma_x^2 = \sigma_y^2$$

$$\bar{x} - \bar{y} \pm t_{1-\alpha/2, v}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \qquad \sigma_x^2, \sigma_y^2 \text{ unknown, do not assume } \sigma_x^2 = \sigma_y^2$$

Two sided C.I. (proportions):

$$\widetilde{p}_x - \widetilde{p}_y \pm z_{1-\alpha/2}\sqrt{\frac{\widetilde{p}_x(1 - \widetilde{p}_x)}{\widetilde{n}_x} + \frac{\widetilde{p}_y(1 - \widetilde{p}_y)}{\widetilde{n}_y}} \text{ where } \widetilde{n}_x = n_x + 2 \text{ and } \widetilde{p}_x = \frac{x + 1}{\widetilde{n}_x}$$

.