Some (possibly) useful formulae

- **Binomial distribution**
  - n **trials, probability** p **of success**:

$$P(X = k) = C_k^n p^k q^{n-k}, k = 0, 1, \ldots, n$$

where $C_k^n = \frac{n!}{k!(n-k)!}$

- **Central Limit Theorem**
  - **Version** 1: If $x_1, \ldots, x_n$ is a random sample from a population with mean μ and standard deviation σ, then for large enough n the distribution of $\bar{x}$ is approximately normal with mean μ, standard deviation $\sigma/\sqrt{n}$.
  - **Version** 2: If $x_1, \ldots, x_n$ is a random sample from a population with mean μ and standard deviation σ, then for large enough n the distribution of $\sum_{i=1}^n x_i$ is approximately normal with mean $n\mu$, standard deviation $\sqrt{n}\sigma$.

- **Sampling from a general population**
  - **Sample mean of a sample of size** n: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
  - **Sample standard deviation of sample of size** n:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

  - **Mean and standard deviation of** $\bar{x}$: μ and $\frac{\sigma}{\sqrt{n}}$
  - **Distribution of** $\bar{x}$ ($n \geq 30$):

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \approx z, \text{ a standard normal}$$

  - **Distribution of** $\bar{x}$ **(population normal)**:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} = t, \text{ a } t \text{ distribution with } n-1 \text{ degrees of freedom}$$

  - **Distribution of** $\bar{x}_1 - \bar{x}_2$ ($n_1, n_2 \geq 30$):

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx z$$

- **Pooled estimator for s^2, the common variance**:

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

- **Distribution of $\bar{x}_1 - \bar{x}_2$ (populations normal, variances equal)**:

$$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = t \text{ with } n_1+n_2-2 \text{ degrees of freedom}$$

  where s^2 is the pooled estimator for variance

- **Distribution of s^2 (population normal)**:

$$\frac{(n-1)s^2}{\sigma^2} = \chi^2, \text{ a } \chi^2 \text{ distribution with } n-1 \text{ degrees of freedom}$$

- **Drawing samples from a binomial population**
  - **Sample proportion of a sample of size n**: $\hat{x} = \frac{\text{Number of successes}}{n}$
  - **Mean and standard deviation of $\hat{p}$**: p and $\sqrt{\frac{pq}{n}}$
  - **Distribution of $\hat{p}$ $(n\hat{p}, n\hat{q} > 5)$**:

$$\frac{\hat{p}-p}{\sqrt{pq/n}} \approx \frac{\hat{p}-p}{\sqrt{\hat{p}\hat{q}/n}} \approx z$$

  - **Distribution of $\hat{p}_1 - \hat{p}_2$ $(n_1\hat{p}_1, n_1\hat{q}_1, n_2\hat{p}_2, n_2\hat{q}_2 > 5)$**:

$$\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}} \approx z$$

  - **Pooled estimator for p, the common proportion**:

$$\hat{p} = \frac{x_1+x_2}{n_1+n_2}$$

  - **Distribution of $\hat{p}_1 - \hat{p}_2$ $(n_1\hat{p}_1, n_1\hat{q}_1, n_2\hat{p}_2, n_2\hat{q}_2 > 5$, proportions equal)**:

$$\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1}+\frac{\hat{p}\hat{q}}{n_2}}} \approx z$$

  where $\hat{p}$ is pooled estimator for variance