Summary of Formulae

Sample mean and variance

$$\text{Mean: } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{Variance: } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Probability Rules

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B)$$
$$\Pr(A \text{ and } B) = \Pr(A)\Pr(B|A) = \Pr(B)\Pr(A|B)$$
$$\Pr(A|B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)} = \frac{\Pr(B|A)\Pr(A)}{\Pr(B|A)\Pr(A) + \Pr(B|A^{\mathsf{C}})\Pr(A^{\mathsf{C}})}$$

Random Variables
If X and Y are random variables with means $E[X]$ and $E[Y]$ respectively, then

$$E[aX + bY] = a\,E[X] + b\,E[Y].$$

If X and Y are independent random variables with variances $Var(X)$ and $Var(Y)$ respectively, then

$$Var(aX + bY) = a^2\,Var(X) + b^2\,Var(Y).$$

Discrete Distributions

$$\text{Mean: } E[Y] = \sum_{i=1}^{k} y_i \Pr(Y = y_i) \qquad \text{Variance: } Var(Y) = \sum_{i=1}^{k} (y_i - E[Y])^2 \Pr(Y = y_i)$$

Normal Distribution
A normal random variable, Y, has mean $E[Y] = \mu$ and variance $Var(Y) = \sigma^2$. A standard normal random variable, Z, has mean 0 and variance 1. To transform a normal random variable Y into a standard normal Z (and vice versa):

$$Z = \tfrac{Y-\mu}{\sigma} \qquad Y = Z\sigma + \mu$$

Binomial Distribution
A binomial random variable X has mean $E[X] = np$ and variance $Var(X) = np(1-p)$.

Distributions of Statistics

- The distribution of $\bar{y}$ has mean μ and standard error $\frac{\sigma}{\sqrt{n}}$
  - Estimate of standard error $\frac{s}{\sqrt{n}}$
- The distribution of $\bar{y}_1 - \bar{y}_2$ has mean $\mu_1 - \mu_2$ and standard error $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.
  - Estimate of standard error $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.
- The distribution of $\hat{p} = \frac{x}{n}$ has mean p and standard error $\sqrt{\frac{p(1-p)}{n}}$
  - For hypothesis test with $\mathrm{H}_0 : p = p_0$ the standard error is $\sqrt{\frac{p_0(1-p_0)}{n}}$
- The distribution of $\hat{p}_1 - \hat{p}_2$ has mean $p_1 - p_2$ and standard error $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$
  - For hypothesis test with $\mathrm{H}_0 : p_1 - p_2 = 0$ the estimated standard error is $\sqrt{\frac{\hat{p}^*(1-\hat{p}^*)}{n_1} + \frac{\hat{p}^*(1-\hat{p}^*)}{n_2}}$
  - $\hat{p}^* = \frac{x_1+x_2}{n_1+n_2} = \frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$.

Contingency Tables

- Test statistic: $X^2 = \sum \frac{(\text{observed}-\text{expected})^2}{\text{expected}}$
- df $= (R-1)(C-1)$, where R and C are the number of rows and columns respectively
- The expected count: expected $= \frac{\text{row total}\times\text{column total}}{\text{table total}}$

Regression
The simple linear regression model is: $y = \beta_0 + \beta_1 x + \varepsilon$, where ε is normally distributed with mean 0 and variance σ_ε^2.
The fitted model is: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$