Descriptive Statistics & Probability

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N} \qquad \sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)}{N} \qquad \rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \qquad s_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1} \qquad r = \frac{s_{xy}}{s_x s_y}$$

Discrete Distributions

$$P(x) = P(X=x) \text{ , for all } x \qquad F(x_0) = P(X \le x_0) = \sum_{x \le x_0} P(x)$$

$$E[X] = \sum_x xP(x) \qquad Var(X) = \sum_x (x-\mu)^2 P(x) \qquad E[g(X)] = \sum_x g(x)P(x)$$

Continuous Distributions

$$F(x_0) = P(X \le x_0) = \int_{-\infty}^{x_0} f(x)dx \qquad E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

$$Var[X] = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx \qquad E[g(x)] = \int_x g(x)f(x)dx$$

Joint Distributions

$$E[aX+bY] = aE[X]+bE[Y] \qquad Var[aX+bY] = a^2Var[X]+b^2Var[Y]+2abCov(X,Y)$$

$$Cov(X,Y) = E[(x-\mu_X)(Y-\mu_Y)] = E[XY]-\mu_X\mu_Y$$

Sampling Distributions

$$E[\bar{X}] = \mu \qquad \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

CLT: $Z = \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} \sim N(0,1)$ for $n > 25$ for $n > 100, t \to z$

Confidence Intervals

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \bar{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

Hypothesis Testing

One sample

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \qquad t_{n-1} = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

Two sample

$$t = \frac{(\bar{Y}_m-\bar{Y}_w)-d_0}{SE(\bar{Y}_m-\bar{Y}_w)} \qquad SE(\bar{Y}_m-\bar{Y}_w) = \sqrt{\frac{s_m^2}{n_m}+\frac{s_w^2}{n_w}}$$