Descriptive Statistics

- $\bar{x} = \frac{\sum x}{n}$
- $s^2 = \frac{1}{n-1}\sum(x-\bar{x})^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1}$

Probability

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A|B) = \frac{P(A\cap B)}{P(B)}, P(B) > 0$
- $P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)}, \quad j = 1, 2, \ldots, k$

Random Variables

- $\mu = E(X) = \sum x\, P(X = x)$
- $\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$
- $\sigma^2 = E(X-\mu)^2 = E(X^2) - \left(E(X)\right)^2$
- Binomial

$$P(X = x) = C_x^n p^x (1-p)^{n-x}, \;\; x = 0, 1, 2, \ldots, n;$$

$$\mu = np \;\;\&\;\; \sigma^2 = np(1-p)$$

- Hypergeometric

$$P(X = x) = \frac{C_x^K C_{n-x}^{N-K}}{C_n^N}, \;\; x = 0, \cdots, min\{K, n\}$$

$$\mu = n\frac{K}{N} \;\;\&\;\; \sigma^2 = n\frac{K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1}$$

- Poisson

$$P(X = x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}, \;\; x = 0, 1, 2, \ldots \;;$$

$$\mu = \lambda t \quad \& \quad \sigma^2 = \lambda t$$

- Uniform

$$f(x) = \frac{1}{b-a}, \quad a \le x \le b; \quad \mu = \frac{a+b}{2} \quad \& \quad \sigma^2 = \frac{(b-a)^2}{12}$$

- Exponential

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0; \mu = \frac{1}{\lambda} \quad \& \quad \sigma^2 = \frac{1}{\lambda^2}$$

- Normal

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Weibull

$$f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-\left(\frac{x}{\delta}\right)^{\beta}}, \quad x > 0; \;\; F(x) = 1 - e^{-\left(\frac{x}{\delta}\right)^{\beta}}; \; \mu = \delta\Gamma\left(1 + \frac{1}{\beta}\right)$$

- Lognormal

$$f(x) = \frac{1}{\sqrt{2\pi}x\omega} e^{-\frac{1}{2}\left(\frac{\ln x - \theta}{\omega}\right)^2} \; 0 < x < \infty; \; \mu = e^{\theta + \omega^2/2}$$

Confidence Interval Estimation

$$\bar{x} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \quad or \quad \bar{x} \pm z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \quad or \quad \bar{x} \pm t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}$$

Required sample size

$$n \ge \left(\frac{z_{\frac{\alpha}{2}}\sigma}{e}\right)^2, \quad or \;\; n \ge \left(\frac{z_{\frac{\alpha}{2}}s}{e}\right)^2,$$

$$or \quad n \ge \left(\frac{z_{\frac{\alpha}{2}}}{e}\right)^2 p(1-p)$$

Hypothesis Testing

Test statistics

$$z = \frac{(\bar{x}-\mu_0)}{\sigma/\sqrt{n}} \;\; \text{or} \;\; t = \frac{(\bar{x}-\mu_0)}{s/\sqrt{n}} \;\; \text{or} \;\; z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

Hypothesis type	P – value	
Lower tail	$P(Z < z)$	$P(T_v < t)$
Upper tail	$P(Z > z)$	$P(T_v > t)$
2-tailed	$2P(Z > \lvert z\rvert)$	$2P(T_v > \lvert t\rvert)$

Sample correlation coefficient

$$r = \frac{Sxy}{\sqrt{SxxSyy}} \quad where \quad Sxx = \sum(x-\bar{x})^2$$

$Syy = \sum(y-\bar{y})^2$ and $Sxy = \sum(x-\bar{x})(y-\bar{y})$

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x$$

$$where \;\; \widehat{\beta_1} = \frac{Sxy}{Sxy} \;\;\&\;\; \widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$$

Simple Linear Regression

$$SST = Syy = \sum(y-\bar{y})^2 = \sum y^2 - n\bar{y}^2$$

$$SSR = \widehat{\beta_1}Sxy \quad \& \quad SSE = SST - SSR$$

Coefficient of Determination

$R^2 = \frac{SSR}{SST}$ and $R_{adj}^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-k-1}\right)$

$$\hat{\sigma} = MSE = \sqrt{\frac{SSE}{n-p}}$$

$$s.e(\widehat{\beta_1}) = \frac{\hat{\sigma}}{\sqrt{Sxx}}$$

Inferences about β_1

$$\widehat{\beta_1} \pm t_{\frac{\alpha}{2}} s.e(\widehat{\beta_1})$$

$H_0: \beta_1 = \beta_{10}$ vs $H_1: \beta_1 \ne \beta_{10}$, $t = \frac{\widehat{\beta_1} - \beta_{10}}{s.e(\widehat{\beta_1})}$

C.I. for the **mean** of y given a particular x_0

$$\hat{y} \pm t_{\frac{\alpha}{2}}\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{Sxx}}$$

P.I. estimate for an **Individual value of y** at a particular x_p

$$\hat{y} \pm t_{\frac{\alpha}{2}}\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_p-\bar{x})^2}{Sxx}}$$

Multiple Linear Regression

$\hat{\beta} = (X'X)^{-1}X'y$

$Var(\hat{\beta}) = (X'X)^{-1}\sigma^2$

$SSR = \hat{\beta}X'y$

$H_0: \beta_1 = \beta_2 = \cdots = \beta_k$ vs H_1: at least one $\beta_i \ne 0$

Test statistic : $F = \frac{MSR}{MSE}$

Inferences for individual parameters:

Similar to Simple Linear Regression

Inferences for a subset of the parameters:

$$\text{Partial F} = \frac{\dfrac{[SSR(Full) - SSR(Reduced)]}{number\ of\ parameters\ tested}}{MSE(Full)}$$