

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

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

Random Variables

- $\mu = E(X) = \sum x P(X = x)$
- $\mu = E(X) = \int_{-\infty}^{\infty} xf(x) dx$
- $\sigma^2 = E(X - \mu)^2 = E(X^2) - (E(X))^2$
- Binomial
 $P(X = x) = C_x^n p^x (1-p)^{n-x}, x = 0, 1, 2, \dots, n;$

$$\mu = np \text{ \& } \sigma^2 = np(1-p)$$

- Hypergeometric

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

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

- Poisson

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

$$\mu = \lambda t \text{ \& } \sigma^2 = \lambda t$$

- Uniform

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

- Exponential

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

- Normal

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

- Weibull

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

- Lognormal

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

Confidence Interval Estimation

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

Required sample size

$$n \geq \left(\frac{z_{\frac{\alpha}{2}} \sigma}{e}\right)^2, \text{ or } n \geq \left(\frac{z_{\frac{\alpha}{2}} s}{e}\right)^2,$$

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

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

P-value

type

Lower tail	$P(Z < z)$	$P(T_v < t)$
Upper tail	$P(Z > z)$	$P(T_v > t)$
2-tailed	$2P(Z > z)$	$2P(T_v > t)$

Sample correlation coefficient

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \text{ where } S_{xx} = \sum (x - \bar{x})^2$$

$$S_{yy} = \sum (y - \bar{y})^2 \text{ and } S_{xy} = \sum (x - \bar{x})(y - \bar{y})$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\text{where } \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \text{ \& } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Simple Linear Regression

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

$$SSR = \hat{\beta}_1 S_{xy} \text{ \& } SSE = SST - SSR$$

Coefficient of Determination

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

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

$$s.e(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{S_{xx}}}$$

Inferences about β_1

$$\hat{\beta}_1 \pm t_{\frac{\alpha}{2}} s.e(\hat{\beta}_1)$$

$$H_0: \beta_1 = \beta_{10} \text{ vs } H_1: \beta_1 \neq \beta_{10}, t = \frac{\hat{\beta}_1 - \beta_{10}}{s.e(\hat{\beta}_1)}$$

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

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

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

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

Multiple Linear Regression

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

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

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

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k \text{ vs } H_1: \text{at least one } \beta_i \neq 0$$

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

Inferences for individual parameters:

Similar to Simple Linear Regression

Inferences for a subset of the parameters:

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