Statistics for Management and Economics, Sixth Edition

Formulas

Numerical Descriptive Techniques

Population mean

\[
\mu = \frac{\sum_{i=1}^{N} x_i}{N}
\]

Sample mean

\[
\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}
\]

Range

Largest observation - Smallest observation

Population variance

\[
\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}
\]

Sample variance

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

Population standard deviation

\[
\sigma = \sqrt{\sigma^2}
\]

Sample standard deviation

\[
s = \sqrt{s^2}
\]

Population covariance
\[
\text{COV}(X,Y) = \frac{\sum_{i=1}^{N} (x_i - \mu_X)(y_i - \mu_Y)}{N}
\]

Sample covariance

\[
\text{cov}(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n - 1}
\]

Population coefficient of correlation

\[
\rho = \frac{\text{COV}(X,Y)}{\sigma_X \sigma_Y}
\]

Sample coefficient of correlation

\[
r = \frac{\text{cov}(x,y)}{s_x s_y}
\]

Least Squares: Slope coefficient

\[
b_1 = \frac{\text{cov}(x,y)}{s_x^2}
\]

Least Squares: y-Intercept

\[
b_0 = \bar{y} - b_1 \bar{x}
\]

**Probability**

Conditional probability

\[
P(A|B) = \frac{P(A \text{ and } B)}{P(B)}
\]

Complement rule

\[
P(A^C) = 1 - P(A)
\]

Multiplication rule

\[
P(A \text{ and } B) = P(A|B)P(B)
\]

Addition rule

\[
P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)
\]
Random Variables and Discrete Probability Distributions

Expected value (mean)

\[ E(X) = \mu = \sum_{x} x p(x) \]

Variance

\[ V(x) = \sigma^2 = \sum_{x} (x - \mu)^2 p(x) \]

Standard deviation

\[ \sigma = \sqrt{\sigma^2} \]

Covariance

\[ \text{COV}(X, Y) = \sum (x - \mu_x)(y - \mu_y) p(x, y) \]

Coefficient of Correlation

\[ \rho = \frac{\text{COV}(X, Y)}{\sigma_x \sigma_y} \]

Laws of expected value

1. \( E(c) = c \)
2. \( E(X + c) = E(X) + c \)
3. \( E(cX) = cE(X) \)

Laws of variance

1. \( V(c) = 0 \)
2. \( V(X + c) = V(X) \)
3. \( V(cX) = c^2 V(X) \)

Laws of expected value and variance of the sum of two variables

1. \( E(X + Y) = E(X) + E(Y) \)
2. \( V(X + Y) = V(X) + V(Y) + 2\text{COV}(X, Y) \)
Laws of expected value and variance for the sum of more than two independent variables

1. \[ E(\sum_{i=1}^{k} X_i) = \sum_{i=1}^{k} E(X_i) \]

2. \[ V(\sum_{i=1}^{k} X_i) = \sum_{i=1}^{k} V(X_i) \]

Mean and variance of a portfolio of two stocks

\[ E(R_p) = w_1 E(R_1) + w_2 E(R_2) \]

\[ V(R_p) = w_1^2 V(R_1) + w_2^2 V(R_2) + 2 w_1 w_2 \text{COV}(R_1, R_2) \]

\[ = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \rho \sigma_1 \sigma_2 \]

Mean and variance of a portfolio of k stocks

\[ E(R_p) = \sum_{i=1}^{k} w_i E(R_i) \]

\[ V(R_p) = \sum_{i=1}^{k} w_i^2 \sigma_i^2 + 2\sum_{i=1}^{k} \sum_{j=i+1}^{k} w_i w_j \text{COV}(R_i, R_j) \]

Binomial probability

\[ P(X = x) = \frac{n!}{x!(n-x)!} p^x (1 - p)^{n-x} \]

\[ \mu = np \]

\[ \sigma^2 = np(1-p) \]

\[ \sigma = \sqrt{np(1-p)} \]

Poisson probability

\[ P(X = x) = \frac{e^{-\mu} \mu^x}{x!} \]
Continuous Probability Distributions

Expected value of the sample mean

\[ E(\bar{X}) = \mu_{\bar{X}} = \mu \]

Variance of the sample mean

\[ V(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \]

Standard error of the sample mean

\[ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \]

Standardizing the sample mean

\[ Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \]

Expected value of the sample proportion

\[ E(\hat{P}) = \mu_\hat{P} = p \]

Variance of the sample proportion

\[ V(\hat{P}) = \sigma_{\hat{P}}^2 = \frac{p(1-p)}{n} \]

Standard error of the sample proportion

\[ \sigma_{\hat{P}} = \sqrt{\frac{p(1-p)}{n}} \]

Standardizing the sample proportion

\[ Z = \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} \]

Expected value of the difference between two means

\[ E(\bar{X}_1 - \bar{X}_2) = \mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 \]

Variance of the difference between two means

\[ V(\bar{X}_1 - \bar{X}_2) = \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2} \]
Standard error of the difference between two means

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Standardizing the difference between two sample means

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

**Introduction to Estimation**

Confidence interval estimator of $\mu$

$$\bar{X} \pm z_{a/2} \frac{\sigma}{\sqrt{n}}$$

Sample size to estimate $\mu$

$$n = \left(\frac{z_{a/2} \sigma}{W}\right)^2$$

**Introduction to Hypothesis Testing**

Test statistic for $\mu$

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

**Inference about One Population**

Test statistic for $\mu$

$$t = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

Interval estimator of $\mu$

$$\bar{X} \pm t_{a/2} \frac{s}{\sqrt{n}}$$
Test statistic for $\sigma^2$

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2}$$

Interval Estimator of $\sigma^2$

$$\text{LCL} = \frac{(n - 1)s^2}{\chi^2_{a/2}}$$

$$\text{UCL} = \frac{(n - 1)s^2}{\chi^2_{1-a/2}}$$

Test statistic for $p$

$$z = \frac{\hat{p} - p}{\sqrt{p(1 - p) / n}}$$

Interval estimator of $p$

$$\hat{p} \pm z_{a/2} \sqrt{\hat{p}(1 - \hat{p}) / n}$$

Sample size to estimate $p$

$$n = \left( \frac{z_{a/2} \sqrt{\hat{p}(1 - \hat{p})}}{W} \right)^2$$

**Inference about Two Populations**

Equal-variances t-test of $\mu_1 - \mu_2$

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

$$\nu = n_1 + n_2 - 2$$

Equal-variances interval estimator of $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{a/2} \sqrt{s^2_p \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$\nu = n_1 + n_2 - 2$$
Unequal-variances t-test of $\mu_1 - \mu_2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$  \hspace{1cm} \nu = \frac{(s_1^2 / n_1 + s_2^2 / n_2)^2}{\left(\frac{(s_1^2 / n_1)^2}{n_1 - 1} + \frac{(s_2^2 / n_2)^2}{n_2 - 1}\right)}$$

Unequal-variances interval estimator of $\mu_1 - \mu_2$

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$  \hspace{1cm} \nu = \frac{(s_1^2 / n_1 + s_2^2 / n_2)^2}{\left(\frac{(s_1^2 / n_1)^2}{n_1 - 1} + \frac{(s_2^2 / n_2)^2}{n_2 - 1}\right)}$$

t-Test of $\mu_D$

$$t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n_D}}$$  \hspace{1cm} \nu = n_D - 1$$

t-Estimator of $\mu_D$

$$\bar{x}_D \pm t_{\alpha/2} \frac{s_D}{\sqrt{n_D}}$$  \hspace{1cm} \nu = n_D - 1$$

F-test of $\sigma_1^2 / \sigma_2^2$

$$F = \frac{s_1^2}{s_2^2}$$  \hspace{1cm} \nu_1 = n_1 - 1 \text{ and } \nu_2 = n_2 - 1$$

F-Estimator of $\sigma_1^2 / \sigma_2^2$

$$\begin{align*}
\text{LCL} &= \left(\frac{s_1^2}{s_2^2}\right) \frac{1}{F_{n_1 - 1, n_2 - 1}} \\
\text{UCL} &= \left(\frac{s_1^2}{s_2^2}\right) F_{n_1 - 1, n_2 - 1}
\end{align*}$$

z-Test and estimator of $p_1 - p_2$

Case 1: $z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
Case 2:  
\[ z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \]

z-Interval estimator of \( p_1 - p_2 \)

\[ (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \]

Analysis of Variance

One-Way Analysis of variance

\[ \text{SST} = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2 \]

\[ \text{SSE} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \]

\[ \text{MST} = \frac{\text{SST}}{k - 1} \]

\[ \text{MSE} = \frac{\text{SSE}}{n - k} \]

\[ F = \frac{\text{MST}}{\text{MSE}} \]

Two-way analysis of Variance (randomized block design of experiment)

\[ \text{SS(Total)} = \sum_{j=1}^{k} \sum_{i=1}^{b} (x_{ij} - \bar{x})^2 \]

\[ \text{SST} = \sum_{i=1}^{k} b(\bar{x}[T]_i - \bar{x})^2 \]

\[ \text{SSB} = \sum_{i=1}^{b} k(\bar{x}[B]_i - \bar{x})^2 \]

\[ \text{SSE} = \sum_{j=1}^{k} \sum_{i=1}^{b} (x_{ij} - \bar{x}[T]_j - \bar{x}[B]_i + \bar{x})^2 \]
MST = \frac{SST}{k - 1}

MSB = \frac{SSB}{b - 1}

MSE = \frac{SSE}{n - k - b + 1}

F = \frac{MST}{MSE}

F = \frac{MSB}{MSE}

F = \frac{MS(A)}{MSE}

F = \frac{MS(B)}{MSE}

F = \frac{MS(AB)}{MSE}

Two-factor experiment

SS(Total) = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} (x_{ijk} - \bar{x})^2

SS(A) = rb \sum_{i=1}^{a} (\bar{x}[A]_i - \bar{x})^2

SS(B) = ra \sum_{j=1}^{b} (\bar{x}[B]_j - \bar{x})^2

SS(AB) = r \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{x}[AB]_{ij} - \bar{x}[A]_i - \bar{x}[B]_j + \bar{x})^2

SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} (x_{ijk} - \bar{x}[AB]_{ij})^2
Least Significant Difference Comparison Method

$$\text{LSD} = t_{\alpha/2} \sqrt{\text{MSE} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

Tukey's multiple comparison method

$$\omega = q_{\alpha} (k, \nu) \sqrt{\frac{\text{MSE}}{n_g}}$$

Chi-Squared Tests

Test statistic for all procedures

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

Nonparametric Statistical Techniques

Wilcoxon rank sum test statistic

$$T = T_1$$

$$\text{E}(T) = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma_T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

$$z = \frac{T - \text{E}(T)}{\sigma_T}$$

Sign test statistic

$$x = \text{number of positive differences}$$

$$z = \frac{x - .5n}{.5\sqrt{n}}$$

Wilcoxon signed rank sum test statistic

$$T = T^*$$
\[ E(T) = \frac{n(n+1)}{4} \]

\[ \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} \]

\[ z = \frac{T - E(T)}{\sigma_T} \]

**Kruskal-Wallis Test**

\[ H = \left[ \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{T_j^2}{n_j} \right] - 3(n+1) \]

**Friedman Test**

\[ F_r = \left[ \frac{12}{b(k)(k+1)} \sum_{j=1}^{k} T_j^2 \right] - 3b(k+1) \]

**Simple Linear Regression**

Sample slope

\[ b_1 = \frac{\text{cov}(x,y)}{s_x^2} \]

Sample y-intercept

\[ b_0 = \bar{y} - b_1 \bar{x} \]

Sum of squares for error

\[ SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

Standard error of estimate

\[ s_e = \sqrt{\frac{SSE}{n-2}} \]

Test statistic for the slope

\[ t = \frac{b_1 - \beta_1}{s_{b_1}} \]
Standard error of $b_1$

$$s_{b_1} = \frac{s_y}{\sqrt{(n-1)s_x^2}}$$

Coefficient of determination

$$R^2 = \frac{[\text{cov}(x, y)]^2}{s_x^2 s_y^2} = 1 - \frac{SSE}{\sum (y_i - \bar{y})^2}$$

Prediction interval

$$\hat{y} \pm t_{n-2, n-2} s_y \sqrt{1 + \frac{(x_{g} - \bar{x})^2}{n} + \frac{1}{(n-1)s_x^2}}$$

Confidence interval estimator of the expected value of $y$

$$\hat{y} \pm t_{n-2, n-2} s_y \sqrt{1 + \frac{(x_{g} - \bar{x})^2}{n} + \frac{1}{(n-1)s_x^2}}$$

Sample coefficient of correlation

$$r = \frac{\text{cov}(x, y)}{s_x s_y}$$

Test statistic for testing $\rho = 0$

$$t = r \sqrt{\frac{n-2}{1 - r^2}}$$

Sample Spearman rank correlation coefficient

$$r_s = \frac{\text{cov}(a, b)}{s_a s_b}$$

Test statistic for testing $\rho_s = 0$ when $n > 30$

$$z = \frac{r_s - 0}{1/\sqrt{n-1}} = r_s \sqrt{n-1}$$
Multiple Regression

Standard Error of Estimate

\[ s_e = \sqrt{\frac{SSE}{n - k - 1}} \]

Test statistic for \( \beta_i \)

\[ t = \frac{b_i - \beta_i}{s_{b_i}} \]

Coefficient of Determination

\[ R^2 = \frac{[\text{cov}(x, y)]^2}{s_x^2 s_y^2} = 1 - \frac{SSE}{\sum (y_i - \bar{y})^2} \]

Adjusted Coefficient of Determination

\[
\text{Adjusted } R^2 = 1 - \frac{SSE/(n - k - 1)}{\sum (y_i - \bar{y})^2 / (n - 1)}
\]

Mean Square for Error

\[ \text{MSE} = \frac{SSE}{k} \]

Mean Square for Regression

\[ \text{MSR} = \frac{SSR}{(n-k-1)} \]

F-statistic

\[ F = \frac{\text{MSR}}{\text{MSE}} \]

Durbin-Watson statistic

\[ d = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \]

Time Series Analysis and Forecasting

Exponential smoothing

\[ S_t = w y_t + (1 - w) S_{t-1} \]
**Statistical Process Control**

Centerline and control limits for $\bar{x}$ chart using $S$

Centerline = $\bar{x}$

Lower control limit = $\bar{x} - 3 \frac{S}{\sqrt{n}}$

Upper control limit = $\bar{x} + 3 \frac{S}{\sqrt{n}}$

Centerline and control limits for the $p$ chart

Centerline = $\bar{p}$

Lower control limit = $\bar{p} - 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$

Upper control limit = $\bar{p} + 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$

**Decision Analysis**

Expected Value of perfect Information

$$EVPI = EPPI - EMV^*$$

Expected Value of Sample Information

$$EVSI = EMV^* - EMV^*$$