

Important Formulas

Chapter 3 Data Description

Mean for individual data: $\bar{X} = \frac{\sum X}{n}$

Mean for grouped data: $\bar{X} = \frac{\sum f \cdot X_m}{n}$

Standard deviation for a sample:

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}} \quad \text{or} \quad s = \sqrt{\frac{n(\sum X^2) - (\sum X)^2}{n(n - 1)}}$$

(Shortcut formula)

Standard deviation for grouped data:

$$s = \sqrt{\frac{n(\sum f \cdot X_m^2) - (\sum f \cdot X_m)^2}{n(n - 1)}}$$

Range rule of thumb: $s \approx \frac{\text{range}}{4}$

Chapter 4 Probability and Counting Rules

Addition rule 1 (mutually exclusive events):

$$P(A \text{ or } B) = P(A) + P(B)$$

Addition rule 2 (events not mutually exclusive):

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Multiplication rule 1 (independent events):

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

Multiplication rule 2 (dependent events):

$$P(A \text{ and } B) = P(A) \cdot P(B | A)$$

Conditional probability: $P(B | A) = \frac{P(A \text{ and } B)}{P(A)}$

Complementary events: $P(\bar{E}) = 1 - P(E)$

Fundamental counting rule: Total number of outcomes of a sequence when each event has a different number of possibilities: $k_1 \cdot k_2 \cdot k_3 \cdot \dots \cdot k_n$

Permutation rule: Number of permutations of n objects

$$\text{taking } r \text{ at a time is } {}_n P_r = \frac{n!}{(n - r)!}$$

Permutation rule of n objects with r_1 objects identical,

r_2 objects identical, etc.

$$\frac{n!}{r_1! r_2! \cdot \dots \cdot r_p!}$$

Combination rule: Number of combinations of r objects

$$\text{selected from } n \text{ objects is } {}_n C_r = \frac{n!}{(n - r)! r!}$$

Chapter 5 Discrete Probability Distributions

Mean for a probability distribution: $\mu = \sum [X \cdot P(X)]$

Variance and standard deviation for a probability distribution:

$$\sigma^2 = \sum [X^2 \cdot P(X)] - \mu^2$$

$$\sigma = \sqrt{\sum [X^2 \cdot P(X)] - \mu^2}$$

Expectation: $E(X) = \sum [X \cdot P(X)]$

Binomial probability: $P(X) = \frac{n!}{(n - X)! X!} \cdot p^X \cdot q^{n - X}$

Mean for binomial distribution: $\mu = n \cdot p$

Variance and standard deviation for the binomial distribution:

$$\sigma^2 = n \cdot p \cdot q \quad \sigma = \sqrt{n \cdot p \cdot q}$$

Multinomial probability:

$$P(X) = \frac{n!}{X_1! X_2! X_3! \cdot \dots \cdot X_k!} \cdot p_1^{X_1} \cdot p_2^{X_2} \cdot p_3^{X_3} \cdot \dots \cdot p_k^{X_k}$$

Poisson probability: $P(X; \lambda) = \frac{e^{-\lambda} \lambda^X}{X!}$ where

$$X = 0, 1, 2, \dots$$

Hypergeometric probability: $P(X) = \frac{{}_a C_X \cdot {}_b C_{n - X}}{{}_{a + b} C_n}$

Geometric probability: $P(n) = p(1 - p)^{n - 1}$ where n is the number of the trial in which the first success occurs and p is the probability of a success

Chapter 6 The Normal Distribution

Standard score: Population: $z = \frac{X - \mu}{\sigma}$ or Sample: $z = \frac{X - \bar{X}}{s}$

Mean of sample means: $\mu_{\bar{X}} = \mu$

Standard error of the mean: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Central limit theorem formula: $z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$

Chapter 7 Confidence Intervals and Sample Size

z confidence interval for means:

$$\bar{X} - z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right) < \mu < \bar{X} + z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right)$$

t confidence interval for means:

$$\bar{X} - t_{\alpha/2} \left(\frac{s}{\sqrt{n}} \right) < \mu < \bar{X} + t_{\alpha/2} \left(\frac{s}{\sqrt{n}} \right)$$

Sample size for means: $n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2$ where E is the margin of error

Confidence interval for a proportion:

$$\hat{p} - (z_{\alpha/2}) \sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + (z_{\alpha/2}) \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Sample size for a proportion: $n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$

where $\hat{p} = \frac{X}{n}$ and $\hat{q} = 1 - \hat{p}$

Confidence interval for variance:

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}}$$

Confidence interval for standard deviation:

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\text{right}}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{\text{left}}}}$$

Chapter 8 Hypothesis Testing

z test: $z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ for any value n . If $n < 30$, population must be normally distributed.

t test: $t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ (d.f. = $n - 1$)

z test for proportions: $z = \frac{\hat{p} - p}{\sqrt{pq/n}}$

Chi-square test for a single variance: $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ (d.f. = $n - 1$)

Chapter 9 Testing the Difference Between Two Means, Two Proportions, and Two Variances

z test for comparing two means (independent samples):

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

Formula for the confidence interval for difference of two means (large samples):

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2$$

$$< (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

t test for comparing two means (independent samples, variances not equal):

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

(d.f. = the smaller of $n_1 - 1$ or $n_2 - 1$)

Formula for the confidence interval for difference of two means (small independent samples, variances unequal):

$$(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2$$

$$< (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(d.f. = smaller of $n_1 - 1$ and $n_2 - 1$)

t test for comparing two means for dependent samples:

$$t = \frac{\bar{D} - \mu_D}{s_D/\sqrt{n}} \quad \text{where} \quad \bar{D} = \frac{\sum D}{n} \quad \text{and}$$

$$s_D = \sqrt{\frac{n \sum D^2 - (\sum D)^2}{n(n-1)}} \quad (\text{d.f.} = n - 1)$$

Formula for confidence interval for the mean of the difference for dependent samples:

$$\bar{D} - t_{\alpha/2} \frac{s_D}{\sqrt{n}} < \mu_D < \bar{D} + t_{\alpha/2} \frac{s_D}{\sqrt{n}}$$

(d.f. = $n - 1$)

z test for comparing two proportions:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\text{where} \quad \bar{p} = \frac{X_1 + X_2}{n_1 + n_2} \quad \hat{p}_1 = \frac{X_1}{n_1}$$

$$\bar{q} = 1 - \bar{p} \quad \hat{p}_2 = \frac{X_2}{n_2}$$

Formula for the confidence interval for the difference of two proportions:

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < p_1 - p_2$$

$$< (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

F test for comparing two variances: $F = \frac{s_1^2}{s_2^2}$ where s_1^2 is the

larger variance and d.f.N. = $n_1 - 1$, d.f.D. = $n_2 - 1$

Chapter 10 Correlation and Regression

Correlation coefficient:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$

t test for correlation coefficient: $t = r \sqrt{\frac{n-2}{1-r^2}}$
(d.f. = $n - 2$)

The regression line equation: $y' = a + bx$

where $a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Coefficient of determination: $r^2 = \frac{\text{explained variation}}{\text{total variation}}$

Standard error of estimate:

$$s_{\text{est}} = \sqrt{\frac{\sum y^2 - a \sum y - b \sum xy}{n - 2}}$$

Prediction interval for y :

$$y' - t_{\alpha/2} s_{\text{est}} \sqrt{1 + \frac{1}{n} + \frac{n(x - \bar{X})^2}{n \sum x^2 - (\sum x)^2}}$$

$$< y < y' + t_{\alpha/2} s_{\text{est}} \sqrt{1 + \frac{1}{n} + \frac{n(x - \bar{X})^2}{n \sum x^2 - (\sum x)^2}}$$

(d.f. = $n - 2$)

Formula for the multiple correlation coefficient:

$$R = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

Formula for the F test for the multiple correlation coefficient:

$$F = \frac{R^2/k}{(1 - R^2)/(n - k - 1)}$$

(d.f.N. = $n - k$ and d.f.D. = $n - k - 1$)

Formula for the adjusted R^2 :

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

Chapter 11 Other Chi-Square Tests

Chi-square test for goodness-of-fit:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

(d.f. = no. of categories - 1)

Chi-square test for independence and homogeneity of proportions:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

[d.f. = (rows - 1)(columns - 1)]

Chapter 12 Analysis of Variance

ANOVA test: $F = \frac{s_B^2}{s_W^2}$ where $\bar{X}_{GM} = \frac{\sum X}{N}$

d.f.N. = $k - 1$ where $N = n_1 + n_2 + \dots + n_k$

d.f.D. = $N - k$ where k = number of groups

$$s_B^2 = \frac{\sum n_i (\bar{X}_i - \bar{X}_{GM})^2}{k - 1}$$

$$s_W^2 = \frac{\sum (n_i - 1) s_i^2}{\sum (n_i - 1)}$$

Scheffé test: $F_S = \frac{(\bar{X}_i - \bar{X}_j)^2}{s_W^2(1/n_i + 1/n_j)}$ and

$$F' = (k - 1)(C.V.)$$

Tukey test: $q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{s_W^2/n}}$

Formulas for two-way ANOVA:

$$MS_A = \frac{SS_A}{a - 1}$$

$$F_A = \frac{MS_A}{MS_W}$$

$$MS_B = \frac{SS_B}{b - 1}$$

$$F_B = \frac{MS_B}{MS_W}$$

$$MS_{A \times B} = \frac{SS_{A \times B}}{(a - 1)(b - 1)}$$

$$F_{A \times B} = \frac{MS_{A \times B}}{MS_W}$$

$$MS_W = \frac{SS_W}{ab(n - 1)}$$

Chapter 13 Nonparametric Statistics

$$z \text{ test value in the sign test: } z = \frac{(X + 0.5) - 0.5n}{\sqrt{n/2}}$$

where n = sample size (greater than or equal to 26)

X = smaller number of + or - signs

$$\text{Wilcoxon rank sum test: } z = \frac{R - \mu_R}{\sigma_R}$$

where

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2}$$

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

R = sum of the ranks for the smaller sample size (n_1)

n_1 = smaller of the sample sizes

n_2 = larger of the sample sizes

$n_1 \geq 10$ and $n_2 \geq 10$

$$\text{Wilcoxon signed-rank test: } z = \frac{w_s - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

where

n = number of pairs where the difference is not 0
and $n \geq 30$

w_s = smaller sum in absolute value of the signed ranks

Kruskal-Wallis test:

$$H = \frac{12}{N(N+1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \cdots + \frac{R_k^2}{n_k} \right) - 3(N+1)$$

where

R_1 = sum of the ranks of sample 1

n_1 = size of sample 1

R_2 = sum of the ranks of sample 2

n_2 = size of sample 2

$\vdots$

R_k = sum of the ranks of sample k

n_k = size of sample k

$N = n_1 + n_2 + \cdots + n_k$

k = number of samples

Spearman rank correlation coefficient:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

where

d = difference in the ranks

n = number of data pairs

Procedure Table

Solving Hypothesis-Testing Problems (Traditional Method)

- Step 1** State the hypotheses and identify the claim.
- Step 2** Find the critical value(s) from the appropriate table in Appendix A.
- Step 3** Compute the test value.
- Step 4** Make the decision to reject or not reject the null hypothesis.
- Step 5** Summarize the results.

Procedure Table

Solving Hypothesis-Testing Problems (P-Value Method)

- Step 1** State the hypotheses and identify the claim.
- Step 2** Compute the test value.
- Step 3** Find the P -value.
- Step 4** Make the decision.
- Step 5** Summarize the results.