INTRODUCTORY STATISTICS, 6/E

Neil A. Weiss

FORMULAS

NOTATION The following notation is used on this card:

n = sample size	σ = population stdev
$\bar{x}$ = sample mean	d = paired difference
s = sample stdev	$\hat{p}$ = sample proportion
Q_j = jth quartile	p = population proportion
N = population size	O = observed frequency
μ = population mean	E = expected frequency

CHAPTER 3 Descriptive Measures

- Sample mean: $\bar{x} = \dfrac{\Sigma x}{n}$
- Range: Range = Max − Min
- Sample standard deviation:

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} \quad \text{or} \quad s = \sqrt{\frac{\Sigma x^2 - (\Sigma x)^2/n}{n-1}}$$

- Interquartile range: IQR $= Q_3 - Q_1$
- Lower limit $= Q_1 - 1.5 \cdot$ IQR, Upper limit $= Q_3 + 1.5 \cdot$ IQR
- Population mean (mean of a variable): $\mu = \dfrac{\Sigma x}{N}$
- Population standard deviation (standard deviation of a variable):

$$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}} \quad \text{or} \quad \sigma = \sqrt{\frac{\Sigma x^2}{N} - \mu^2}$$

- Standardized variable: $z = \dfrac{x-\mu}{\sigma}$

CHAPTER 4 Probability Concepts

- Probability for equally likely outcomes:

$$P(E) = \frac{f}{N},$$

  where f denotes the number of ways event E can occur and N denotes the total number of outcomes possible.

- Special addition rule:

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

  (A, B, C, ... mutually exclusive)

- Complementation rule: $P(E) = 1 - P(\text{not } E)$
- General addition rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \,\&\, B)$
- Conditional probability rule: $P(B \mid A) = \dfrac{P(A \,\&\, B)}{P(A)}$
- General multiplication rule: $P(A \,\&\, B) = P(A) \cdot P(B \mid A)$
- Special multiplication rule:

$$P(A \,\&\, B \,\&\, C \,\&\, \cdots) = P(A) \cdot P(B) \cdot P(C) \cdots$$

  (A, B, C, ... independent)

- Rule of total probability:

$$P(B) = \sum_{j=1}^{k} P(A_j) \cdot P(B \mid A_j)$$

  ($A_1, A_2, \ldots, A_k$ mutually exclusive and exhaustive)

- Bayes's rule:

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A_i)}{\sum_{j=1}^{k} P(A_j) \cdot P(B \mid A_j)}$$

  ($A_1, A_2, \ldots, A_k$ mutually exclusive and exhaustive)

- Factorial: $k! = k(k-1)\cdots 2 \cdot 1$
- Permutations rule: ${}_mP_r = \dfrac{m!}{(m-r)!}$
- Special permutations rule: ${}_mP_m = m!$
- Combinations rule: ${}_mC_r = \dfrac{m!}{r!\,(m-r)!}$
- Number of possible samples: ${}_NC_n = \dfrac{N!}{n!\,(N-n)!}$

CHAPTER 5 Discrete Random Variables

- Mean of a discrete random variable X: $\mu = \Sigma x P(X = x)$
- Standard deviation of a discrete random variable X:

$$\sigma = \sqrt{\Sigma(x-\mu)^2 P(X=x)} \quad \text{or} \quad \sigma = \sqrt{\Sigma x^2 P(X=x) - \mu^2}$$

- Factorial: $k! = k(k-1)\cdots 2 \cdot 1$
- Binomial coefficient: $\dbinom{n}{x} = \dfrac{n!}{x!\,(n-x)!}$
- Binomial probability formula:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x},$$

  where n denotes the number of trials and p denotes the success probability.

- Mean of a binomial random variable: $\mu = np$
- Standard deviation of a binomial random variable: $\sigma = \sqrt{np(1-p)}$
- Poisson probability formula: $P(X = x) = e^{-\lambda} \dfrac{\lambda^x}{x!}$
- Mean of a Poisson random variable: $\mu = \lambda$
- Standard deviation of a Poisson random variable: $\sigma = \sqrt{\lambda}$

CHAPTER 7 The Sampling Distribution of the Sample Mean

- Mean of the variable $\bar{x}$: $\mu_{\bar{x}} = \mu$
- Standard deviation of the variable $\bar{x}$: $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

INTRODUCTORY STATISTICS, 6/E

Neil A. Weiss

FORMULAS

CHAPTER 8 Confidence Intervals for One Population Mean

- Standardized version of the variable $\bar{x}$:

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

- z-interval for μ (σ known, normal population or large sample):

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- Margin of error for the estimate of μ: $E = z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$

- Sample size for estimating μ:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2,$$

rounded up to the nearest whole number.

- Studentized version of the variable $\bar{x}$:

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

- t-interval for μ (σ unknown, normal population or large sample):

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

with df $= n - 1$.

CHAPTER 9 Hypothesis Tests for One Population Mean

- z-test statistic for H_0: $\mu = \mu_0$ (σ known, normal population or large sample):

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- t-test statistic for H_0: $\mu = \mu_0$ (σ unknown, normal population or large sample):

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

with df $= n - 1$.

- Wilcoxon signed-rank test statistic for H_0: $\mu = \mu_0$ (symmetric population):

$$W = \text{sum of the positive ranks}$$

CHAPTER 10 Inferences for Two Population Means

- Pooled sample standard deviation:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

- Pooled t-test statistic for H_0: $\mu_1 = \mu_2$ (independent samples, normal populations or large samples, and equal population standard deviations):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{(1/n_1) + (1/n_2)}}$$

with df $= n_1 + n_2 - 2$.

- Pooled t-interval for $\mu_1 - \mu_2$ (independent samples, normal populations or large samples, and equal population standard deviations):

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \cdot s_p\sqrt{(1/n_1) + (1/n_2)}$$

with df $= n_1 + n_2 - 2$.

- Degrees of freedom for nonpooled-t procedures:

$$\Delta = \frac{\left[\left(s_1^2/n_1\right) + \left(s_2^2/n_2\right)\right]^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2 - 1}},$$

rounded down to the nearest integer.

- Nonpooled t-test statistic for H_0: $\mu_1 = \mu_2$ (independent samples, and normal populations or large samples):

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

with df $= \Delta$.

- Nonpooled t-interval for $\mu_1 - \mu_2$ (independent samples, and normal populations or large samples):

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \cdot \sqrt{(s_1^2/n_1) + (s_2^2/n_2)}$$

with df $= \Delta$.

- Mann–Whitney test statistic for H_0: $\mu_1 = \mu_2$ (independent samples and same-shape populations):

$$M = \text{sum of the ranks for sample data from Population 1}$$

- Paired t-test statistic for H_0: $\mu_1 = \mu_2$ (paired sample, and normal differences or large sample):

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

with df $= n - 1$.

- Paired t-interval for $\mu_1 - \mu_2$ (paired sample, and normal differences or large sample):

$$\bar{d} \pm t_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}$$

with df $= n - 1$.

- Paired Wilcoxon signed-rank test statistic for H_0: $\mu_1 = \mu_2$ (paired sample and symmetric differences):

$$W = \text{sum of the positive ranks}$$

INTRODUCTORY STATISTICS, 6/E

Neil A. Weiss

FORMULAS

CHAPTER 11 Inferences for Population Standard Deviations

- χ^2-test statistic for H_0: $\sigma = \sigma_0$ (normal population):

$$\chi^2 = \frac{n-1}{\sigma_0^2} s^2$$

with df $= n - 1$.

- χ^2-interval for σ (normal population):

$$\sqrt{\frac{n-1}{\chi^2_{\alpha/2}}} \cdot s \quad \text{to} \quad \sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}}} \cdot s$$

with df $= n - 1$.

- F-test statistic for H_0: $\sigma_1 = \sigma_2$ (independent samples and normal populations):

$$F = s_1^2/s_2^2$$

with df $= (n_1 - 1, n_2 - 1)$.

- F-interval for σ_1/σ_2 (independent samples and normal populations):

$$\frac{1}{\sqrt{F_{\alpha/2}}} \cdot \frac{s_1}{s_2} \quad \text{to} \quad \frac{1}{\sqrt{F_{1-\alpha/2}}} \cdot \frac{s_1}{s_2}$$

with df $= (n_1 - 1, n_2 - 1)$.

CHAPTER 12 Inferences for Population Proportions

- Sample proportion:

$$\hat{p} = \frac{x}{n},$$

where x denotes the number of members in the sample that have the specified attribute.

- One-sample z-interval for p:

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

(*Assumption:* both x and $n - x$ are 5 or greater)

- Margin of error for the estimate of p:

$$E = z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}$$

- Sample size for estimating p:

$$n = 0.25\left(\frac{z_{\alpha/2}}{E}\right)^2 \quad \text{or} \quad n = \hat{p}_g(1-\hat{p}_g)\left(\frac{z_{\alpha/2}}{E}\right)^2$$

rounded up to the nearest whole number (g = "educated guess")

- One-sample z-test statistic for H_0: $p = p_0$:

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

(*Assumption:* both np_0 and $n(1 - p_0)$ are 5 or greater)

- Pooled sample proportion: $\hat{p}_p = \dfrac{x_1 + x_2}{n_1 + n_2}$

- Two-sample z-test statistic for H_0: $p_1 = p_2$:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_p(1-\hat{p}_p)}\sqrt{(1/n_1) + (1/n_2)}}$$

(*Assumptions:* independent samples; $x_1, n_1 - x_1, x_2, n_2 - x_2$ are all 5 or greater)

- Two-sample z-interval for $p_1 - p_2$:

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

(*Assumptions:* independent samples; $x_1, n_1 - x_1, x_2, n_2 - x_2$ are all 5 or greater)

- Margin of error for the estimate of $p_1 - p_2$:

$$E = z_{\alpha/2} \cdot \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}$$

- Sample size for estimating $p_1 - p_2$:

$$n_1 = n_2 = 0.5\left(\frac{z_{\alpha/2}}{E}\right)^2$$

or

$$n_1 = n_2 = \left(\hat{p}_{1g}(1-\hat{p}_{1g}) + \hat{p}_{2g}(1-\hat{p}_{2g})\right)\left(\frac{z_{\alpha/2}}{E}\right)^2$$

rounded up to the nearest whole number (g = "educated guess")

CHAPTER 13 Chi-Square Procedures

- Expected frequencies for a chi-square goodness-of-fit test:

$$E = np$$

- Test statistic for a chi-square goodness-of-fit test:

$$\chi^2 = \Sigma(O - E)^2/E$$

with df $= k - 1$, where k is the number of possible values for the variable under consideration.

- Expected frequencies for a chi-square independence test:

$$E = \frac{R \cdot C}{n}$$

where R = row total and C = column total.

- Test statistic for a chi-square independence test:

$$\chi^2 = \Sigma(O - E)^2/E$$

with df $= (r - 1)(c - 1)$, where r and c are the number of possible values for the two variables under consideration.

CHAPTER 14 Descriptive Methods in Regression and Correlation

- S_{xx}, S_{xy}, and S_{yy}:

$$S_{xx} = \Sigma(x - \bar{x})^2 = \Sigma x^2 - (\Sigma x)^2/n$$

$$S_{xy} = \Sigma(x - \bar{x})(y - \bar{y}) = \Sigma xy - (\Sigma x)(\Sigma y)/n$$

$$S_{yy} = \Sigma(y - \bar{y})^2 = \Sigma y^2 - (\Sigma y)^2/n$$

- Regression equation: $\hat{y} = b_0 + b_1 x$, where

$$b_1 = \frac{S_{xy}}{S_{xx}} \quad \text{and} \quad b_0 = \frac{1}{n}(\Sigma y - b_1 \Sigma x) = \bar{y} - b_1\bar{x}$$

INTRODUCTORY STATISTICS, 6/E

Neil A. Weiss

FORMULAS

- Total sum of squares: $SST = \Sigma(y - \bar{y})^2 = S_{yy}$
- Regression sum of squares: $SSR = \Sigma(\hat{y} - \bar{y})^2 = S_{xy}^2/S_{xx}$
- Error sum of squares: $SSE = \Sigma(y - \hat{y})^2 = S_{yy} - S_{xy}^2/S_{xx}$
- Regression identity: $SST = SSR + SSE$
- Coefficient of determination: $r^2 = \dfrac{SSR}{SST}$
- Linear correlation coefficient:

$$r = \frac{\frac{1}{n-1}\Sigma(x - \bar{x})(y - \bar{y})}{s_x s_y} \quad \text{or} \quad r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

CHAPTER 15 Inferential Methods in Regression and Correlation

- Population regression equation: $y = \beta_0 + \beta_1 x$
- Standard error of the estimate: $s_e = \sqrt{\dfrac{SSE}{n-2}}$
- Test statistic for H_0: $\beta_1 = 0$:

$$t = \frac{b_1}{s_e/\sqrt{S_{xx}}}$$

  with df $= n - 2$.
- Confidence interval for β_1:

$$b_1 \pm t_{\alpha/2} \cdot \frac{s_e}{\sqrt{S_{xx}}}$$

  with df $= n - 2$.
- Confidence interval for the conditional mean of the response variable corresponding to x_p:

$$\hat{y}_p \pm t_{\alpha/2} \cdot s_e\sqrt{\frac{1}{n} + \frac{(x_p - \Sigma x/n)^2}{S_{xx}}}$$

  with df $= n - 2$.
- Prediction interval for an observed value of the response variable corresponding to x_p:

$$\hat{y}_p \pm t_{\alpha/2} \cdot s_e\sqrt{1 + \frac{1}{n} + \frac{(x_p - \Sigma x/n)^2}{S_{xx}}}$$

  with df $= n - 2$.
- Test statistic for H_0: $\rho = 0$:

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

  with df $= n - 2$.
- Test statistic for a correlation test for normality:

$$R_p = \frac{\Sigma xw}{\sqrt{S_{xx}\,\Sigma w^2}}$$

  where x and w denote observations of the variable and the corresponding normal scores, respectively.

CHAPTER 16 Analysis of Variance (ANOVA)

- Notation in one-way ANOVA:

  k = number of populations
  n = total number of observations
  $\bar{x}$ = mean of all n observations
  n_j = size of sample from Population j
  $\bar{x}_j$ = mean of sample from Population j
  s_j^2 = variance of sample from Population j
  T_j = sum of sample data from Population j

- Defining formulas for sums of squares in one-way ANOVA:

$$SST = \Sigma(x - \bar{x})^2$$
$$SSTR = \Sigma n_j(\bar{x}_j - \bar{x})^2$$
$$SSE = \Sigma(n_j - 1)s_j^2$$

- One-way ANOVA identity: $SST = SSTR + SSE$
- Computing formulas for sums of squares in one-way ANOVA:

$$SST = \Sigma x^2 - (\Sigma x)^2/n$$
$$SSTR = \Sigma(T_j^2/n_j) - (\Sigma x)^2/n$$
$$SSE = SST - SSTR$$

- Mean squares in one-way ANOVA:

$$MSTR = \frac{SSTR}{k-1}, \qquad MSE = \frac{SSE}{n-k}$$

- Test statistic for one-way ANOVA (independent samples, normal populations, and equal population standard deviations):

$$F = \frac{MSTR}{MSE}$$

  with df $= (k - 1, n - k)$.
- Confidence interval for $\mu_i - \mu_j$ in the Tukey multiple-comparison method (independent samples, normal populations, and equal population standard deviations):

$$(\bar{x}_i - \bar{x}_j) \pm \frac{q_\alpha}{\sqrt{2}} \cdot s\sqrt{(1/n_i) + (1/n_j)},$$

  where $s = \sqrt{MSE}$ and q_α is obtained for a q-curve with parameters k and $n - k$.
- Test statistic for a Kruskal–Wallis test (independent samples, same-shape populations, all sample sizes 5 or greater):

$$H = \frac{SSTR}{SST/(n-1)} \quad \text{or} \quad H = \frac{12}{n(n+1)}\sum_{j=1}^{k}\frac{R_j^2}{n_j} - 3(n+1),$$

  where $SSTR$ and SST are computed for the ranks of the data, and R_j denotes the sum of the ranks for the sample data from Population j. H is approximately chi-square with df $= k - 1$.