Key Formulas

From Larson/Farber *Elementary Statistics: Picturing the World,* Sixth Edition

CHAPTER 2

Class Width $= \dfrac{\text{Range of data}}{\text{Number of classes}}$

(round up to next convenient number)

Midpoint $= \dfrac{(\text{Lower class limit}) + (\text{Upper class limit})}{2}$

Relative Frequency $= \dfrac{\text{Class frequency}}{\text{Sample size}} = \dfrac{f}{n}$

Population Mean: $\mu = \dfrac{\sum x}{N}$

Sample Mean: $\bar{x} = \dfrac{\sum x}{n}$

Weighted Mean: $\bar{x} = \dfrac{\sum (x \cdot w)}{\sum w}$

Mean of a Frequency Distribution: $\bar{x} = \dfrac{\sum (x \cdot f)}{n}$

Range $=$ (Maximum entry) $-$ (Minimum entry)

Population Variance: $\sigma^2 = \dfrac{\sum (x - \mu)^2}{N}$

Population Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

Sample Variance: $s^2 = \dfrac{\sum (x - \bar{x})^2}{n - 1}$

Sample Standard Deviation: $s = \sqrt{s^2} = \sqrt{\dfrac{\sum (x - \bar{x})^2}{n - 1}}$

Empirical Rule (or 68-95-99.7 Rule) For data sets with distributions that are approximately symmetric and bell-shaped:

1. About 68% of the data lie within one standard deviation of the mean.
2. About 95% of the data lie within two standard deviations of the mean.
3. About 99.7% of the data lie within three standard deviations of the mean.

Chebychev's Theorem The portion of any data set lying within k standard deviations $(k > 1)$ of the mean is at least $1 - \dfrac{1}{k^2}$.

Sample Standard Deviation of a Frequency Distribution:

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

Standard Score: $z = \dfrac{\text{Value} - \text{Mean}}{\text{Standard deviation}} = \dfrac{x - \mu}{\sigma}$

CHAPTER 3

Classical (or Theoretical) Probability:

$$P(E) = \frac{\text{Number of outcomes in event } E}{\text{Total number of outcomes in sample space}}$$

Empirical (or Statistical) Probability:

$$P(E) = \frac{\text{Frequency of event } E}{\text{Total frequency}} = \frac{f}{n}$$

Probability of a Complement: $P(E') = 1 - P(E)$

Probability of occurrence of both events A and B:

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

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

Probability of occurrence of either A or B:

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

$P(A \text{ or } B) = P(A) + P(B)$ if A and B are mutually exclusive

Permutations of n objects taken r at a time:

$${}_nP_r = \frac{n!}{(n - r)!}, \text{where } r \le n$$

Distinguishable Permutations: n_1 alike, n_2 alike, $\ldots$, n_k alike:

$$\frac{n!}{n_1! \cdot n_2! \cdot n_3! \cdots n_k!},$$

where $n_1 + n_2 + n_3 + \cdots + n_k = n$

Combinations of n objects taken r at a time:

$${}_nC_r = \frac{n!}{(n - r)!r!}, \text{ where } r \le n$$

Key Formulas

From Larson/Farber *Elementary Statistics: Picturing the World,* Sixth Edition

Chapter 4

Mean of a Discrete Random Variable: $\mu = \sum xP(x)$

Variance of a Discrete Random Variable:

$$\sigma^2 = \sum (x - \mu)^2 P(x)$$

Standard Deviation of a Discrete Random Variable:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum (x - \mu)^2 P(x)}$$

Expected Value: $E(x) = \mu = \sum xP(x)$

Binomial Probability of x successes in n trials:

$$P(x) = {}_nC_x p^x q^{n-x} = \frac{n!}{(n-x)!x!} p^x q^{n-x}$$

Population Parameters of a Binomial Distribution:

Mean: $\mu = np$ Variance: $\sigma^2 = npq$

Standard Deviation: $\sigma = \sqrt{npq}$

Geometric Distribution: The probability that the first success will occur on trial number x is $P(x) = pq^{x-1}$, where $q = 1 - p$.

Poisson Distribution: The probability of exactly x occurrences in an interval is $P(x) = \frac{\mu^x e^{-\mu}}{x!}$, where $e \approx 2.71828$ and μ is the mean number of occurences per interval unit.

Chapter 5

Standard Score, or z-Score:

$$z = \frac{\text{Value} - \text{Mean}}{\text{Standard deviation}} = \frac{x - \mu}{\sigma}$$

Transforming a z-Score to an x-Value: $x = \mu + z\sigma$

Central Limit Theorem ($n \geq 30$ or population is normally distributed):

Mean of the Sampling Distribution: $\mu_{\bar{x}} = \mu$

Variance of the Sampling Distribution: $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$

Standard Deviation of the Sampling Distribution (Standard Error): $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$$z\text{-Score} = \frac{\text{Value} - \text{Mean}}{\text{Standard Error}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Chapter 6

c-Confidence Interval for μ: $\bar{x} - E < \mu < \bar{x} + E$, where $E = z_c \frac{\sigma}{\sqrt{n}}$ when σ is known, the sample is random, and either the population is normally distributed or $n \geq 30$, or $E = t_c \frac{s}{\sqrt{n}}$ when σ is unknown, the sample is random, and either the population is normally distributed or $n \geq 30$.

Minimum Sample Size to Estimate μ: $n = \left(\frac{z_c \sigma}{E}\right)^2$

Point Estimate for p, the population proportion of successes: $\hat{p} = \frac{x}{n}$

c-Confidence Interval for Population Proportion p (when $np \geq 5$ and $nq \geq 5$): $\hat{p} - E < p < \hat{p} + E$, where

$$E = z_c \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Minimum Sample Size to Estimate p: $n = \hat{p}\hat{q}\left(\frac{z_c}{E}\right)^2$

c-Confidence Interval for Population Variance σ^2:

$$\frac{(n-1)s^2}{\chi_R^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2}$$

c-Confidence Interval for Population Standard Deviation σ:

$$\sqrt{\frac{(n-1)s^2}{\chi_R^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi_L^2}}$$

Key Formulas

From Larson/Farber *Elementary Statistics: Picturing the World,* Sixth Edition

Chapter 7

z-Test for a Mean μ: $z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}}$, when σ is known, the sample is random, and either the population is normally distributed or $n \geq 30$.

t-Test for a Mean μ: $t = \dfrac{\overline{x} - \mu}{s/\sqrt{n}}$, when σ is unknown, the sample is random, and either the population is normally distributed or $n \geq 30$. (d.f. $= n - 1$)

z-Test for a Proportion p (when $np \geq 5$ and $nq \geq 5$):

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

Chi-Square Test for a Variance σ^2 or Standard Deviation σ:

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2} \quad (\text{d.f.} = n - 1)$$

Chapter 8

Two-Sample z-Test for the Difference Between Means (σ_1 and σ_2 are known, the samples are random and independent, and either the populations are normally distributed or both $n_1 \geq 30$ and $n_2 \geq 30$):

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\overline{x}_1 - \overline{x}_2}},$$

where $\sigma_{\overline{x}_1 - \overline{x}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Two-Sample t-Test for the Difference Between Means (σ_1 and σ_2 are unknown, the samples are random and independent, and either the populations are normally distributed or both $n_1 \geq 30$ and $n_2 \geq 30$):

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{s_{\overline{x}_1 - \overline{x}_2}}$$

If population variances are equal, d.f. $= n_1 + n_2 - 2$ and

$$s_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

If population variances are not equal, d.f. is the smaller of $n_1 - 1$ or $n_2 - 1$ and $s_{\overline{x}_1 - \overline{x}_2} = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$.

t-Test for the Difference Between Means (the samples are random and dependent, and either the populations are normally distributed or $n \geq 30$):

$$t = \frac{\overline{d} - \mu_d}{s_d/\sqrt{n}}, \text{ where } \overline{d} = \frac{\Sigma d}{n},\ s_d = \sqrt{\frac{\Sigma (d - \overline{d})^2}{n - 1}},$$

and d.f. $= n - 1$.

Two-Sample z-Test for the Difference Between Proportions (the samples are random and independent, and $n_1\overline{p}$, $n_1\overline{q}$, $n_2\overline{p}$, and $n_2\overline{q}$ are at least 5):

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ where } \overline{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

and $\overline{q} = 1 - \overline{p}$.

Key Formulas

From Larson/Farber *Elementary Statistics: Picturing the World,* Sixth Edition

Chapter 9

Correlation Coefficient:

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

t-Test for the Correlation Coefficient:

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}} \quad (\text{d.f.} = n - 2)$$

Equation of a Regression Line: $\hat{y} = mx + b$,

where $m = \dfrac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$ and

$$b = \overline{y} - m\overline{x} = \frac{\Sigma y}{n} - m\frac{\Sigma x}{n}.$$

Coefficient of Determination:

$$r^2 = \frac{\text{Explained variation}}{\text{Total variation}} = \frac{\Sigma(\hat{y}_i - \overline{y})^2}{\Sigma(y_i - \overline{y})^2}$$

Standard Error of Estimate: $s_e = \sqrt{\dfrac{\Sigma(y_i - \hat{y}_i)^2}{n - 2}}$

c-Prediction Interval for y: $\hat{y} - E < y < \hat{y} + E$, where

$$E = t_c s_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \overline{x})^2}{n\Sigma x^2 - (\Sigma x)^2}} \quad (\text{d.f.} = n - 2)$$

Chapter 10

Chi-Square: $\chi^2 = \Sigma \dfrac{(O - E)^2}{E}$

Goodness-of-Fit Test: d.f. $= k - 1$

Independence Test:

d.f. $=$ (no. of rows $- 1$)(no. of columns $- 1$)

Two-Sample F-Test for Variances: $F = \dfrac{s_1^2}{s_2^2}$, where

$s_1^2 \geq s_2^2$, $\text{d.f.}_N = n_1 - 1$, and $\text{d.f.}_D = n_2 - 1$.

One-Way Analysis of Variance Test:

$$F = \frac{MS_B}{MS_W}, \text{ where } MS_B = \frac{SS_B}{\text{d.f.}_N} = \frac{\Sigma n_i(\overline{x}_i - \overline{\overline{x}})^2}{k - 1}$$

$$\text{and } MS_W = \frac{SS_W}{\text{d.f.}_D} = \frac{\Sigma(n_i - 1)s_i^2}{N - k}.$$

$(\text{d.f.}_N = k - 1, \text{d.f.}_D = N - k)$