Math 115 - Elementary Statistics Summary*

CHAPTER 1

Section 1.1

Data Sets

Population (parameter is numerical characteristic)
Sample (statistic is numerical characteristic)

Branches of Statistics

Descriptive and Inferential

Section 1.2

Types of Data

Qualitative and Quantitative

Levels of Measure

nominal, ordinal, interval, and ratio

Section 1.3

Data Collection Methods

1. Observational Study
2. Experiment
3. Simulation
4. Survey

Types of Sampling Techniques

1. Random sample
2. Stratified sample
3. Cluster sample
4. Systematic sample
5. Convenience sample

CHAPTER 2

Section 2.1

Frequency Distribution Columns

Class, Class Boundaries, Frequency, Midpoint, Relative Frequency, Cumulative Frequency.

Class Width = $\frac{\text{range}}{\text{\# of classes}}$

Midpoint = $\frac{\text{lower limit+upper limit}}{2}$

Relative Frequency = $\frac{\text{class frequency}}{\text{sample size(n)}}$

Frequency Histogram (horizontal = midpoints, vertical = frequencies)

Section 2.3

Pop. Mean: $\mu = \frac{\sum X}{N}$ **Sample Mean**: $\overline{X} = \frac{\sum X}{n}$ **Weighted Mean:** $\overline{x} = \frac{\sum(x \bullet w)}{\sum w}$

*Reproduced with permission from Stacey Buck

Math 115 - Elementary Statistics Summary*

Mean of Grouped Data (mean of a frequency distribution)

$$\bar{x} = \frac{\sum(x \cdot f)}{n}$$ x = midpoints, f = frequencies, $n = \sum f$

Section 2.4

Population deviation of x = $x - \mu$ Sample deviation of x = $x - \bar{X}$

Sum of Squares: $\sum(x - \mu)^2$

Population Standard Deviation:

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

Sample Standard Deviation:

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

Calculator: Computing Standard Deviation

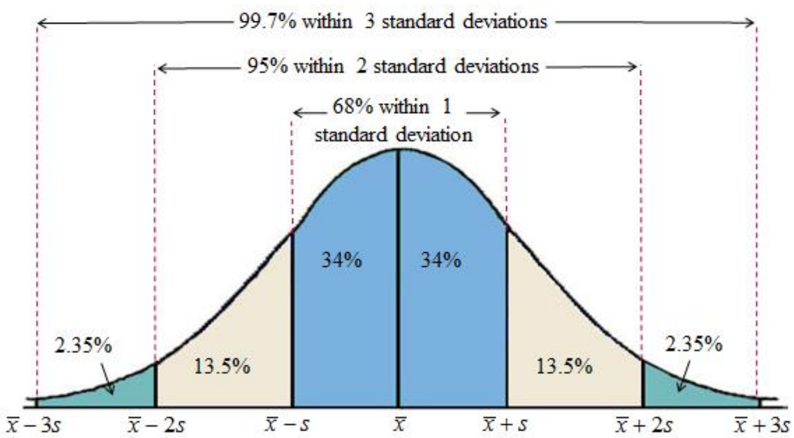
To enter the data list into the calculator:

STAT → EDIT Menu → enter data into L_1

To compute mean and standard deviation

STAT → CALC Menu → 1:1 Var Stats

Empirical Rule



Chebychev's Theorem

The portion of any data set lying within K (K > 1) standard deviations from the mean is **at least** $1 - \frac{1}{k^2}$

If K = 2 then at least 75% of data lies within 2 standard deviation of the mean.

If K = 3 then at least 88.9% of data lies within 3 standard deviations of the mean.

*Reproduced with permission from Stacey Buck

Math 115 - Elementary Statistics Summary*

Standard Deviation of Grouped Data (s.d. of a frequency distribution)

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

Calculator: L_1 = midpoints (x-values), L_2= frequencies; then use 1-var stats then L_1, L_2

<u>Section 2.5</u>

IQR = $Q_3 - Q_1$

Outlier: any entry beyond: $Q_1 - 1.5(IQR)$ or $Q_3 + 1.5(IQR)$

Percentile of x = $\frac{\text{\# of data values less than x}}{\text{total number of data values}} \cdot 100$

z-score = $\frac{x-\mu}{\sigma}$ (A z-score is considered unusual if it is outside of the -2 to 2 range)

<u>CHAPTER 3</u>

<u>Section 3.1</u>

Fundamental Counting Principle: multiple events occurring in sequence m•n ways

Classical (Theoretical) Probability

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

Empirical Probability

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

Compliment: P(E)' = 1 – P(E)

<u>Section 3.2</u>

Independent Events: P(B/A) = P(B) and P(A/B) = P(A)

Multiplication Rule (probability that two events will occur in sequence)

P(A and B) = P(A) • P(B/A)

independent events: P(A andB) = P(A) • P(B)

<u>Section 3.3</u>

Addition Rule

P(A or B) = P(A) + P(B) – P(A and B)

mutually exclusive: P(A or B) = P(A) + P(B)

<u>CHAPTER 4</u>

<u>Section 4.1</u>

Mean of a Discrete Probability Distribution: $\mu = \sum x \cdot p(x)$

Standard Deviation of a Discrete Probability Distribution (Discr. Random Variable)

σ = $\sqrt{\sum(x-\mu)^2 \cdot p(x)}$

Calculator for Standard Deviation of Discrete Probability Distribution:

L_1 – discrete random variables (x); L_2 – probabilities p(x); then 1-Var stats then L_1, L_2

*Reproduced with permission from Stacey Buck

Math 115 - Elementary Statistics Summary*

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

Section 4.2
Binomial Experiments
n = number of trials; p = p(success); q = p(failure); x = # of successes in n trials

Binomial Probability Formula: $\frac{n!}{(n-x)! \cdot x!} \cdot p^x \cdot q^{n-x}$ p(exactly x successes in n trials)

Calculator for Binomial Probabilities:
Probability of exactly x success: binompdf(n, p, x)
Probability of "at most x successes" binomcdf(n, p, x)

Unusual Probabilities: $p \le .05$

Population Parameters of a Binomial Distribution
Mean: $\mu = n \cdot p$
Variance: $\sigma^2 = n \cdot p \cdot q$
Standard Deviation: $\sigma = \sqrt{n \cdot p \cdot q}$

CHAPTER 5

Section 5.1
To transform any x-value to a z-score use:

z-score $= \frac{x - \mu}{\sigma} = \frac{\text{value} - \text{mean}}{\text{standard deviation}}$

Calculator to find an area that corresponds to a given z-score:
normalcdf(-10,000,z) = area to the left of z
normalcdf(z, 10,000) = area to the right of z
normalcdf(z_1, z_2) = area between two z's

Section 5.2
Finding Normal Distribution Probabilities
Finding the probability that x will fall in a given interval by finding the area under the normal curve for that interval
Calculator: normalcdf(x_1, x_2, μ, σ) (Probability from raw data (x's))

Section 5.3
Calculator to find the z-score for a given area or a percentile:
invNorm(area)

Finding an x-value for a corresponding z-score
$x = \mu + z\sigma$

Calculator to find an x-value for a given probability:
Calculator: invNorm(area, μ, σ)

*Reproduced with permission from Stacey Buck

Math 115 - Elementary Statistics Summary*

<u>Section 5.4</u>
Central Limit Theorem
If n ≥ 30 or population is normally distributed, then:

$$\mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

To transform $\bar{x}$ to a *z*-score:

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

Calculator: normalcdf (x_1, x_2, $\mu_{\bar{x}}$, $\frac{\sigma}{\sqrt{n}}$)

<u>Section 5.5</u>
You can use a normal distribution to approximate a binomial distribution if np ≥ 5 and nq ≥ 5. If this is true, then do the following:

1. Find μ = np and σ = $\sqrt{npq}$
2. Apply the continuity correction (Add or subtract 0.5 from the endpoints).
3. Use the **calculator** to find the binomial probability:
   normalcdf: (x_1, x_2, μ, σ)

<u>CHAPTER 6</u>

<u>Section 6.1</u> (Confidence interval for the mean - large samples)
Margin of Error (E): The greatest possible distance between $\bar{x}$ and μ

$$E = z_c \frac{\sigma}{\sqrt{n}}$$

Confidence Interval: where "c" is the probability that the confidence interval contains μ

$$\bar{x} - E < \mu < \bar{x} + E$$

Calculator: STAT → TESTS Menu → 7:Zinterval

Minimum Sample Size:

$$n = \left(\frac{z_c \sigma}{E}\right)^2$$

<u>Section 6.2</u> (Confidence interval for the mean - small samples)
Use when: σ is unknown, n < 30 and population is (approx.) normally distributed
Degrees of Freedom:
d.f. = n – 1

Critical Value = t_c is found in Table 5 using d.f. and the confidence interval wanted.

Margin of Error (E):

$$E = t_c \frac{s}{\sqrt{n}}$$

*Reproduced with permission from Stacey Buck

Math 115 - Elementary Statistics Summary*

Confidence Interval: $\bar{x} - E < \mu < \bar{x} + E$

Calculator: STAT → TESTS Menu → 8:Tinterval

<u>Section 6.3</u> (Confidence intervals for population proportions)
Population Proportion (p):
- probability of success in a single trial of a binomial experiment
- proportion of the population included in a "success" outcome (we are estimating this)

$$\hat{p} = \frac{x}{n} = \frac{\text{\# of successes in the sample}}{\text{sample size}}$$

$$\hat{q} = 1 - \hat{p}$$

Confidence Interval for p: $\hat{p} - E < p < \hat{p} + E$

Margin of Error (E):

$E = z_c\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$ ($n\hat{p} \geq 5$ and $n\hat{q} \geq 5$ for a normal approximation)

Calculator: STAT → TESTS Menu → A:1-PropZint

Minimum Sample Size:

$$n = \hat{p}\,\hat{q}\left(\frac{z_c}{E}\right)^2$$

<u>CHAPTER 7</u>

<u>Section 7.1</u>

Hypothesis Testing: Uses sample statistics to test a claim about the value of a population parameter.

H_0: $\mu \geq k$	H_0: $\mu \leq k$	H_0: $\mu = k$
H_a: $\mu < k$	H_a: $\mu > k$	H_a: $\mu \neq k$
left-tailed	right-tailed	two-tailed

Level of Significance = α
The maximum allowable probability of making a Type I error.
P-Value (probability value)
-The estimated probability of rejecting Ho when it is true (Type I error)
-The smaller the P-value the more evidence to reject H_o.

*Reproduced with permission from Stacey Buck

Math 115 - Elementary Statistics Summary*

Section 7.2 (Hypothesis testing for mean - large sample)

z-Test

$z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ (if n ≥ 30, the $\sigma \approx s$)

Guidelines for Using P-Values

1. Find the z-score and then area of your data and compare it to α.
   can use normalcdf($\infty, \bar{x}, \mu_{\bar{x}}, \sigma_{\bar{x}}$) = area of data
2. If P ≤ α then reject H_o.
   If P > α then fail to reject H_o.

Calculator: STAT → TESTS Menu → 1:Z-Test

Rejection Regions

-Range of values for which H_o is not probable; If z-score for data is in this region reject H_o.

Guidelines for Using Rejection Regions

1. Find the z-score that goes with α and sketch. (This delineates rejection region)
2. Find z-score for given data and add to sketch
3. Reject H_o if data z-score is in rejection region.

Section 7.3 Hypothesis Testing for the mean - small samples using t-Distribution)

Using t-Test Guidelines

1. Find critical values (t-scores) for α using d.f. = n – 1, and table 5 then sketch
2. Compute ***t*** for data and add to sketch
3. Reject H_o if ***t*** for data is in rejection region delineated by critical values.

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

Using P-Values with t-Test

This can be done only with a graphing calculator
Calculator: STAT → TESTS Menu → 2:T-Test

*Reproduced with permission from Stacey Buck

Math 115 - Elementary Statistics Summary*

Section 7.4 (Hypothesis testing for a population proportion (p))

Test statistic = $\hat{p}$ and standardized test statistic = z
Must have: $np \geq 5$ and $nq \geq 5$ then use z-Test:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Guidelines for Hypothesis Testing For a Population Proportion

1. check np and nq then find rejection regions for α and sketch
2. Find z-scores for data and add to sketch
3. Reject H_o if data z-score is in rejection region.

Calculator: STAT → TESTS Menu → 5:PropZTest

CHAPTER 8

Section 8.1 (Testing the difference between sample means - large sample)
Necessary z-Test Conditions

1. Samples are randomly selected
2. Samples are independent
3. $n \geq 30$ **or** each population is normally distributed and σ is known.

Then $\bar{x}_1$ - $\bar{x}_2$ is normally distributed so you can use a z-Test
(s_1 and s_2 can be used for σ_1 and σ_2)

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

Calculator: STAT → TESTS Menu → 3:2-SampZTest

Section 8.2 (Testing the difference between sample means - small sample)
-n<30 and σ is unknown
-Samples must be independent, randomly selected and normally distributed

If the variances are equal use the following to compute t (pooled estimate):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)s_1{}^2 + (n_2 - 1)s_2{}^2}{n_1 + n_2 - 2}} \bullet \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{and} \quad \text{d.f.} = n_1 + n_2 - 2$$

*Reproduced with permission from Stacey Buck

If the variances are not equal use the following to compute t :

$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ and d.f. = smaller of $(n_1 - 1)$ and $(n_2 - 1)$

Calculator: STAT → TESTS Menu → 4:2-SampTTest Pooled: Yes or no

Section 8.4 (Testing the difference between population proportions)
To use a z-Test
1. The samples are independent and randomly selected.
2. n_1p_1, n_1q_1, n_2p_2, n_2q_2 all ≥ 5 (large enough to use a normal sampling distribution)

Weighted Estimate of p_1 and p_2

$\bar{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$ $x_1 = n_1\widehat{p_1}$ and $x_2 = n_2\widehat{p_2}$ (assume that $p_2 - p_1 = 0$)

$\bar{q} = 1 - \bar{p}$ (Condition needed: $n_1\overline{p_1}$, $n_1\overline{q_1}$, $n_2\overline{p_2}$, $n_2\overline{q_2}$ all ≥ 5)

$Z = \dfrac{(\widehat{p_1} - \widehat{p_2}) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

Calculator: STAT → TESTS Menu → 6:2-PropZTest

CHAPTER 9

Section 9.1
Correlation Coefficient (r)
-measures the direction and strength of a linear correlation between two variables
-range: $-1 \le r \le 1$

Correlation Coefficient Formula

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

Calculator:
STAT → Edit → L_1 (enter x-values) and L_2 (enter y-values), then
STAT → CALC Menu → 4: LinReg (ax + b) → enter

Testing a Population Correlation Coefficient With Table 11
1. Determine n = # of pairs.
2. Find the critical values for α using Table 11.
3. If |r| > c.v. the correlation coefficient of the population can be determined to be significant.

*Reproduced with permission from Stacey Buck

Hypothesis Testing for a Population Correlation Coefficient ρ
H_o: $\rho = 0$ (no significant correlation)
H_a: $\rho \neq 0$ (significant correlation)

$t = \dfrac{r}{\sqrt{\frac{1-r^2}{n-2}}}$ d.f = n – 2

Section 9.2
Equation of a Regression Line:

$\hat{y} = mx + b$

CHAPTER 10
Section 10.1
Chi-Square Goodness-of-fit Test: Used to test whether a frequency distribution fits an expected distribution.
H_o: The frequency distribution fits the specified distribution
H_a: The frequency distribution does not fit the specified distribution.
$E_i = np_i$

n = the number of trials (sample size)

p_i = the assumed probability of the specific category.

Conditions Needed:
1. The observed frequencies must be obtained using a random sample
2. Each $E \geq 5$

$x^2 = \sum \dfrac{(O-E)^2}{E}$ d.f. = k – 1 (k = # of categories in the distribution)

Guidelines For Performing a Chi-Square Goodness-o-Fit Test
1. Use d.f. and Table 6 to find the critical values and sketch the rejection region
2. Compute x^2 and add to sketch.
3. If x^2 is in rejection region reject Ho.

Section 10.2
Chi-Square Independence Test: Used to determine whether the occurrence of one variable affects the probability of the occurrences of the other variable.

$x^2 = \sum \dfrac{(O-E)^2}{E}$ d.f. = (r - 1)(c - 1) (r = # of rows and c = # of columns)

Guidelines For Performing a Chi-Square Independence Test
1. Use d.f. and Table 6 to find the critical values and sketch the rejection region
2. Compute x^2 and add to sketch.
3. If x^2 is in rejection region reject Ho.

*Reproduced with permission from Stacey Buck