**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** = $\frac{range}{\# of classes}$

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

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

**Frequency Histogram** (horizontal = midpoints, vertical = frequencies)

**Section 2.3**

**Pop. Mean**: µ = $\frac{∑X}{N}$ **Sample Mean**: $\overbar{X}= \frac{∑X}{n}$ **Weighted Mean:** $\overbar{x }=\frac{∑(x •w)}{∑w}$

**Mean of Grouped Data** (mean of a frequency distribution)

 $\overbar{x }= \frac{∑(x •f)}{n}$ x = midpoints, f = frequencies, n = ∑f

**Section 2.4**

Population deviation of x = x - µ Sample deviation of x = x - $\overbar{X}$

**Sum of Squares**: ∑( x - µ)2

**Population Standard Deviation:** **Sample Standard Deviation:**

σ = $\sqrt{\frac{∑(x- µ)^{2}}{N}}$ s = $\sqrt{\frac{∑(x- \overbar{x})^{2}}{n-1}}$

**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.

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

S = $\sqrt{\frac{∑(x- \overbar{x})^{2}•f}{n-1}}$

 **Calculator:** L1 = midpoints (x-values), L2= frequencies; then use 1-var stats then L1, L2

**Section 2.5**

**IQR** = Q3 – Q1

**Outlier:** any entry beyond: Q1 – 1.5(IQR) or Q3 + 1.5(IQR)

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

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

**CHAPTER 3**

**Section 3.1**

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

**Classical (Theoretical) Probability Empirical Probability**

P(E) = $\frac{\# of outcomes in event E}{\# of outcomes in sample space}$ P(E) = $\frac{frequency of event}{total frequency}= \frac{f}{n}$

**Compliment:** P(E)’ = 1 – P(E)

**Section 3.2**

**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)

**Section 3.3**

**Addition Rule**

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

**CHAPTER 4**

**Section 4.1**

**Mean of a Discrete Probability Distribution:**  µ = ∑ x • p(x)

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

σ = $\sqrt{∑\left(x- µ\right)^{2 }• p(x) }$

**Calculator for Standard Deviation of Discrete Probability Distribution:**

 L1 – discrete random variables (x); L2 – probabilities p(x); then 1-Var stats then L1, L2

**Expected Value:** E(x) = µ = ∑ x • 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:** $\frac{n!}{\left(n-x\right)! •x!}$• px • qn-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 ≤ .05

**Population Parameters of a Binomial Distribution**

Mean: µ = n•p

Variance: σ2 = n•p•q

Standard Deviation: σ = $\sqrt{n•p•q}$

**CHAPTER 5**

**Section 5.1**

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

**z-score** = $\frac{x- μ}{σ}$ = $\frac{value-mean}{standard deviation}$

**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(z1, z2) = 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(x1, x2, 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 = µ + zσ

**Calculator to find an x-value for a given probability:**

**Calculator:** invNorm(area, µ, σ)

**Section 5.4**

**Central Limit Theorem**

If n ≥ 30 or population is normally distributed, then:

 $μ\_{\overbar{x}}= μ$ and $σ\_{\overbar{x}}^{2}$ = $\frac{σ^{2}}{n}$ and $σ\_{\overbar{x}}$ = $\frac{σ}{\sqrt{n}}$

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

 Z = $\frac{\overbar{x}- μ\_{\overbar{x}}}{\frac{σ}{\sqrt{n}}}$

**Calculator**: normalcdf (x1, x2, $μ\_{\overbar{x}}$, $\frac{σ}{\sqrt{n}}$)

**Section 5.5**

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: (x1, x2, µ, σ)

**CHAPTER 6**

**Section 6.1** (Confidence interval for the mean - large samples)

**Margin of Error (E):** The greatest possible distance between $\overbar{x}$ and µ

E = zc $\frac{σ}{\sqrt{n}}$

**Confidence Interval:** where “c” is the probability that the confidence interval contains µ

 $\overbar{x}$ – E < µ < $\overbar{x}$ + E

**Calculator:** STAT → TESTS Menu → 7:Zinterval

**Minimum Sample Size:**

n = $\left(\frac{z\_{c}σ}{E}\right)^{2}$

**Section 6.2** (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 = tc** is found in Table 5 using d.f. and the confidence interval wanted.

**Margin of Error (E):**

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

**Confidence Interval:** $\overbar{x}$ – E < µ < $\overbar{x}$ + E

**Calculator:** STAT → TESTS Menu → 8:Tinterval

**Section 6.3**  (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)

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

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

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

**Margin of Error ( E):**

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

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

**Minimum Sample Size:**

n = $\hat{p}$ $\hat{q}$ $\left(\frac{z\_{c}}{E}\right)^{2}$

**CHAPTER 7**

**Section 7.1**

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

H0: μ ≥ k H0: μ ≤ k H0: μ = k

Ha: μ < k Ha: μ > k Ha: μ ≠ 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 Ho.

**Section 7.2** (Hypothesis testing for mean - large sample)

**z-Test**

z = $\frac{\overbar{x}- μ}{\frac{σ}{\sqrt{n}}} $ = $\frac{\overbar{x}- μ}{\frac{s}{\sqrt{n}}}$ (if n ≥ 30, the σ $≈$ 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( $\infty , \overbar{x, } µ\_{\overbar{x}, } σ\_{\overbar{x}}$ ) = area of data

2. If P ≤ $α$ then reject Ho.

 If P > $α$ then fail to reject Ho.

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

**Rejection Regions**

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

 reject Ho.

**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 Ho 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 Ho if ***t*** for data is in rejection region delineated by critical values.

t = $\frac{\overbar{x}- μ}{\frac{s}{\sqrt{n}}}$

**Using P-Values with t-Test**

This can be done only with a graphing calculator

**Calculator:** STAT → TESTS Menu → 2:T-Test

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

Test statistic = $\hat{p}$ and standardized test statistic = z

Must have: np≥5 and nq≥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 Ho 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≥30 **or** each population is normally distributed and σ is known.

Then $\overbar{x}\_{1}$ - $\overbar{x}\_{2}$ is normally distributed so you can use a z-Test

 (s1 and s2 can be used for σ1 and σ2)

z = $\frac{\left(\overbar{x}\_{1} - \overbar{x}\_{2}\right)- \left(µ\_{1}- µ\_{2}\right)}{\sqrt{\frac{σ\_{1}^{2}}{n\_{1} }+\frac{σ\_{2}^{2}}{n\_{2} } }}$

**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 = $\frac{\left(\overbar{x}\_{1} - \overbar{x}\_{2}\right)- \left(µ\_{1}- µ\_{2}\right)}{\sqrt{\frac{\left(n\_{1}-1 \right)s\_{1}^{2}+ \left(n\_{2}-1\right)s\_{2}^{2}}{n\_{1}+ n\_{2}-2}}•\sqrt{\frac{1}{n\_{1} }+\frac{1}{n\_{2} } }}$ and d.f. = $n\_{1}+ n\_{2}-2$

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

t = $\frac{\left(\overbar{x}\_{1} - \overbar{x}\_{2}\right)- \left(µ\_{1}- µ\_{2}\right)}{\sqrt{\frac{s\_{1}^{2}}{n\_{1} }+\frac{s\_{2}^{2}}{n\_{2} } }}$ and d.f. = smaller of (n1 – 1) and (n2 – 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. n1p1, n1q1, n2p2, n2q2 all ≥ 5 (large enough to use a normal sampling distribution)

**Weighted Estimate of p1 and p2**

$\overbar{p}$ = $\frac{x\_{1}+ x\_{2}}{n\_{1}+ n\_{2}}$ x1 = n1$\hat{p\_{1}}$ and x2 = n2$\hat{p\_{2}}$ (assume that p2 – p1 = 0)

$\overbar{q}$ = 1 - $\overbar{p}$ (Condition needed: n1$\overbar{p\_{1}}$, n1$\overbar{q\_{1}}$, n2$\overbar{p\_{2}}$, n2$\overbar{q\_{2}}$ all ≥ 5)

Z = $\frac{(\hat{p\_{1}} - \hat{p\_{2}}) – (p\_{1} - p\_{2})}{\sqrt{\overbar{p}\overbar{q}\left(\frac{1}{n\_{1} }+\frac{1}{n\_{2} }\right)}}$

**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 ≤ r ≤ 1

**Correlation Coefficient Formula**

r = $\frac{n∑xy – (∑x)(∑y)}{\sqrt{n∑x^{2}- \left(∑x\right)^{2} } \sqrt{n∑y^{2}- (∑y)^{2}}} $

**Calculator:**

STAT → Edit → L1 (enter x-values) and L2 (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.

**Hypothesis Testing for a Population Correlation Coefficient** $ρ$

Ho: $ρ=0$ (no significant correlation)

Ha: $ρ \ne 0$ (significant correlation)

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

**Ho:** The frequency distribution fits the specified distribution

**Ha:** The frequency distribution does not fit the specified distribution.

**Ei = npi**

 n = the number of trials (sample size)

 pi = the assumed probability of the specific category.

**Conditions Needed:**

1. The observed frequencies must be obtained using a random sample

2. Each E ≥ 5

x2 = ∑ $\frac{\left(O-E\right)^{2}}{E}$ 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 x2 and add to sketch.

3. If x2 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.

x2 = ∑ $\frac{\left(O-E\right)^{2}}{E}$ 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 x2 and add to sketch.

3. If x2 is in rejection region reject Ho.