Chapter 1: Stats Starts Here
Chapter 2: Data

Key Vocabulary:

- Statistics
- data, datum
- variation
- individual
- respondent
- subject
- participant
- experimental unit
- observation
- variable
- categorical
- quantitative

Calculator Skills:

- enter data in a list
- change a datum
- delete a datum
- name a new list
- clear a list
- delete a list
- recreate a list
- copy a list

1. Name three things you learned about *Statistics* in Chapter 1.

   •

   •

   •

2. The authors claim that this book is very different from a typical mathematics textbook. Would you agree or disagree, based on what you read in Chapter 1? Explain.

3. According to the authors, what are the "three simple steps to doing *Statistics* right?"

4. What do the authors refer to as the "W's of data?"

5. Why must data be in context (the W's)?

6. Explain the difference between a *categorical variable* and a *quantitative variable*. Give an example of each.

Chapter 3: Displaying and Describing Categorical Data

Key Vocabulary:

- frequency table
- relative frequency table
- distribution
- bar chart
- pie chart
- contingency table
- marginal distribution
- conditional distribution
- independent
- segmented bar chart
- Simpson's Paradox

1. According to the authors, what are the three rules of data analysis?

2. Explain the difference between a frequency table and a relative frequency table.

3. When is it appropriate to use a bar chart?

4. When is it appropriate to use a pie chart?

5. When is it appropriate to use a contingency table?

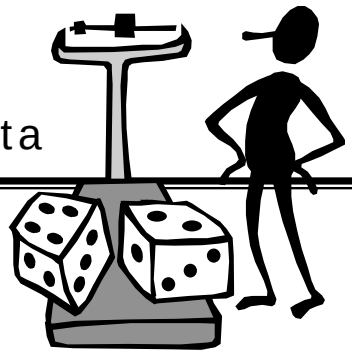
6. What does a marginal distribution show?

7. When is it appropriate to look at a conditional distribution?

8. What does it mean for two variables to be independent?

9. How does a segmented bar chart compare to a pie chart?

10. Explain what is meant by Simpson's Paradox.



Chapter 4: Displaying Quantitative Data

Key Vocabulary:

- distribution
- histogram
- relative frequency histogram
- stem-and-leaf display
- dotplot
- shape
- center
- spread
- mode
- unimodal
- bimodal
- multimodal
- uniform
- symmetric
- tail
- skewed
- outliers
- gaps
- time plot
- re-expressing data

Calculator Skills:

- display a histogram
- SortA (

1. What is meant by a *distribution*?

2. Explain the difference between a *histogram* and a *relative frequency histogram*.

3. In what ways are *histograms* similar to *stem-and-leaf displays*?

4. Name some advantages and disadvantages of *stem-and-leaf displays.*

5. When is it more appropriate to use a *histogram* rather than a *stem-and-leaf display*?

6. Name some advantages and disadvantages of *dotplots.*

7. When describing a *distribution*, what three things should you always mention?

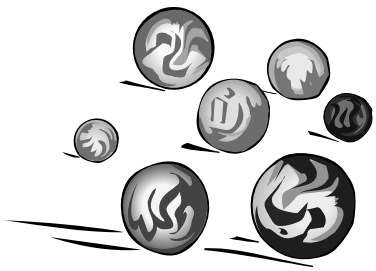
8. What should you look for when describing the *shape* of a *distribution*?

9. In general, what is meant by the *center* of a *distribution*?

10. In general, what is meant by the *spread* of a *distribution*?

11. When is it appropriate to use a *time plot* to display quantitative data?

12. What is meant by *re-expressing* or *transforming* data? What is the purpose of *re-expressing* or *transforming* data?



Chapter 5: Describing Distributions Numerically

Key Vocabulary:

- center
- spread
- midrange
- median
- range
- quartile
- interquartile range
- percentile
- five-number summary
- boxplot
- mean
- standard deviation
- variance

Calculator Skills:

- boxplot
- modified boxplot
- 1-Var Stats

1. Explain the difference between *range* and *interquartile range*. Why is the *interquartile range* often a better measure of the spread of a distribution?

2. What are some advantages of *boxplots*?

3. What are some disadvantages of *boxplots*?

4. When is it more appropriate to use the *mean* as a measure of center rather than the *median*? Why?

5. When is it more appropriate to use the *median* as a measure of center rather than the *mean*? Why?

6. When do the *mean* and *median* have the same value?

7. Describe the relationship between *variance* and *standard deviation*.

Chapter 6: The Standard Deviation as a Ruler and the Normal Model

Key Vocabulary:

- standard deviation
- standardized value
- rescaling
- *z*-score
- normal model
- parameter
- statistic
- standard Normal model
- 68-95-99.7 Rule
- normal probability plot
- $N(\mu, \sigma)$

Calculator Skills:

- normalpdf(
- normalcdf(
- invNorm(
- normal probability plot
- -1E99 and 1E99

1. What unit of measurement is used to describe how far a set of values are from the mean?

2. Explain how to *standardize* a value.

3. Briefly describe why *standardized units* are used to compare values that are measured using different scales, different units, or different populations.

4. How does adding or subtracting a constant amount to each value in a set of data affect the mean? Why does this happen?

5. How does multiplying or dividing a constant amount by each value in a set of data (also called rescaling) affect the mean? Why does this happen?

6. How does adding or subtracting a constant amount to each value in a set of data affect the standard deviation? Why does this happen?

7. How does multiplying or dividing a constant amount by each value in a set of data (also called rescaling) affect the standard deviation? Why does this happen?

8. How does *standardizing* a variable affect the shape, center, and spread of its distribution?

9. In what way does a *z-score* give an indication of how unusual a value is?

10. How would you describe the shape of a *normal curve*? Draw several examples.

11. Where on the *normal curve* are *inflection points* located?

12. When is it appropriate to use a *normal model* to model a set of data?

13. Explain the difference between $\bar{y}$ and μ.

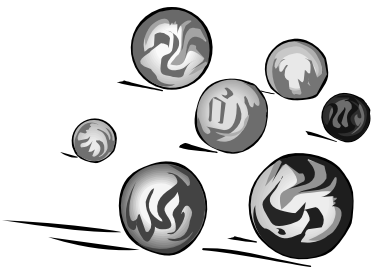
14. Explain the difference between s and σ.

15. Briefly explain the *68-95-99.7 Rule*.

16. What is a *percentile*?

17. Is there a difference between the 80^{th} percentile and the top 80%? Explain.

18. Describe two methods for assessing whether or not a distribution is *approximately normal*.



Chapter 7: Scatterplots, Association, and Correlation

Key Vocabulary:

- scatterplot
- association
- direction
- form
- scatter
- explanatory variable
- response variable
- correlation coefficient

Calculator Skills:

- r-value
- linear regression
- make a scatterplot

1. What type of graph is used to show the relationship between two quantitative variables?

2. When describing a *scatterplot*, what four things should you always mention?

3. What is meant by an *explanatory variable*?

4. What is meant by a *response variable*?

5. What does *correlation* measure?

6. Explain the difference between *association* and *correlation*?

7. What three conditions are necessary in order to use *correlation* as a measure of *association*?

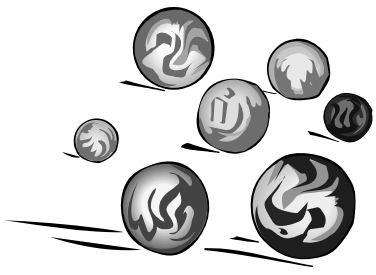
8. What does the sign of the *correlation coefficient* tell you about the association?

9. What does a *correlation* near 1 or – 1 indicate?

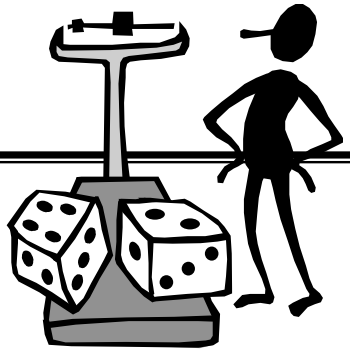
10. What does a *correlation* near 0 indicate?

11. Sketch an example of a scatterplot that shows two variables with a strong *association* but a weak *correlation*.

12. Is correlation *resistant* or *nonresistant* to outliers? Explain.



Chapter 8: Linear Regression



Key Vocabulary:

- parameter
- linear model
- predicted value
- residual
- line of best fit
- slope
- $\hat{y}$
- mean-mean point
- regression line
- R^2
- coefficient of determination

Calculator Skills:

- LinReg (a + bx)
- RESID

1. Explain the quote (by George Box, a famous statistician), "All models are wrong, but some are useful."

2. What are the *parameters* of the Normal model?

3. Describe the difference in notation between y and $\hat{y}$.

4. What is a *residual* and how is it calculated?

5. What does a negative *residual* indicate? A positive *residual*? A *residual* of zero?

6. How many *residuals* does a set of data have?

7. What is meant by a *line of best fit*?

8. The *line of best fit* always passes through which point?

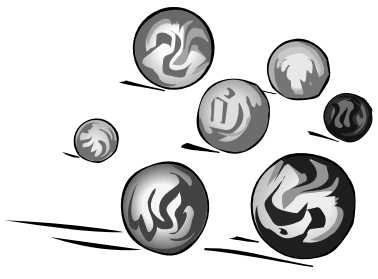
9. The R^2 value shows how much of the *variation* in the response variable can be accounted for by the linear regression model. If $R^2 = 0.95$, what can be concluded about the relationship between x and y?

   _____ % of the variability in _____ is accounted for by the linear relationship with _____.

10. What conditions are necessary before using a *linear model* for a set of data?

11. Explain how to construct a *residual plot*.

12. If a *least-squares regression line* fits the data well, what characteristics should the *residual plot* exhibit? Sketch a well-labeled example.



Chapter 9: Regression Wisdom

Key Vocabulary:

- extrapolation
- interpolation
- leverage
- influential point
- lurking variable
- homogeneous

1. If a histogram of residuals appears to be bimodal or multimodal, what might you conclude (or at least investigate) about the data?

2. Why is it important to examine a residual plot even if a scatterplot appears to be linear?

3. What is meant by extrapolation?

4. Why is extrapolation dangerous?

5. How is interpolation different than extrapolation?

6. How do you identify an outlier in a scatterplot?

7. What is meant by a point that has high leverage?

8. Explain the difference between an outlier and an influential point in a scatterplot.

9. Even if a strong association is observed between two variables, it may be wrong to assume that changing one variable will cause a change in the other variable. Explain why.

10. Why do scatterplots of summary statistics show less scatter than those of individual observations.

Chapter 10: Re-expressing Data

Key Vocabulary:

- Re-expression

Calculator Skills:

- log(
- ln(
- LnReg
- ExpReg
- PwrReg
- QuadReg
- CubicReg

1. What is meant by re-expressing data?

2. One of the goals of re-expressing data is to make the distribution appear more symmetric. Why is this advantageous?

3. Another goal of re-expressing data is to make the spread of several groups more alike. Why is this advantageous?

4. Why is it advantageous to make the form of a scatterplot more nearly linear?

5. What type of data often benefits from re-expression by squaring values?

6. What type of data often benefits from re-expression by taking the square root of values?

7. What type of data often benefits from re-expression by taking the logarithm of values?

8. What type of data often benefits from re-expression by taking the reciprocal of values?

9. If your data contain zeroes, what must you do before re-expressing using logarithms or reciprocals? Explain.

10. If a scatterplot of the *x*-values vs. the logarithm of the *y*-values appears to be linear, what type of relationship is there between the original *x*- and *y*-values?

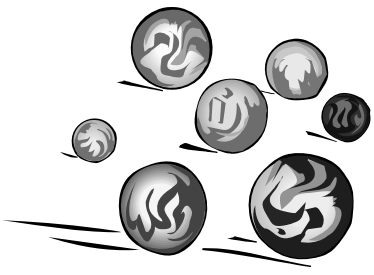
11. Rewrite $\hat{y} = ab^x$ in linear form.

12. If a scatterplot of the logarithm of the *x*-values vs. the logarithm of the *y*-values appears to be linear, what type of relationship is there between the original *x*- and *y*-values?

13. Rewrite $\hat{y} = ax^b$ in linear form.

14. If a scatterplot of the logarithm of the *x*-values vs. the *y*-values appears to be linear, what type of relationship is there between the original *x*- and *y*-values?

15. Rewrite $\hat{y} = a + b \ln x$ in linear form.



Chapter 11: Understanding Randomness

Key Vocabulary:

- random behavior
- random numbers
- pseudorandom numbers
- simulation
- component
- outcome
- trial

Calculator Skills:

- randInt(

1. What is meant by *random* behavior?

2. What is the purpose of a *simulation*?

3. What are the steps for conducting a *simulation*?

4. What is meant by a *trial*?

5. Why is it necessary to conduct a large number of *trials*?

Chapter 12: Sample Surveys

Key Vocabulary:

- population
- sample
- sample survey
- biased
- randomization
- census
- parameter
- statistic
- Simple Random Sample (SRS)
- sampling frame
- sampling variability
- homogeneous groups
- heterogeneous groups
- strata
- stratified random sample
- cluster sampling
- multistage sampling
- systematic sampling
- respondents
- voluntary response sample
- convenience sampling
- undercoverage
- nonresponse bias
- response bias

1. Explain the difference between a *population*, a *sampling frame*, and a *sample*.

2. What does it mean for a sample to be *representative* of a *population*?

3. What is meant by a *biased* sample?

4. What is the role of *randomization* in selecting a sample?

5. What is meant by a *census*? Why is a *census* often impractical?

6. Explain the difference between a *parameter* and a *statistic*.

7. A *Simple Random Sample* (SRS) must satisfy what two conditions?

8. What is meant by *sampling variability*?

9. When is *stratified random sampling* useful?

10. When is *cluster sampling* useful?

11. What is meant by a *multistage sampling*?

12. When is *systematic sampling* appropriate?

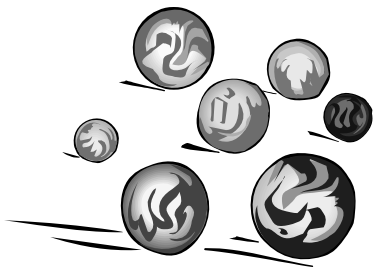
13. In what way are *voluntary response samples* often *biased*?

14. Why is *convenience sampling* unreliable?

15. What is meant by *undercoverage*? Give an example.

16. Explain the difference between *nonresponse bias* and *response bias*.

17. How can the wording of questions cause bias in a survey?



Chapter 13: Experiments

Key Vocabulary:

- observational study
- retrospective study
- prospective study
- experiment
- random assignment
- subjects
- participants
- experimental units
- factors
- levels
- treatment
- block
- completely randomized experiment
- statistically significant
- control
- control group
- single-blind
- double-blind
- placebo
- placebo effect
- matching
- confounding

1. Explain the difference between an *observational study* and an *experiment*.

2. Explain the difference between *experimental units* and *subjects*.

3. Why is it necessary to assign *subjects* to *treatments* at random?

4. Describe the four *Principles of Experimental Design.*
   - *Control*
   - *Randomization*
   - *Replication*
   - *Blocking*

5. Explain what is meant by a *control group*.

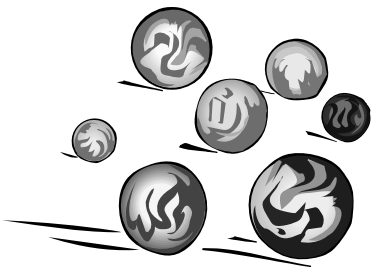
6. Define *statistically significant*.

7. What is the purpose of using a *single-blind* or *double-blind experiment*?

8. What is a *placebo*? What is meant by the *placebo effect*?

9. What is the purpose of using *blocking* in an *experiment*?

10. How might *confounding* affect the results of an *experiment*?



Chapter 14: From Randomness to Probability

Key Vocabulary:

- probability
- trial
- outcome
- event
- independent
- Law of Large Numbers
- complement
- disjoint
- mutually exclusive

1. In statistics, what is meant by the term *random*?

2. In statistics, what is meant by *probability*?

3. What does it mean for trials to be *independent*?

4. How is the *Law of Large Numbers* related to probability?

5. In statistics, what is an *event*?

6. Explain why the probability of any *event* is a number between 0 and 1.

7. What is the sum of the probabilities of all possible *outcomes*?

8. Describe the probability that an *event* does not occur?

9. What is meant by the *complement* of an event?

10. When are two events considered *disjoint*?

11. What is the probability of two *disjoint* events?

12. What is the *Multiplication Rule* for *independent* events?

13. Can *disjoint* events be *independent*?

14. If two events A and B are *independent*, what must be true about A^c and B^c?

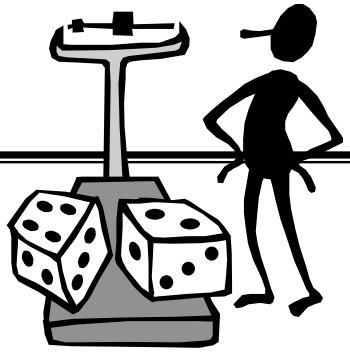
15. What is meant by the *union* of two or more events? Draw a diagram.

16. State the addition rule for *disjoint* events.

17. What is meant by the *intersection* of two or more events? Draw a diagram.

18. Explain the difference between the *union* and the *intersection* of two or more events.

Chapter 15: Probability Rules!



Key Vocabulary:

- trial
- outcome
- event
- sample space
- disjoint
- mutually exclusive
- independent
- conditional probability
- tree diagram

1. If events A and B are *disjoint*, then $P(A \cup B) = P(A) + P(B)$. If events A and B are NOT *disjoint*, explain why this formula does not work.

2. In general (whether events are *disjoint* or not), what is the formula for finding $P(A \cup B)$?

3. Explain the difference between the *Addition Rule* for *disjoint events* and the *General Addition Rule*.

4. What is meant by *joint probability*?

5. What is meant by *conditional probability*?

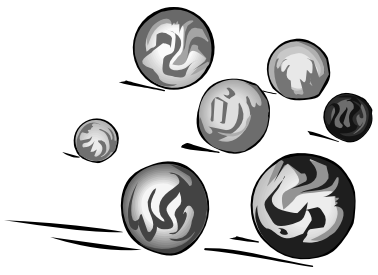
6. State the formula for finding *conditional probability*.

7. Is the probability of "A given B" the same as the probability of "B given A?" Explain.

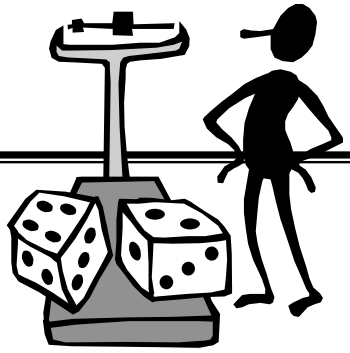
8. In general (whether events are *independent* or not), what is the formula for finding $P(A \cap B)$?

9. Explain the difference between the *Multiplication Rule* for *independent events* and the *General Multiplication Rule*.

10. State the formula used to determine whether or not two events are *independent*.



Chapter 16: Random Variables



Key Vocabulary:

- random variable
- discrete random variable
- continuous random variable
- standard deviation
- expected value
- E(*X*)
- V(*X*)

Calculator Skills:

- 1-VarStats L_1, L_2

1. What is meant by a random variable?

2. Explain the difference between a discrete random variable and a continuous random variable.

3. What information does a probability model give?

4. What is the expected value of a random variable?

5. How do you calculate the expected value of a random variable?

6. Explain the difference between the notations $\overline{x}$ and μ_X.

7. Suppose $\mu_X = 5$ and $\mu_Y = 10$. According to the rules for means, what is μ_{X+Y}?

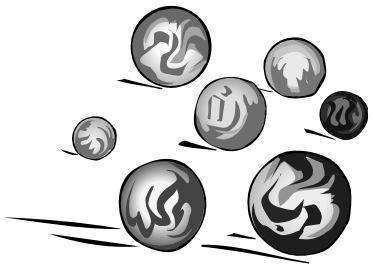
8. Suppose $\mu_X = 2$. According to the rules for means, what is μ_{3+4X}?

9. Explain how to calculate the variance of a discrete random variable X using the formula $\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i$.

10. Given the variance of a random variable, explain how to calculate the standard deviation.

11. Suppose $\sigma_X^2 = 2$ and $\sigma_Y^2 = 3$ and X and Y are independent random variables. According to the rules for variances, what is σ_{X+Y}^2? What is σ_{X+Y}?

12. Suppose $\sigma_X^2 = 4$. According to the rules for variances, what is σ_{3+2X}^2? What is σ_{3+2X}?



Chapter 17: Probability Models

Key Vocabulary:

- Bernoulli trials
- Geometric model
- Binomial model

Calculator Skills:

- geometpdf(
- geometcdf(
- binompdf(
- binomcdf(

1. List three characteristics of Bernoulli trials.

2. What is the variable of interest in a *geometric model*?

3. How do you find the *expected value* and *standard deviation* of a *geometric random variable*?

4. In the *geometic distribution*, what does the parameter p represent?

5. If X has a *geometric distribution*, what does $(1-p)^{n-1}p$ represent?

6. What is the difference between a *probability distribution function* (pdf) and a *cumulative distribution function* (cdf)?

7. What is the variable of interest in a *binomial model*?

8. Explain the difference between the *binomial setting* and the *geometric setting*.

9. How do you find the *expected value* and *standard deviation* of a *binomial random variable*?

10. In the *binomial distribution,* what do parameters *n* and *p* represent?

11. What is meant by *B(n, p)* ?

12. In the formula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, what does *n* represent? What does *k* represent?

What does the value of $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ represent?

13. Complete the following table of values:

1!	1	1
2!	2 x 1	2
3!	3 x 2 x 1	6
4!	4 x 3 x 2 x 1	24

5!	5 x 4 x 3 x 2 x 1	
6!		
7!		
n!		

14. What is the value of $\frac{n!}{(n-1)!}$?

Chapter 18: Sampling Distribution Models

Key Vocabulary:

- parameter
- statistic
- proportion
- sampling distribution model
- Central Limit Theorem
- Standard error

1. Explain the difference between a *parameter* and a *statistic.*

2. Explain the difference between p and $\hat{p}$?

3. What is meant by *sampling variability*?

4. What is meant by the *sampling distribution model* of a statistic?

5. How is the size of a sample related to the *spread* of the sampling distribution?

6. In an SRS of size n, what is true about the sampling distribution of $\hat{p}$ when the sample size n increases?

7. In an SRS of size n, what is the mean of the sampling distribution of $\hat{p}$?

8. In an SRS of size n, what is the standard deviation of the sampling distribution of $\hat{p}$?

9. What happens to the standard deviation of $\hat{p}$ as the sample size *n* increases?

10. When does the formula $\sqrt{\frac{pq}{n}}$ apply to the standard deviation of $\hat{p}$?

11. When the sample size *n* is large, the sampling distribution of $\hat{p}$ is approximately normal. What test can you use to determine if the sample is large enough to assume that the sampling distribution is approximately normal?

12. The mean and standard deviation of a population are *parameters*.

    What symbols are used to represent these *parameters*?

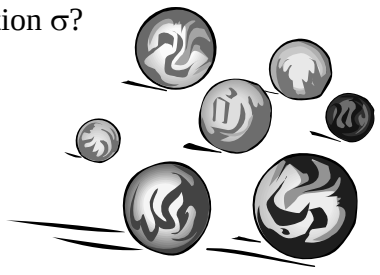
13. The mean and standard deviation of a sample are *statistics*.

    What symbols are used to represent these *statistics*?

14. Because averages are less variable than individual outcomes, what is true about the standard deviation of the sampling distribution of $\bar{x}$?

15. What is the mean of the sampling distribution of $\bar{x}$, if $\bar{x}$ is the mean of an SRS of size *n* drawn from a large population with mean μ and standard deviation σ?

16. What is the standard deviation of the sampling distribution of $\bar{x}$, if $\bar{x}$ is the mean of an SRS of size *n* drawn from a large population with mean μ and standard deviation σ?



17. To cut the standard deviation of $\bar{x}$ in half, you must take a sample _____ times as large.

18. When should you use $\frac{\sigma}{\sqrt{n}}$ to calculate the standard deviation of $\bar{x}$?

19. If σ is not known, what can you use to estimate the standard deviation of $\bar{x}$? What is this called?

20. What does the central limit theorem say about the shape of the sampling distribution of $\bar{x}$?

21. What does the law of large numbers state?

Chapter 19: Confidence Intervals for Proportions

Key Vocabulary:

- standard error
- confidence level
- confidence interval
- margin of error
- critical value
- one-proportion z-interval

Calculator Skills:

- 1-PropZInt

1. Describe the *sampling distribution model* of $\hat{p}$. What *assumptions* must you make for this description to be reasonable?

2. What is the *standard error* of $\hat{p}$? When would you use *standard error* in place of *standard deviation*?

3. Explain the meaning of the following statement: "We are 95% *confident* that between 42.1% and 61.7% of sea fans are infected."

4. What is meant by a *confidence interval*?

5. What is the general form of a *confidence interval* for a one-proportion z-interval?

6. Explain how to calculate *margin of error*.

7. As the *confidence level* increases, what happens to the *margin of error*? What happens to the *confidence interval*?

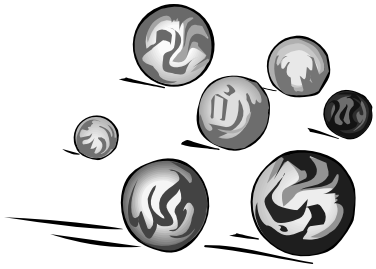
8. By how many times must the sample size *n* increase in order to cut the *margin of error* in half?

9. Why is it best to have high *confidence* and a small *margin of error*?

10. What is the *critical value* z^* for a 90% *confidence interval*? Draw a sketch.

11. What is the *critical value* z^* for a 95% *confidence interval*? Draw a sketch.

12. What is the *critical value* z^* for a 99% *confidence interval*? Draw a sketch.



13. What assumptions and conditions must you consider before creating a *confidence interval* for a proportion?

14. What effect does increasing your sample size have on the *margin of error*? What effect does it have on the *confidence level*? What effect does it have on the *confidence interval*?

15. The formula used to determine the sample size *n* that will yield a confidence interval for a population proportion with a specified margin of error *m* is $m = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$. Solve for *n*.

Chapter 20: Testing Hypotheses About Proportions

Key Vocabulary:

- hypothesis
- null hypothesis
- reject (the null hypothesis)
- fail to reject (the null hypothesis)
- alternative hypothesis

Calculator Skills:

- 1-Prop ZTest

1. What is a *hypothesis*?

2. After analyzing a set of data, if the results support a *hypothesis,* does that prove the *hypothesis* is true? Explain.

3. After analyzing a set of data, if the results are inconsistent with a *hypothesis,* does that prove the *hypothesis* is false? Explain.

4. What does it mean to *reject* a *hypothesis*?

5. When testing *hypotheses*, always start by assuming that the *null hypothesis* is true. What is meant by a *null hypothesis*?

6. Given a *null hypothesis* $H_0 : p = p_0$, what are the parameters of the Normal sampling distribution model? Under what conditions is this model appropriate?

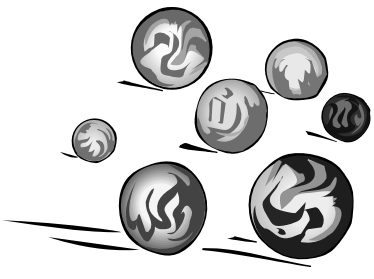
7. How would you determine whether a particular value of $\hat{p}$ is so unlikely to have occurred (assuming $p = p_0$) that you would *reject* the *null hypothesis*?

8. Why do we say we "*fail to reject*" the *null hypothesis* rather than "accept" the *null hypothesis*?

9. What is meant by an *alternative hypothesis*?

10. What is meant by a P-value?

11. Explain the difference between a two-sided alternative hypothesis and a one-sided alternative hypothesis. Draw a sketch



Chapter 21: More About Tests

Key Vocabulary:

- P-value
- statistically significant
- alpha level
- significance level
- Type I Error
- Type II Error
- power

1. Explain what the *p-value* represents.

2. What is meant by an *alpha level*?

3. What does it mean for a result to be *statistically significant*?

4. A 95% confidence interval corresponds to a two-sided hypothesis test at what *alpha level*?

5. A 90% confidence interval corresponds to a one-sided hypothesis test at what *alpha level*?

6. Explain the difference between a *Type I* and *Type II Error*.

7. What is the probability of a *Type I Error*?

8. What is meant by the *power* of a test?

9. How do you calculate the *power* of a test?

Chapter 22: Comparing Two Proportions

Key Vocabulary:

- pooling

Calculator Skills:

- 2-Prop Z-Int
- 2-Prop Z-Test

1. What conditions and assumptions are necessary for the sampling model of $\hat{p}_1 - \hat{p}_2$ to be approximately Normal?

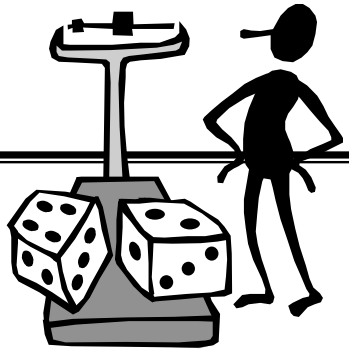
2. If the above conditions and assumptions are met, what is the mean and standard deviation of the sampling model?

3. Describe how to construct a level C confidence interval for the difference between two proportions, $p_1 - p_2$.

4. Explain what is meant by pooling two samples. When is it appropriate to pool samples?

5. For a two-sample hypothesis test where $H_0 : p_1 - p_2 = 0$, show how to calculate the z test statistic?

Chapter 23: Inferences About Means



Key Vocabulary:

- t-distribution
- t-table
- degrees of freedom
- one-sample t-interval
- one-sample t-test

Calculator Skills:

- T-Interval
- T-test
- tcdf (leftend, rightend, df)

1. What is the *standard deviation* of the sample mean $\bar{x}$?

2. What is the *standard error* of the sample mean $\bar{x}$?

3. Describe the similarities between a *standard normal distribution* and a *t distribution.*

4. Describe the differences between a *standard normal distribution* and a *t distribution.*

5. How do you calculate the *degrees of freedom* for a *t distribution*?

6. What happens to the *t distribution* as the *degrees of freedom* increase?

7. How would you construct a level C confidence interval for μ if σ is unknown?

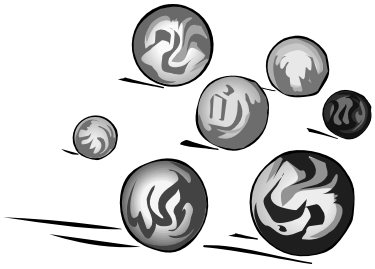
8. The *z*-Table gives the area under the standard normal curve to the left of *z*. What does the *t*-Table give?

9. Samples from normal distributions have very few outliers. If your data contains outliers, what does this suggest?

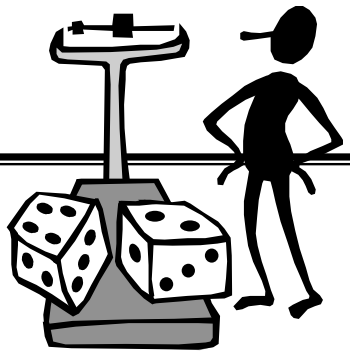
10. If the size of the SRS is less than 15, when can we use *t procedures* on the data?

11. If the size of the SRS is between 15 and 40, when can we use *t procedures* on the data?

12. If the size of the SRS is at least 40, when can we use *t procedures* on the data?



Chapter 24: Comparing Means



Key Vocabulary:

- two-sample t-interval
- two-sample t-test
- pooled t-test

Calculator Skills:

- 2-SampTInt
- 2-SampTTest

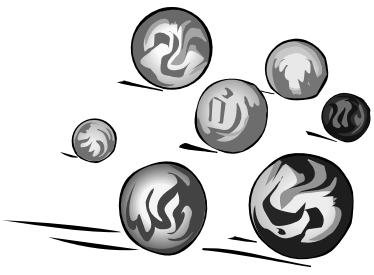
1. Describe the assumptions and conditions that must be met in order to use two-sample t-procedures.

2. In a two-sample problem, must/should the two sample sizes be equal?

3. In a two-sample problem, what is the null hypothesis for comparing two means?

4. Explain how to standardize $\overline{x}_1 - \overline{x}_2$ if σ_1 and σ_2 are unknown.

5. Explain why we ALWAYS pool our statistics when testing two proportions, but almost NEVER pool our statistics when testing two means.



Chapter 25: Paired Samples and Blocks

Key Vocabulary:

- paired t-test
- paired t-interval

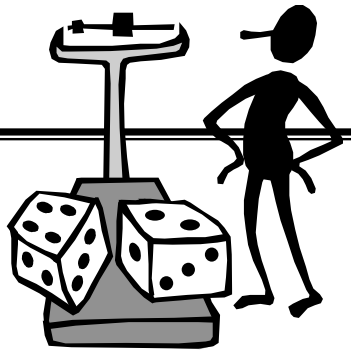
1. Explain why a two-sample hypothesis test is not appropriate for paired data.

2. What type of test is used instead of a two-sample t-test?

3. What conditions and assumptions are necessary in order to use a paired t-test?

4. Explain how to construct a confidence interval for paired data.

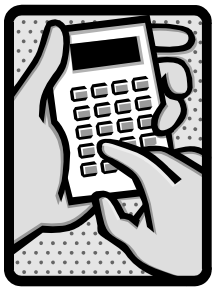
Chapter 26: Comparing Counts



Key Vocabulary:

- chi-square test for goodness of fit
- chi-square statistic
- expected count
- observed count
- degrees of freedom
- chi-square distribution
- components of chi-square
- cell counts
- r x c table

Calculator Skills:



- sum ( )
- χ^2cdf (leftbound, rightbound, df)
- χ^2pdf (X, df)
- χ^2-Test

1. State the null and alternative hypotheses for a *goodness of fit test*.

2. What conditions must be met in order to use a *goodness of fit test*?

3. What is the *chi-square statistic*?

4. What is the difference between the notation X^2 and χ^2?

5. How many degrees of freedom does the *chi-square distribution* have?

6. As the *chi-square statistic* increases, what happens to the P-value?

7. In order to reject the null hypothesis, what will be true about the *chi-square statistic*?

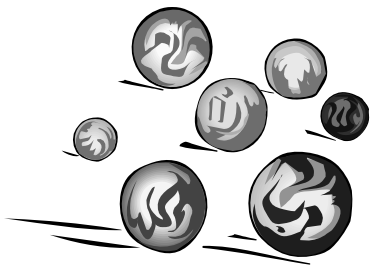
8. Describe the domain of a *chi-square distribution*?

9. What is the shape of a *chi-square distribution*? What happens to the shape as the degrees of freedom increases?

10. Explain the difference between a *chi-square test for goodness-of-fit* and a *chi-square test of homogeneity*.

11. State the null and alternative hypotheses for comparing more than two proportions.

12. How do you calculate the expected count in any cell of a two-way table when the null hypothesis is true?



13. How many degrees of freedom does a chi-square test for a two-way table with r rows and c columns have?

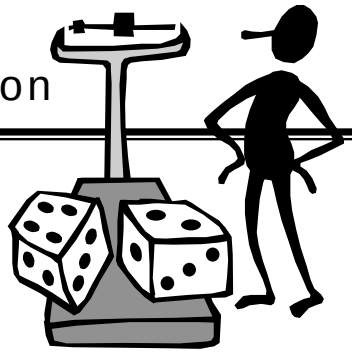
14. If you decide to reject the null hypothesis when performing a chi-square test of homogeneity, it is necessary to follow up by examining the residuals of the components. Explain how to standardize the residuals for the components of the chi-square test.

15. What would you look for when interpreting these standardized residuals?

16. State the null and alternative hypotheses for a *chi-square test for independence*.

17. Explain the difference between a chi-square test of homogeneity and a *chi-square test for independence* (think about the hypotheses for each).

Chapter 27: Inferences for Regression



Key Vocabulary:

- true regression line
- standard error about the line
- degrees of freedom
- prediction interval

Calculator Skills:

- LinRegTTest

1. What is the equation for the *idealized regression line*? Label each part of the equation.

2. What is the equation for the *least-squares regression line*? Label each part of the equation.

3. What are the four assumptions and conditions that must be met in order to perform inference for regression? Explain how to check each one.

4. What three aspects of the scatterplot affect the standard error of the regression slope?

5. If there is no association between two variables, what should be the value of the slope of the regression line?

6. When performing inference on the regression slope, what are the null and alternative hypotheses?

7. How many degrees of freedom does the standardized extimated regression slope have?

8. What is the formula for constructing a confidence interval for the regression slope β?

9. How is this similar to the formula for constructing other confidence intervals?

10. If you want to test for no correlation between your variables *x* and *y*, what should be used for the null hypothesis?

11. When using the LinRegTTest on the TI-83+, what does the symbol ρ represent?