Francis Howell School District
Mission Statement

Francis Howell School District is a learning community where all students reach their full potential.

Vision Statement

Francis Howell School District is an educational leader that builds excellence through a collaborative culture that values students, parents, employees, and the community as partners in learning.

Values

Francis Howell School District is committed to:

- Providing a consistent and comprehensive education that fosters high levels of academic achievement for all
- Operating safe and well-maintained schools
- Promoting parent, community, student, and business involvement in support of the school district
- Ensuring fiscal responsibility
- Developing character and leadership

Francis Howell School District Graduate Goals

Upon completion of their academic study in the Francis Howell School District, students will be able to:

1. Gather, analyze and apply information and ideas.
2. Communicate effectively within and beyond the classroom.
3. Recognize and solve problems.
4. Make decisions and act as responsible members of society.

Mathematics Graduate Goals

Upon completion of their mathematics study in the Francis Howell School District, students will be able to:

1. Communicate mathematically
2. Reason mathematically
3. Make mathematical connections
4. Use mathematical representations to model and interpret practical situations

Mathematics Rationale for AP Statistics

In today's global and technological society, production and consumption of data continues to increase, necessitating statistical literacy skills of reading, analyzing and interpreting data. Today's citizens must possess an understanding of data in order be a viable part of our world. As information travels with greater speed, statistics empowers us to make better, more accurate and faster decisions affecting multiple facets of our lives. Statistical analysis plays a key role in the fields of psychology, sociology, education, health-related professions, mathematics, physical and life sciences as well as in business, computer sciences and more. AP Statistics provides students with the necessary skills and meaningful applications to compete in today's society.

Course Description for Statistics

Students will study the major concepts and tools for collecting, analyzing, and drawing conclusions from data. The four broad conceptual themes are: exploring data, planning study, anticipating patterns, and statistical inference. TI-83 Graphing calculator or higher is required

Curriculum Team

Keith Looten
Steve Willott

Secondary Content Leader	Keiren Greenhouse
Director of Student Learning	Travis Bracht
Chief Academic Office	Dr. Pam Sloan
Superintendent	Dr. Renee Schuster

Curriculum Map (organized by chapters in primary textbook):

PART I: Organizing Data: Looking for Patterns and Departures from Patterns

Chapter 1: Exploring Data (12 days—one test)
- 1.1: Displaying Distributions with Graphs (including all types as referenced in Course Description)
- 1.2: Describing Distributions with Numbers

Chapter 2: The Normal Distributions (8 days—one test)
- 2.1: Density Curves and the Normal Distributions
- 2.2: Standard Normal Calculations

Chapter 3: Examining Relationships (15 days—one test)
- 3.1: Scatterplots
- 3.2: Correlation
- 3.3: Least-Squares Regression

Chapter 4: More on Two Variable Data (14 days—one test)
- 4.1: Transforming Relationships
- 4.2: Cautions about Correlation and Regression
- 4.3: Relations in Categorical Data

PART II: Producing Data: Samples, Experiments, and Simulations

Chapter 5: Producing Data (10 days—one test) Gummy Bear Project (see attached)
- 5.1: Designing Samples
- 5.2: Designing Experiments
- 5.3: Simulating Experiments

PART III: Probability: Foundations of Inference

Chapter 6: Probability: The Study of Randomness (8 days—one test)
- 6.1: The Idea of Probability
- 6.2: Probability Models
- 6.3: General Probability Rules

Chapter 7: Random Variables (7 days—one test)
- 7.1: Discrete and Continuous Random Variables
- 7.2: Means and Variances of Random Variables

End of First Semester (Comprehensive Semester Exam)

Chapter 8: The Binomial and Geometric Distributions (10 days—one test)
- 8.1: The Binomial Distributions
- 8.2: The Geometric Distributions

Chapter 9: Sampling Distributions (12 days—one test) Sampling Distribution Project (see attached)
- 9.1: Sampling Distributions
- 9.2: Sample Proportions
- 9.3: Sample Means

PART IV: Inference: Conclusions with Confidence

Chapter 10: Introduction to Inference (8 days—one test)
- 10.1: Estimating with Confidence
- 10.2: Tests of Significance
- 10.3: Making Sense of Statistical Significance
- 10.4: Inference as Decision

Chapter 11: Inference for Distributions (11 days—one test)
- 11.1: Inference for the Mean of a Population
- 11.2: Comparing Two Means

Chapter 12: Inference for Proportions (8 days—one test)
- 12.1: Inference for a Population Proportion
- 12.2: Comparing Two Proportions

Chapter 13: Inference for Tables: Chi-Square Procedures (7 days—one test)
- 13.1: Test for Goodness of Fit
- 13.2: Inference for Two-Way Tables

Chapter 14: Inference for Regression (4 days—one test)
14.1: Inference about the Model
14.2: Predictions and Conditions

End of AP test material (Comprehensive Semester Exam)

Content Area: Mathematics	**Course: AP Statistics**	**Strand: Data and Probability 1**
Learner Objectives: Students will explore data by describing patterns and departures from patterns.		

Concepts:
A: Construct and interpret graphical displays of distributions of univariate data
B: Summarize distributions of univariate data
C: Compare distributions of univariate data

Students Should Know	Students Should Be Able to
• The effect of changing units on summary measures • How shape determines the appropriate measures of center and spread	• Identify appropriate center and spread (IA1) • Identify outliers and other unusual features (IA2) • Assess shape including any cluster and gaps (IA3) • Calculate center: median, mean (IB1) • Calculate spread: range, interquartile range, standard deviation (IB2) • Identify measures of position: quartiles, percentiles, standardized scores (z-scores) (IB3) • Create boxplots and use them to compare distributions (IB4) • Select the appropriate graphical display (IC1, IC2, IC3, IC4)

Instructional Support

Student Essential Vocabulary					
Data	Parameter	Population	Sample	Statistic(s)	Outlier(s)
Quantitative Variable	Standardized Value	Transformations	Variability	Z Score	Extrapolation
Transforming Data	Proportion	5 Number Summary	68-95-99.7 Rule	Bimodal	Boxplot

Center	Degrees of Freedom	Distribution	Dotplot	Expected Value	Histogram
Interquartile Range (IQR)	Mean	Median	Midrange	Mode	Normal
Percentile	Quartile	Range	Shape	Skew	Spread
Standard Deviation	Standard Normal Model	Stem-and-Leaf	Symmetric	Tails	Time Plot
Uniform	Unimodel	Variance			

Sample Learning Activities	Sample Assessments
Learning Activity #1 : The data (***See Appendix***) shows the various roller coasters that Mr. Wills rode from October of 2008 through February of 2010. The variables reported include the name of each roller coaster, the amusement park where it is located, the material of which it is made, the type of roller coaster, its length, its height, its largest drop, the number of inversions, its maximum speed, its duration, its maximum G-Force, and its maximum angle of descent. Identify whether each variable is categorical or quantitative and its units. Select a quantitative variable and make an appropriate graph of your choice. Describe the distribution of this variable. Make box plots to compare the length of roller coasters at the various parks. Use comparative language to compare the collections of rides at these parks. height of roller coasters at the various parks. Use comparative language to compare the collections of rides at these parks.	**Assessment #1:** Before deciding whether to drill more wells, it is key to decide how much oil the present wells will produce. Here are the estimated total amounts of oil recovered from 64 wells in the Devonian Richmond Dolomite area of the Michigan basin, in thousands of barrels. 21.7 53.2 46.4 42.7 50.4 97.7 103.1 51.9 43.4 69.5 156.5 34.6 37.9 12.9 2.5 31.4 79.5 26.9 18.5 14.7 32.9 196 24.9 118.2 82.2 35.1 47.6 54.2 63.1 69.8 57.4 65.6 56.4 49.4 44.9 34.6 92.2 37.0 58.8 21.3 36.6 64.9 14.8 17.6 29.1 61.4 38.6 32.5 12.0 28.3 204.9 44.5 10.3 37.7 33.7 81.1 12.1 20.1 30.5 7.1 10.1 18.0 3.0 2.0

speed of roller coasters at the various parks. Use comparative language to compare the collections of rides at these parks.

duration of roller coasters at the various parks. Use comparative language to compare the collections of rides at these parks.

Make box plots to compare the

length of roller coasters made of steel vs. those made of wood. Use comparative language to compare the rides made of these materials.

height of roller coasters made of steel vs. those made of wood. Use comparative language to compare the rides made of these materials.

speed of roller coasters made of steel vs. those made of wood. Use comparative language to compare the rides made of these materials.

duration of roller coasters made of steel vs. those made of wood. Use comparative language to compare the rides made of these materials.

Make boxplots to compare the

length of the type of roller coasters (sit down, inverted, stand up, etc.). Use comparative language to compare the types of rides.

height of the type of roller coasters (sit down, inverted, stand up, etc.). Use comparative language to compare the types of rides.

speed of the type of roller coasters (sit down, inverted, stand up, etc.). Use comparative language to compare the types of rides.

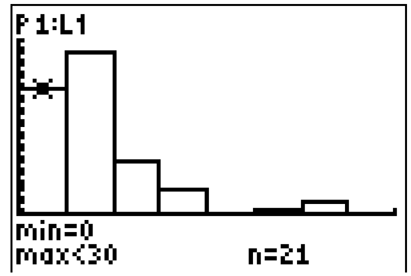
duration of the type of roller coasters (sit down, inverted, stand up, etc.). Use comparative language to compare the types of rides.

Solution:
See Appendix

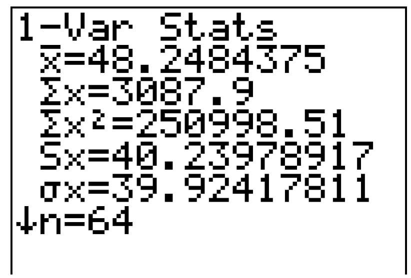
Activity's Alignment

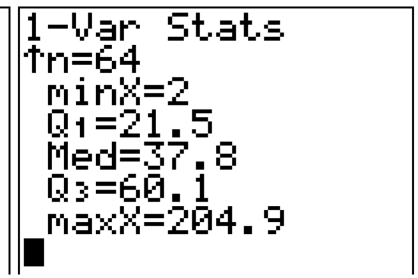
- Construct an appropriate graph of the distribution and describe its shape, center, and spread.

Solutions may vary:



or other appropriate graphical display. The shape is skewed right, with a couple of outliers to the right.





The center would be the median of 37.8 with IQR of 60.1-21.5 = 38.6.

Assessment's Alignment	
CONTENT	MA3 Data analysis

CONTENT	MA3 Data analysis
PROCESS	1.8 Organize data and ideas
DOK	2 Skill/concept
INSTRUCTIONAL STRATEGIES	Nonlinguistic representation

Learning Activity #2:

Death Rate ***See Appendix***

With your 1.69oz. bag of M&Ms; carefully dump them onto your desk. The ones that land with the **M** side up have an incurable **M**alady and have passed away. Count the number that have the **M**alady and the total number in your package and calculate the death rate (to the nearest whole percent).

Record your death rate here:

____

Mark the appropriate chart with the death rate.

Count the number of each color of M&M and record them below

Red	Orange	Yellow	Green	Blue	Brown	Totals

Mark the appropriate chart with the count for each color from your bag.

PROCESS	1.8 Organize data and ideas
DOK	2 Skill/concept
LEVEL OF EXPECTATION	Mastery Level – 85%

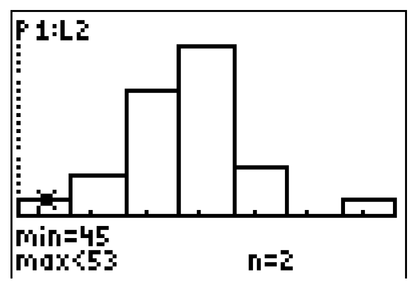
Assessment #2:

The following data represent scores of 50 students on a calculus test.

72	72	93	70	59	78	74	65	73
80	57	67	72	57	83	76	74	56
68	67	74	76	79	72	61	72	73
76	67	49	71	53	67	65	100	83
69	61	72	68	65	51	75	68	75
66	77	61	64	74				

- Construct an appropriate graph of the distribution and describe its shape, center, and spread.

Solutions may vary:



or other appropriate graphical display. The shape is roughly symmetric, with possible outliers to the right.

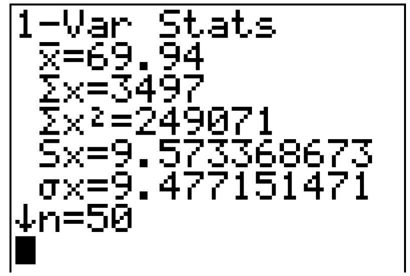
Distributions of data should always address four things: Shape, Center, Spread and possibly Outliers (SOCS).

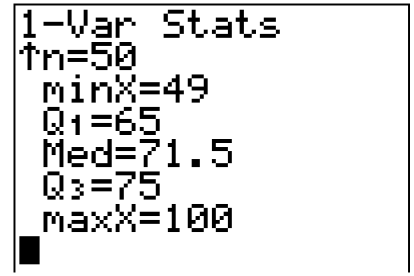
For the class data collected, record the values you feel best represent the above.

	Shape	Center	Spread	Any Outliers?
Death Rate				
Red				
Orange				
Yellow				
Green				
Blue				
Brown				
Total No.				

Sample solutions:
See Appendix

Activity's Alignment		
CONTENT	MA3	Data analysis
PROCESS	1.8	Organize data and ideas
DOK	2	Skill/concept





The center would best be described by the mean of 69.94 and the spread would best be described by the standard deviation of approximately 9.57.

- What would happen to the measure of center and spread if the teacher added five extra points to everyone's grade?

Solution: The measure of center would increase by 5 and the measure of spread would remain the same.

- What would happen to the measure of center and spread if the teacher doubled everyone's score?

Solution: The measure of center and spread would both be doubled.

Assessment's Alignment		
CONTENT	MA3	Data analysis
PROCESS	1.8	Organize data and ideas
DOK	2	Skill/concept
LEVEL OF EXPECTATION	Mastery Level – 85%	

INSTRUCTIONAL STRATEGIES	Nonlinguistic representation	

Student Resources	Teacher Resources
The Practice of Statistics	The Practice of Statistics; Yates, Moore, Starnes; Freeman; ©2003; ISBN # 0-7167-4773-1

Identity Equity and Readiness			
Gender Equity		Technology Skills	
Racial/Ethnic Equity		Research/Information	
Disability Equity		Workplace/Job Prep	

Content Area: Mathematics	**Course: AP Statistics**	**Strand: Data and Probability 2**
Learner Objectives: Students will explore data by describing patterns and departures from patterns		

Concepts: D: Explore bivariate data

Students Should Know	Students Should Be Able to
● Limitations of correlations and when correlation is appropriate ● How to calculate least-squares regression line ● How to create residual plots, outliers, and influential points	● Analyze patterns in scatterplots (ID1) ● Interpret correlation (ID2) ● Assess linearity (ID2) ● Use least-squares regression for predictions (ID3) ● Interpret residual plots, outliers, and influential points (ID4) ● Perform transformations to achieve linearity: logarithmic and power transformations (ID5)

Instructional Support

<table>
<tr><th colspan="6">Student Essential Vocabulary</th></tr>
<tr><td>Data</td><td>Parameter</td><td>Population</td><td>Sample</td><td>Statistic(s)</td><td>Outlier(s)</td></tr>
<tr><td>Quantitative Variable</td><td>Transformations</td><td>Transforming Data</td><td>Variability</td><td>Z Score</td><td>Association</td></tr>
<tr><td>Standardized Value</td><td>Causation</td><td>Central Limit Theorem</td><td>Direction</td><td>Explanatory</td><td>Form</td></tr>
<tr><td>Exponential Model</td><td>Influential Point</td><td>Intercept</td><td>Line of Best Fit</td><td>Linear Model</td><td>Lurking Variables</td></tr>
<tr><td colspan="2">Coefficient of Determination (R^2)</td><td>Correlation Coefficient (r)</td><td>Model</td><td>Power Model</td><td>Predicted Value</td></tr>
<tr><td colspan="2">Least Squares Regression (LSRL)</td><td>Prediction</td><td>Regression</td><td>Regression Outliers</td><td>Residual</td></tr>
<tr><td colspan="2">Regression to the Mean</td><td>Residual Plot</td><td>Response</td><td>Response Variable</td><td>Scatterplot</td></tr>
<tr><td colspan="2">Slope (Rate of Change)</td><td>Strength</td><td>Ladder of Powers</td><td>Logarithmic Model</td><td>Monotonicity</td></tr>
<tr><td>Subset</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Sample Learning Activities	Sample Assessments

Learning Activity #1 :

Modeling the Spread of a Disease

A disease in a community may begin with 1 person, who then spreads the disease to a friend or acquaintance. Eventually, each person may spread the disease to other people. This process continues until there is some intervention to interrupt the spread of the disease or until the patient dies. In this activity, you will simulate the spread of disease in a community.

1. The first student (present) on my alphabetical class list will represent the first infected person. That person moves to one side of the room and rolls a die (singular of dice) repeatedly, with each roll representing a unit of time. The number 5 will signal a transmission of the disease to another uninfected person. When a 5 is rolled, a new student is chosen from the class to receive a die and represent an additional infected person. This additional person joins the first student so that there are now 2 infected individuals at one side of the room, perhaps a corner or at the front.

2. As the die is rolled, everyone should plot points on the graph on the other side of this paper. "Time" is marked as the explanatory variable on the horizontal axis and "number of infected people" is marked as the response variable on the vertical axis. The points that everyone graphs will form a scatter plot.

3. At the signal from the teacher, each "infected person" will roll his or her die. If anyone rolls a 5, a new student will be chosen from the class to join the group of infected people. For each new 5, a new person becomes "infected" (so if 2 students roll 5s, 2 additional students are selected for that time period, if no one rolls a 5, that's a period of time, but with no additional disease transmission). After each signal from the teacher to roll the die/dice, the class counts the number of infected individuals and plots a point for that year. The simulation continues until all students in the class have become "infected".

4. Bring your data back to class tomorrow and think about this:

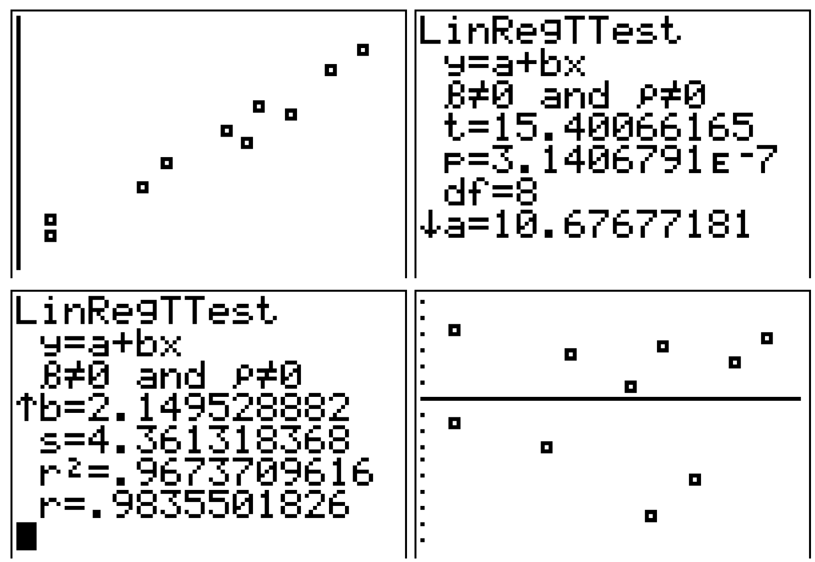
Assessment #1:

The Great Plains Railroad is interested in studying how fuel consumption is related to the number of railcars for its trains on a certain route between Oklahoma City and Omaha. A random sample of 10 trains on this route has yielded the data in the table below.

# of Railcars	20	20	37	31	47	43	39
	50	40	29				
Fuel Cons.	58	52	91	80	114	98	87
1	22	100	70				
(units/mile)							

- Create an appropriate regression equation to represent the data.

Solution**:** The scatter plot of the data shows a relatively strong positive, seemingly linear relationship. The correlation (r = 0.9836), coefficient of determination ($R^2 = 96.7\%$) and having no discernable pattern in the residual plot all support a linear relationship



a. Do the points show a pattern or association?
b. If so, is the pattern linear?
c. What mathematical function would best describe the pattern of points?

Sample Solution:
See Appendix

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.10 Apply information, ideas and skills
DOK	2 Skill/concept
INSTRUCTIONAL STRATEGIES	

Learning Activity #2:

How to Weigh a Gator!

Many wildlife populations are monitored by taking aerial photographs. Information about the number of animals and their whereabouts is important

Therefore the equation is

$$\widehat{fuel}\ consumption = 10.6768 + 2.1495(\text{Rail cars})$$

- What percent of the variation in fuel consumption is explained by the linear equation?

Solution: $R^2 = 96.7\%$

- Predict the fuel consumption for a train with 50 rail cars.

Solution: $\widehat{fuel}\ consumption = 10.6768 + 2.1495(50) = 118.15$

- What is the residual for the eighth point (50, 122) above?

Solution: 122 – 118.15 = 3.85

Assessment's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.10 Apply information, ideas and skills
DOK	2 Skill/concept
LEVEL OF EXPECTATION	Mastery level –

Assessment #2:

The number of motor vehicles registered in the U.S. has grown as follows:

to protecting certain species and to ensuring the safety of surrounding human populations. In addition, it is sometimes possible to monitor certain characteristics of the animals. The length of an alligator can be estimated quite accurately from aerial photographs or from a boat. However, the alligator's weight is much more difficult to determine. ("YOU weigh him." "No, YOU weigh him!")

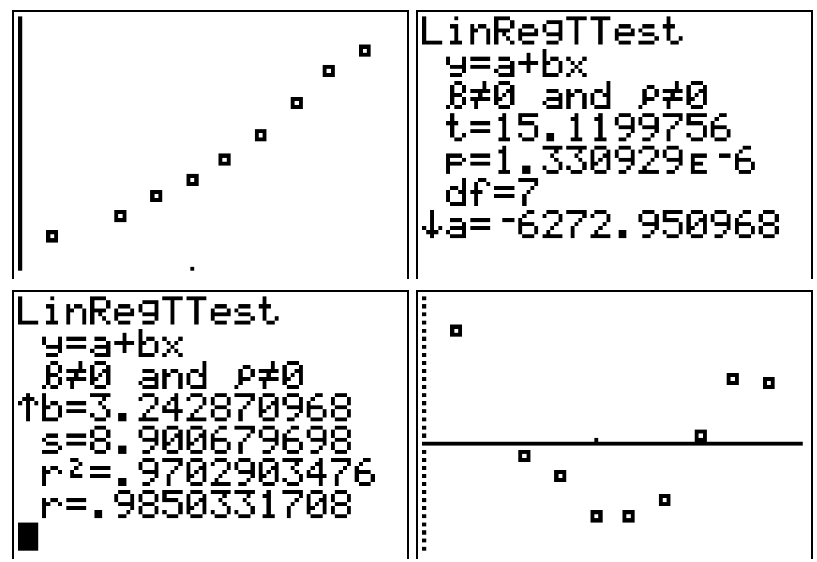
Length (in)	Weight (lbs)
58	2.8
61	44
63	33
68	39
69	36
72	38
72	61
74	54
74	51
76	42
78	57
82	80
85	84
86	80
85	84
86	90
88	70
89	84
90	106
90	102
94	110
94	130
114	197
128	366

In the example below, data on the length (in inches) and weight (in pounds) of alligators captured in central Florida are given. Your task is to develop a model from which the weight of an alligator can be predicted from its length.

Year	Vehicles	Year	Vehicles
1940	32.4	1965	90.4
1945	31.0	1970	108.4
1950	49.2	1975	132.9
1955	62.7	1980	155.8
1960	73.9	1985	171.7

- Create an appropriate regression equation to represent the data. (The number for 1945 is an outlier. Why might it be an outlier? You may delete it.)

Solution**:** With the outlier removed (probably due to WWII), the scatter plot of the data shows a relatively strong positive, possibly non-linear relationship. While the correlation (r = 0.9850) and coefficient of determination ($R^2 = 97.0\%$) are both relatively strong, the residual plot has a clear pattern, meaning the linear relationship is NOT appropriate.

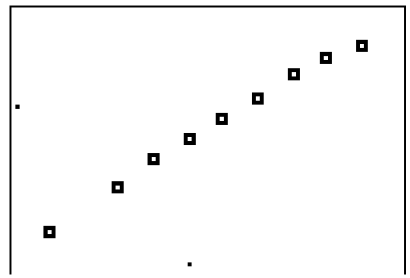


The Report. Describe your investigation in a report. Tell the story, from the introduction to the analysis to the conclusions with all of the necessary supporting calculator screen shots or computer plots and numerical summaries. In particular, write your report so that the reader can follow your reasoning as you proceed through your investigation. Follow the conventions as described in the general guidelines for writing up Special Problems.

Sample Solution:

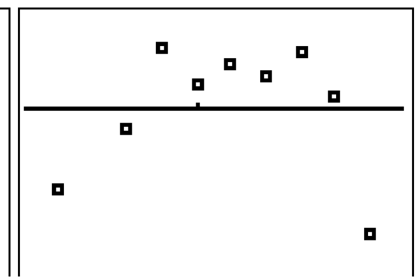
power LinReg (ax + b) L_3 ,L_4

The ratios from 1950 on (0.785, 0.848, 0.817, 0.834, 0.816, 0.853, and 0.907) indicate an exponential model may be appropriate. The scatter plot of the transformed data shows a relatively strong, positive, seemingly linear relationship. The correlation (r = 0.9970), coefficient of determination ($R^2 = 99.4\%$) and having no discernable pattern in the residual plot all support a linear relationship (Although the last point may be an outlier and other methods learned in this class show no improvement to the model).



```
LinRegTTest
 y=a+bx
 β≠0 and ρ≠0
 t=34.0465227
 p=4.8879935E-9
 df=7
↓a=-30.24059802
```

```
LinRegTTest
 y=a+bx
 β≠0 and ρ≠0
↑b=.0163787254
 s=.019964201
 r²=.9939974225
 r=.9969941938
```



Therefore the LSRL on the transformed data is

$$\log \widehat{vehicle\ registration} = -30.2406 + 0.0164(Year)$$

Which yields an exponential regression model:

$$\widehat{vehicle\ registration} = 10^{-30.2406} \cdot 10^{0.0164(Year)}$$

L 3 is log (length), L 4 is log (weight)

$y = 10^{\log x^{3.286} - 4.42}$

$y = 10^{\log x^{3.286}} \cdot 10^{-4.42}$ since that 3.286 is an exponent on the x

Activity's Alignment	
CONTENT	MA3 Data analysis CA 1 Speaking and writing English
PROCESS	1.10 Apply information, ideas and skills
DOK	3 Strategic Thinking
INSTRUCTIONAL STRATEGIES	Summarizing and note taking

Assessment's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.10 Apply information, ideas and skills
DOK	3 Strategic Thinking
LEVEL OF EXPECTATION	Mastery level – 80%

Student Resources	Teacher Resources
The Practice of Statistics	The Practice of Statistics; Yates, Moore, Starnes; Freeman; ©2003; ISBN # 0-7167-4773-1

Identity Equity and Readiness			
Gender Equity		Technology Skills	
Racial/Ethnic Equity		Research/Information	
Disability Equity		Workplace/Job Prep	

Content Area: Mathematics	**Course: AP Statistics**	**Strand: Data and Probability 3**
Learner Objectives: Students will explore data by describing patterns and departures from patterns		

Concepts: E: Explore categorical data

Students Should Know	Students Should Be Able to
• How to select appropriate relative frequency • Features of graphical displays that indicate an association between two variables	• Construct and interpret frequency tables and bar charts (IE1) • Calculate marginal and joint frequencies for two-way tables (IE2) • Calculate conditional relative frequencies and association (IE3) • Compare distributions using bar charts (IE4)

Instructional Support

Student Essential Vocabulary					
Data	Parameter	Population	Sample	Statistic(s)	Extrapolation
Proportion	Bar Chart	Categorical Variable	Conditional Distribution	Contingency Table	Frequency Table
Marginal Distribution	Pie Chart	*Simpson's Paradox*			

Sample Learning Activities

Learning Activity #1 :

A simple random sample of adults living in a suburb of a large city was selected. The age and annual income of each adult in the sample were recorded. The resulting data are summarized in the table below.

	Annual Income			
Age Category	$25,000 - $35,000	$35,001 - $50,000	Over $50,000	Total
21-30	8	15	27	50
31-45	22	32	35	89
46-60	12	14	27	53
Over 60	5	3	7	15
Total	47	64	96	207

What is the probability that a person chosen at random from those in this sample will be in the 31-45 age category?

What is the probability that a person chosen at random from those in this sample whose incomes are over $50,000 will be in the 31-45 age category? Show your work.

Based on your answers to the previous questions, is annual income independent of age category for those in the sample? Explain.

Solution:

$$P(\text{age } 31\text{-}45) = \frac{89}{207} = 0.42995$$

Sample Assessments

Assessment #1:

In a 1980 study, researchers looked at the relationship between the type of college (public or private) attended by 3265 members of the class of 1960 who went into industry and the level of job each member had in 1980. The results were:

Management Level	Public	Private
High	75	107
Middle	962	794
Low	732	595

- Compute the marginal counts.

Solution:

Management Level	Public	Private	Totals
High	75	107	182
Middle	962	794	1756
Low	732	595	1327
Totals	1769	1496	3265

Compute the conditional distributions of management level given college type (in percents). [Write the numbers next to the counts in the above table.]

Solution:

Management Level	Public%	Private%
High	4.24	7.15
Middle	54.38	53.07
Low	41.38	39.77

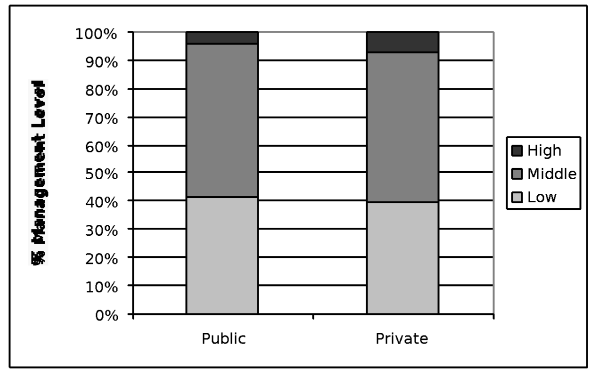
$$P(\text{age 31-45/income over 50,000}) = \frac{35}{96} = 0.36458$$

If annual income and age were independent, the probabilities in the questions above would be equal. Since these probabilities are not equal, annual income and age category are not independent for adults in this sample.

Activity's Alignment		
CONTENT	MA3	Data analysis
PROCESS	1.10	Apply information, ideas, and skills
DOK	3	Strategic Thinking

- Construct a **segmented bar graph**

Solution:



- Comment on the observed relationship.

Solution**:** The percent of graduates that went into low or middle management jobs was about the same for public and private schools. But significantly more private college graduates went into high level management jobs than the public college grads.

Assessment's Alignment		
CONTENT	MA3	Data analysis
PROCESS	1.10	Apply information, ideas, and skills
DOK	3	Strategic Thinking

INSTRUCTIONAL STRATEGIES	Summarizing and note taking

Learning Activity #2:

The ELISA tests whether a patient has contracted HIV. The ELISA is said to be positive if it indicates that HIV is present in a blood sample, and the ELISA is said to be negative if it does not indicate that HIV is present in a blood sample. Instead of directly measuring the presence of HIV, the ELISA measures levels of antibodies in the blood that should be elevated if HIV is present. Because of variability in antibody levels among human patients the ELISA does not always indicate the correct result.

As part of a training program, staff at a testing lab applied the ELISA to 500 blood samples known to contain HIV. The ELISA was positive for 489 of those blood samples and negative for the other 11 samples. As part of the same training program, the staff also applied the ELISA to 500 other blood samples known to not contain HIV. The ELISA was positive for 37 of those blood samples and negative for the other 463 samples.

When a new blood arrives at the lab, it will be tested to determine whether HIV is present. Using the data from the training program, estimate the probability that the ELISA would be positive when it is applied to a blood sample that does not contain HIV.

Among the blood samples examined in the training program that provided positive ELISA results for HIV, what proportion actually contained HIV?

When a blood sample yields a positive ELISA result, two more ELISAs are performed on the same blood sample. If at least one of the two additional ELISAs is positive, the blood sample is subjected to a more expensive and more accurate test to make a definitive determination of whether HIV is present in the sample. Repeated ELISAs on the same sample are generally assumed to be independent. Under the assumption of independence, what is

LEVEL OF EXPECTATION	Mastery level – 75%

Assessment #2:

The following data are the survival rates of the passengers and crew for the Titanic.

	First	Second	Third	Crew	TOTAL
Lived	202	118	178	212	**710**
Died	123	167	528	673	**1491**
TOTAL	**325**	**285**	**706**	**885**	**2201**

- What % of those aboard were 1st class passengers?

Solution: $\frac{325}{2201} = 14.8\%$

- What % of those aboard survived?

Solution: $\frac{710}{2201} = 32.3\%$

- Of those in 1st class what % survived?

Solution: $\frac{202}{325} = 62.2\%$

- Of those that survived, what % was 1st class?

Solution: $\frac{202}{710} = 28.5\%$

the probability that a new blood sample that comes into the lab will be subjected to the more expensive test if that sample does not contain HIV?

Solution:

The estimated probability of a positive ELISA if the blood sample does not have HIV present is

$$\frac{37}{500} = 0.074$$

A total of 489 + 37 = 526 blood samples resulted in a positive ELISA. Of these, 489 samples actually contained HIV. Therefore the proportion of samples that resulted in a positive ELISA that actually contained HIV is

$$\frac{489}{526} \approx 0.9297$$

The probability that the ELISA will be positive, given that the blood sample does not actually have HIV present, is 0.074. Thus, the probability of a negative ELISA, given that the blood sample does not actually have HIV present, is 1-0.074 = 0.926

P(new blood sample that does not contain HIV will be subjected to the more expensive test)
= P(1^{st} ELISA positive and 2^{nd} ELISA positive OR 1^{st} ELISA positive and 2^{nd} ELISA negative and 3^{rd} ELISA positive/HIV not present in blood)
=P(1^{st} ELISA positive and 2^{nd} ELISA positive/HIV not present in blood) + P(1^{st} ELISA positive and 2^{nd} ELISA negative and 3^{rd} ELISA positive/HIV not present in blood)
=(0.074)(0.074) + (0.074)(0.926)(0.074)
=0.005476 + 0.005070776
=0.010546776
≈0.0105

OR
P(1^{st} ELISA positive and not both the 2^{nd} and 3^{rd} are negative)

- What % of those aboard were 1^{st} class passengers that survived?

Solution: $\frac{202}{2201} = 9.2\%$

$=(0.074)(1 - 0.926^2)$
$=(0.074)(0.142524)$
$=0.010546776$
≈ 0.0105

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.10 Apply information, ideas and skills
DOK	3 Strategic Thinking
INSTRUCTIONAL STRATEGIES	Homework and practice

Assessment's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.10 Apply information, ideas and skills
DOK	3 Strategic Thinking
LEVEL OF EXPECTATION	Mastery level – 90%

Student Resources	Teacher Resources
The Practice of Statistics	The Practice of Statistics; Yates, Moore, Starnes; Freeman; ©2003; ISBN # 0-7167-4773-1

Identity Equity and Readiness			
Gender Equity		Technology Skills	
Racial/Ethnic Equity		Research/Information	
Disability Equity		Workplace/Job Prep	

Content Area: Mathematics	Course: AP Statistics	Strand: Data and Probability 4
Learner Objectives: Students will sample and experiment by planning and conducting a study.		

Concepts: B: Plan and conduct surveys

Students Should Know	Students Should Be Able to
• Characteristics of a well-designed and well-conducted survey • Differences between populations, samples, and methods of random selection • Sources of bias in sampling and surveys	• Create an SRS for surveys (IIB) • Identify potential problems resulting from samples and surveys (IIB)

• Sampling methods, including simple random sampling, stratified random sampling, and cluster sampling	

Instructional Support

Student Essential Vocabulary					
Data	Parameter	Population	Sample	Statistic(s)	Variability
Random	Simulation	Bias	Random Selection	Census	Cluster
Convenience	Stratified	Multistage Design	Non-response Bias	Response Bias	Sampling Frame
Survey	Systematic	Undercoverage	Voluntary Response		

Sample Learning Activities	Sample Assessments
Learning Activity #1 : Survey Fill out the survey. ***See Appendix*** Gender: Male __ Female __ Eye color: _____	**Assessment #1:** A university's financial aid office wants to know how much it can expect students to earn from summer employment. This information will be used to set the level of financial aid. The population contains 3,478 students who have completed at least one year of study but have not yet graduated. A questionnaire will be sent to an SRS of 100 of these students, drawn from an alphabetized list.

Height: ______ (inches)
Birth Date: _______
Left/Right Handed? _______
Amount of sleep you got last night: __________________ hours
Ounces of soda consumed yesterday: _______
Favorite Delivery Pizza: __________
ACT Math Score: _______
ACT Composite Score: ______
GPA: _______
# of Honors Classes you are enrolled in this semester: _____
Pulse rate: __________ beats/minute (See Activity 1 on page 4 of Stats book)

Survey sample solution

Gender:	M
Eye color:	blue
Height:	73 inches
Birth Date:	3-14-1960
Handed:	Right
Amount of sleep last night:	8 hours
Ounces of soda consumed:	0
Favorite Delivery Pizza:	Fox's
ACT Math Score:	32
ACT Composite Score:	31
GPA:	3.97
# of honors classes:	3
Pulse Rate:	84

1. Which variables are categorical and which are quantitative?
2. What are some biases that might be present?
   Don't know some answers, may lie, may not do arithmetic correct
3. Why might a question about weight be biased?
4. Which variables could we make a bar chart from?
5. Which variable could we make a pie chart from?
6. Are any of the variables linked?

- Describe how you will label the students in order to select the sample.

Solution: Answers may vary. We will assign students the numbers 0001-3478. All other numbers are invalid and will be ignored and we will ignore repeats of numbers. The random digits will be parsed into 4 digit numbers and valid numbers selected. We will repeat the procedure until we have an appropriate number for our sample.

- Use Table B, beginning at line 105, to select the first five students in the sample.

Solution: The vertical "pipes" designate the parsed numbers. The underlined numbers are valid. The others (including repeated numbers) are ignored. We will stop after finding 5 valid numbers.

4711| 9112| 3264| 7257| 7007| 1837| 5071| 9635| 9404| 1230| 5859| 6428| 5304| 2669| 5907| 3626| 3076|

The students numbered 1230, 1837, 2669, 3076 and 3264 would be the first five students selected for inclusion in the questionnaire.

Possible Solutions:

1. Categorical: gender, eye color, birth date, handedness, favorite delivery pizza
   Quantitative: height, sleep, soda consumed, ACT scores, GPA, #s of honors, pulse rate
2. don't know some answers, may lie, may not do arithmetic correct
3. In our culture, more likely to underestimate their true weight.
4. Almost any.
5. favorite pizza place, eye color, gender etc. (if combined with rest of class)
6. Probably ACT scores

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.2 Conduct research
DOK	2 Skill/concept
INSTRUCTIONAL STRATEGIES	Summarizing and note taking

Learning Activity #2:
Vocabulary Sort and Notes

Take the student pages with the definitions and examples. Cut the boxes apart and place them in the appropriate box that matches up with the term, definitions, and examples

See Appendix

Assessment's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.6 Discover and evaluate relationships
DOK	2 Skill/concept
LEVEL OF EXPECTATION	Mastery level – 85%

Assessment #2:

In late 1995, a Gallup survey reported that Americans approved sending troops to Bosnia by 46 to 40 percent. The poll did not mention that 20,000 U.S. troops were committed to go. A CBS News poll mentioned the 20,000 figure and got the opposite outcome -- a 58 to 33 percent disapproval rate. Briefly explain why the mention of the number of troops would cause such a big difference in the poll results.

- Write the name for the kind of bias that is at work here.
- Explain the bias.

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.8 Organize data and ideas
DOK	2 Skill/concept
INSTRUCTIONAL STRATEGIES	Homework and practice

Solution: Answers may vary. The bias is in the wording of the question. When you find out that 20000 ARE COMMITTED to go, it may become more REAL to those being asked the question.

A church group interested in promoting volunteerism in a community chooses a SRS of 200 community addresses and sends members to visit these addresses during weekday working hours and inquire about the residents' attitude toward volunteer work. 60% of all respondents say that they would be willing to donate at least an hour a week to some volunteer organization. Bias is present in this sample design. Identify the type of bias involved and state whether you think the sample proportion obtained is higher or lower than the true population proportion.

- Write the name for the kind of bias that is at work here.
- Explain the bias.

Solution: Answers may vary. The bias is in undercoverage. The results are likely to be higher than the true population proportion because the survey was done during the workday. People at home during those times may have more of an opportunity to volunteer time than those who are working.

Assessment's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.7 Evaluate information
DOK	2 Skill/concept
LEVEL OF EXPECTATION	Mastery level – 85%

Student Resources	Teacher Resources
The Practice of Statistics	The Practice of Statistics; Yates, Moore, Starnes; Freeman; ©2003; ISBN # 0-7167-4773-1

Identity Equity and Readiness			
Gender Equity		Technology Skills	
Racial/Ethnic Equity		Research/Information	
Disability Equity		Workplace/Job Prep	

Content Area: Mathematics	Course: AP Statistics	Strand: Data and Probability 5
Learner Objectives: Students will sample and experiment by planning and conducting a study.		

Concepts: C: Plan and conduct experiments

Students Should Know	Students Should Be Able to
● Characteristics of a well-designed and well-conducted experiment ● Treatments, control groups, experimental units, random assignments, and replication ● Sources of bias and confounding, including placebo effect and blinding ● Differences between completely randomized design, randomized block design, including matched pairs design including uses of each	● Design an experiment (IIC) ● Randomly assign subjects to treatments (IIC)

Instructional Support

Student Essential Vocabulary					
Data	Parameter	Population	Sample	Statistic(s)	Variability
Random	Simulation	Bias	Independence	Association	Causation
Statistically Significant	Prediction	Explanatory	Lurking Variables	Response	Blinding
Blocking	Confounding	Common Response	Control	Double Blind	Experiment
Experimental Units	Factor	Level	Matched Pairs	Observational Study	Placebo
Placebo Effect	Prospective	Randomization	Replication	Retrospective	Single Blind
Subjects	Treatment				

Sample Learning Activities	Sample Assessments
Learning Activity #1 : Rolling Down the River A farmer has just cleared a new field for corn. It is a unique plot of land in that a river runs along one side. The corn looks good in some areas of the field but not others. The farmer is not sure that harvesting the field is worth the expense. He has decided to harvest 10 plots and use this information to estimate the total yield. Based on this estimate, he will decide whether to harvest the remaining plots. Sampling: Convenience Sample Simple Random Sample Stratified Sample – Vertical Stratified Sample – Horizontal Observations: 1. You have looked at four different methods of choosing plots. Is there a reason, other than convenience, to choose one method over another? 2. How did your estimates vary according to the different sampling methods you used? 3. Compare your results to someone else in the class. Were your results similar? 4. Pool the results of all students for the mean yields from the simple random samples and make a class box plot. Repeat for means from	**Assessment #1:** A medical study of heart surgery investigates the effect of a drug called a beta-blocker on the pulse rate of the patient during surgery. The pulse rate will be measured at a specific point during the operation. The investigators will use 20 patients facing heart surgery as subjects. You have a list of these patients, numbered 1 to 20, in alphabetical order. ➢ Outline as an algorithm (paragraph form) or in diagram form a randomized experimental design for this study. Solution: Answers may vary. We will assign potential patients the numbers 01-20. All other numbers are invalid and will be ignored and we will ignore repeats of numbers. The random digits will be parsed into 2 digit numbers and valid numbers selected until we have 10 subjects who will receive the beta-blockers. The others will receive no beta blockers. After the procedure we will compare the pulse rates during surgery of the two groups. The selection process for the 10 beta-blocker patients follows. The vertical “pipes” designate the parsed numbers. The underlined numbers are valid. The others (including repeated numbers) are ignored. We will stop after finding 10 valid numbers. 20| 63| 92| 31| 85| 59| 22| 51| 81| 92| 74| 79| 77| 04| 06| 63| 68| 55| 65| 55| 97| 50| 45| 66| 55| 79| 31| 97| 01| 70| 51| 00| 28| 33| 74| 35| 79| 37| 19| 07| 80| 06| 20| 25| 37| 38| 26| 47| 81| 20| 25| 04| 49| 76| 13| 88| 94| 00| 17| 24| 01| 15| 82| 23| 44| 91| 75| 76| 75| 40| 73| 70| 13| 43| 12| 50| 45| 79| 20| 95| 26| 53| 53| 68| 12| 00| 43| 69| 10| 73|

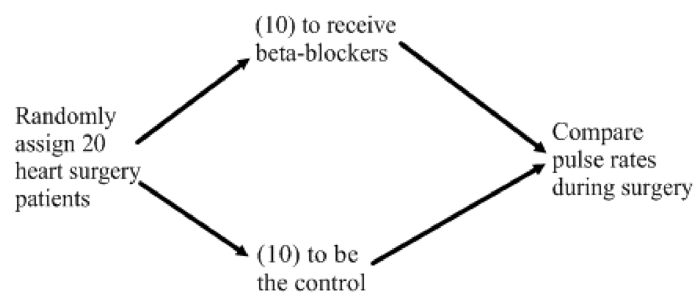
vertical strata and from horizontal strata. Compare the class box plots for each sampling method. What do you see?
5. Which sampling method should you use? Why do you think this method is best?
6. What was the actual yield of the farmer's field? How did the box plots relate to this actual value?

Solution:
1. One needs to choose a method that will give the best estimate of the yield. This can be affected by factors that cannot be controlled: e.g. the placement of the river. That's why one shouldn't choose the ten plots chosen by the farmer.
2. The student will see that the farmer's sample yields a very low estimate compared to the other methods used.
3. Comparing results with a peer helps the student verify that the sampling was done correctly. This does not mean the students will have the same sample, but each student should use the same process of drawing a sample for a given method. Some methods will produce highly variable results while others are much more consistent.
4. The variability of the means of the sample yields, as shown by the length of the box plot and the width of the middle 50%, will reduce drastically once the student has stratified appropriately. Thus the strata that are effective are the vertical ones, in which the values in each stratum are similar. This stratification reduces the variation in the sample means since the values chosen for a particular stratum vary little from sample to sample relative to the variability in the population.
5. Vertical stratification should be used since the sample would then include higher yielding plots as well as lower yielding ones.
6. The actual yield is 5004. The class box plot for the means resulting from the vertical stratification should be centered near 5004/100 or about 50.

See Appendix

The patients numbered 01, 04, 06, 07, 10, 12, 13, 15, 19 & 20 are the ones then that will receive the beta blockers during surgery.

By diagram:



Assessment's Alignment	
CONTENT	MA3 Data analysis

Activity's Alignment		
CONTENT	MA3	Data analysis
PROCESS	1.6	Discover and evaluate relationships
DOK	2	Skill/concept
INSTRUCTIONAL STRATEGIES	Generating and testing hypotheses	

Learning Activity #2:

When Does Blocking Help?

A set of 24 dogs (6 of each of four breeds; 6 from each of four veterinary clinics) has been randomly selected from a population of dogs older than eight years of age whose owners have permitted their inclusion in a study. Each dog will be assigned to exactly one of three treatment groups. Group "Ca" will receive a dietary supplement of calcium, Group "Ex" will receive a dietary supplement of calcium and a daily exercise regimen, and Group "Co" will be a control group that receives no supplement to the ordinary diet and no additional exercise. All dogs will have a bone density evaluation at the beginning and end of the one-year study. (The bone density is measured in Houndsfield units by using a CT scan.) The goals of the study are to determine (i) whether there are different changes in bone density over the year of the study for the dogs in the three treatment groups; and if so, (ii) how much each treatment influences that change in bone density.

Mechanics of the Simulations
The activity consists of three separate simulations, each involving its own particular process for allocating the dogs to the treatment groups.

See Appendix

Activity's Alignment		
CONTENT	MA3	Data analysis
PROCESS	1.6	Discover and evaluate relationships
DOK	2	Skill/concept
LEVEL OF EXPECTATION	Mastery level – 75%	

Assessment #2:

You are trying to determine the difference in dexterity levels between the dominant hand and the non-dominant hand. You decide to do this by having a person place washers one at a time using only one hand on a peg. You count how many the person is able to place on a peg in a 60 second time period.

- Explain how you would design a matched pairs experiment to test the difference in dexterity levels between the dominant and non-dominant hand.
- What purpose does it serve to make this a matched-pairs design?

Solution: Answers may vary. Select a large sample of subjects, determine their dominant hand. Then randomly assign half to use their dominant hand first in the 60 second interval and then the other half will use their non-dominant hand first. You will take the difference in the counts for the dominant hand minus the non-dominant hand to determine the average increase in dexterity. The "matched-pairs" is used to reduce variability within the individual counts due to general dexterity not associated with "handedness".

Assessment's Alignment		
CONTENT	MA3	Data analysis
PROCESS	3.5	Reason logically (inductive/deductive)
DOK	3	Strategic Thinking

PROCESS	1.3	Design/conduct investigations	LEVEL OF EXPECTATION	Mastery level – 85%
DOK	3	Strategic thinking		
INSTRUCTIONAL STRATEGIES	3.1	Identify and define problems		

Student Resources	Teacher Resources
The Practice of Statistics	The Practice of Statistics; Yates, Moore, Starnes; Freeman; ©2003; ISBN # 0-7167-4773-1

Identity Equity and Readiness			
Gender Equity		Technology Skills	
Racial/Ethnic Equity		Research/Information	
Disability Equity		Workplace/Job Prep	

Content Area: Mathematics	Course: AP Statistics	Strand: Data and Probability 6
Learner Objectives: Students will anticipate patterns by exploring random phenomena using probability and simulation.		

Concepts: A: Interpret probability

Students Should Know	Students Should Be Able to
• Interpreting probability, including long-run relative frequency interpretation • "Law of Large Numbers" concept • Effect of linear transformation of a random variable	• Use the addition rule, multiplication rule, conditional probability, and independence to calculate probabilities (IIIA3) • Identify discrete random variables and their probability distributions, including binomial and geometric (IIIA4)

	● Perform simulation of random behavior and probability distributions (IIIA5) ● Calculate mean (expected value) and standard deviation of a random variable, and perform linear transformation of a random variable (IIIA6)

Instructional Support

Student Essential Vocabulary					
Data	Parameter	Population	Sample	Statistic(s)	Variability
Random	Simulation	Bias	Random Selection	Independence	Statistically Significant
Transformations	Extrapolation	Proportion	68-95-99.7 Rule	Degrees of Freedom	Distribution
Dotplot	Histogram	Expected Value	Mean	Normal	Outliers
Shape	Skew	Spread	Standard Deviation	Stem-and-Leaf	Symmetric
Uniform	Variance	Central Limit Theorem	Model	Addition Rule	Complement
Assumptions and Conditions		Conditional Probability	Complement	Event	Multiplication Rule
Binomial Distribution Model		Continuous Random Variable		Discrete Random Variable	
Disjoint (Mutually Exclusive)		Geometric Distribution Model		Law of Large Numbers	Outcome
Normal Approximation to the Binomial		Probability Distribution		Sample Space	Tree Diagram
Probability Model		Random Variable		Trial	Venn Diagram
Conditions and Assumptions					

Sample Learning Activities	Sample Assessments
Learning Activity #1 : ***See Appendix*** After looking at several popular games of chance, it is now your turn to develop your own game using probability. You may work alone or with **<u>one</u>** partner. The following requirements need to be met: Step One: 1. Decide on a game that you would like to develop.	**Assessment #1:** Blood is categorized by two components, type and Rh factor. The following pie chart represents the percentage of the American population that falls into each category. Use the chart to answer the probability questions that follow.

2. Test your ideas through simulations of your game.

3. Will you be able to answer all of the questions in Step Two for your game?

4. Share your thoughts with instructor before proceeding.

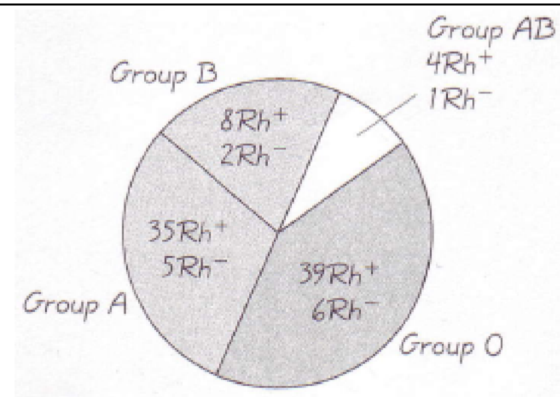
Step Two: Prepare your written project including each of the following:

1. Name your game.
2. Describe the rules of your game.
3. Play your game, **recording the results**, and determine the experimental probability of winning. "Playing of your game" may be done with the actual materials, or simulated with the calculator. Play a realistic number of times in order to feel somewhat confident about your experimental probability.
4. Determine the theoretical probability of winning your game.
5. If it costs $2 to play your game, what must be the payoff in order to make this a fair game?
6. Find two other individuals from our class to play your game. Provide them with the proper forms or worksheets to record their results.
7. How do these results (in problem 6) compare to the experimental and theoretical probabilities that you found in problems 3 and 4 above? Is there a significant difference?
8. Are there any improvements or alterations that you think should be made to your game?

Possible solution:

Game name: Snake-eyes

You roll two dice-If you get 2 ones, you win. Anything else, you lose.



- What is the P(O or A)?

Solution**:** 0.45 + 0.40 = 0.85

- What is the P(not O)?

Solution: 0.55

- What is the P(A or Rh+)?

Solution: 0.40 + 0.86 – 0.35 0.91

Experimental probability: Done using the ProbSim application on the calculator: $\frac{3}{201} = 0.0149$

Theoretical probability: $\frac{1}{36} = 0.0278$

Using the Theoretical probability of winning the probability distribution would be set up as follows:

Outcome	X	P(X)
2	???	1/36
Not 2	-$2	35/36

Since this must be a "fair game", the outcome (mean) should be zero.

Therefore

$$x\left(\frac{1}{36}\right) + -2\left(\frac{35}{36}\right) = 0$$

and solving for x gives us

$$x = 70$$

The payout would have to be $70 for this to be a fair game.

Activity's Alignment		
CONTENT	MA3	Data analysis
PROCESS	3.3	Apply one's own strategies
DOK	3	Strategic Thinking

INSTRUCTIONAL STRATEGIES	Homework and practice

Learning Activity #2:

Thumbtack Tossing Activity ***See Appendix***

1. Define relative frequency probability.

2. Define subjective probability.

3. Explain the Law of Large Numbers.

4. What is the probability that a thumbtack lands point up?

5. Make a table and line graph illustrating your data.

6. As the number of tosses increased, what happened to the relative frequency of the thumbtacks landing up?

7. From the data collected, what is the probability that a thumbtack will land point up?

8. How many tosses do you feel is needed to get an accurate probability for a thumbtack landing point up?

9. How was the Law of Large Numbers used in this activity?

Sample Solution:
See Appendix

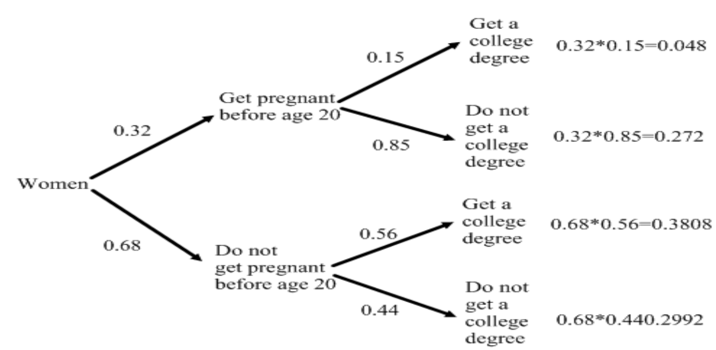
Assessment's Alignment		
CONTENT	MA3	Data analysis
PROCESS	3.3	Apply one's own strategies
DOK	2	Skill/concept
LEVEL OF EXPECTATION	Mastery level – 90%	

Assessment #2:

According to a recent study, 32% of American women will become pregnant before they turn 20. Of those that become pregnant, 85% will not receive a college degree, while 56% of women who don't become pregnant before they turn 20 will get a college degree. Let A be an American woman getting pregnant before the age of 20 and B be a woman receiving a college degree. Use this information for the following problems.

- Create a tree diagram representing the information given.

Solution



- Find the probability that a randomly selected American woman gets pregnant before the age of 20 and gets a college degree.

Solution: $0.32 * 0.15 = 0.048$

- What is the probability that a randomly chosen American woman receives a college degree?

Solution: $0.32 * 0.15 + 0.68 * 0.56 = 0.4288$

- What is the probability that a woman got pregnant before the age of 20, given that she received a college degree?

Solution: $\frac{0.048}{0.4288} = 0.1119$

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.10 Apply information, ideas and skills
DOK	2 Skill/concept
INSTRUCTIONAL STRATEGIES	Homework and practice

Assessment's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.10 Apply information, ideas and skills
DOK	3 Strategic Thinking
LEVEL OF EXPECTATION	Mastery level – 85%

Student Resources	Teacher Resources
The Practice of Statistics	The Practice of Statistics; Yates, Moore, Starnes; Freeman; ©2003; ISBN # 0-7167-4773-1

Identity Equity and Readiness			
Gender Equity		Technology Skills	
Racial/Ethnic Equity		Research/Information	
Disability Equity		Workplace/Job Prep	

Content Area: Mathematics	Course: AP Statistics	Strand: Data and Probability 7
Learner Objectives: Students will anticipate patterns by exploring random phenomena using probability and simulation.		

Concepts: B: Combine independent random variables

Students Should Know	Students Should Be Able to
• Notion of independence versus dependence	• Calculate mean and standard deviation for sums and differences of independent random variables (IIIB2)

Instructional Support

Student Essential Vocabulary					
Data	Sample	Statistic(s)	Variability	Random	Random Selection
Transformations	Mean	68-95-99.7 Rule	Degrees of Freedom	Spread	Standard Deviation
Variance	Model	Central Limit Theorem	Complement	Trial	Sample Space
Conditions and Assumptions		Discrete Random Variable		Law of Large Numbers	Probability
Assumptions and Conditions		Normal Approximation to the Binomial		Probability Distribution	Probability Model
Continuous Random Variable		Transforming Data	Independence	Addition Rule	Event
Multiplication Rule	Outcome	Random Variable			

Sample Learning Activities	Sample Assessments
Learning Activity #1 : Dettmer Dog ***See Appendix*** Students will explore combinations of independent random variables by filling in a data table and then discovering the formulas for handling the transformations of random variables. Solution: ***See Appendix***	**Assessment #1:** A certain type of cereal has a mean caloric content of 110 with a standard deviation of 8 for one serving. Whole milk has a mean caloric content of 140 with a standard deviation of 6 for one serving. ➢ If your breakfast consists of one serving of cereal with a serving of whole milk, what is the mean number of calories you would consume for breakfast? Solution: $\mu_{milk+cereal} = \mu_{milk} + \mu_{cereal} = 110 + 140 = 250cal.$ ➢ What would be the standard deviation for the number of calories in your breakfast? Solution: Assuming independence, $\sigma_{milk+cereal} = \sqrt{\sigma^2_{milk} + \sigma^2_{cereal}} = \sqrt{8^2 + 6^2} = 10cal.$

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.6 Discover/evaluate relationships

Assessment's Alignment	
CONTENT	MA3 Data analysis

DOK	2 Skill/concept
INSTRUCTIONAL STRATEGIES	Cues, questions, and advance organizers

Learning Activity #2:

A car dealer finds that the probability of selling 0, 1, 2, 3, 4, or 5 hybrid cars in any week is .32, .28, .15, .11, .08, and .06 respectfully. Construct a table to represent the probability distribution for the random variable X, the number of hybrid cars sold each day.

What is the mean or expected value for the random variable X?
What is the variance for the random variable X?
What is the standard deviation for the random variable X?
What is the probability that the dealer sells at least 2 hybrid cars in any week?
According to the information provided, what is the value of P(X<4)?

Solution:

x	0	1	2	3	4	5
P(x)	.32	.28	.15	.11	.08	.06

What is the mean or expected value for the random variable X?

$\mu_x = 1.53$

What is the variance for the random variable X?

$\sigma_x^2 = (1.5196)^2 = 2.3091$

PROCESS	1.6 Discover/evaluate relationships
DOK	2 Skill/concept
LEVEL OF EXPECTATION	Mastery level –

Assessment #2:

A grain elevator and shipper has a fleet of tractor trailer trucks that carry corn that are normally distributed with a mean of 46,000 lbs and a standard deviation of 2800 lbs.

One of their customers cannot take full shipments daily because of storage concerns. They take shipments that are normally distributed with a mean of 30,000 lbs and a standard deviation of 4200.

- If the corn left over on the truck is then returned to the grain elevator, what is the mean amount of corn returned to the elevator?

Solution: $\mu_{truck-delivery} = \mu_{truck} - \mu_{delivery} = 46000 - 30000 = 16000 lbs.$

- What is the standard deviation of the corn returned to the elevator above?

What is the standard deviation for the random variable X?

$\sigma_x = 1.5196$

What is the probability that the dealer sells at least 2 hybrid cars in any week?
$P(x \geq 2) = .15 + .11 + .08 + .06 = .4$

According to the information provided, what is the value of $P(X<4)$?
$P(X<4) = .86$

See Appendix for more activities and solutions

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	3.3 Apply one's own strategies
DOK	2 Skill/concept
INSTRUCTIONAL STRATEGIES	Nonlinguistic representation

Solution**:** Assuming independence,

$$\sigma_{truck-delivery} = \sqrt{\sigma^2_{truck} + \sigma^2_{delivery}} = \sqrt{2800^2 + 4200^2} \approx 5047.77 lbs.$$

Assessment's Alignment	
CONTENT	MA3 Data analysis
PROCESS	3.3 Apply one's own strategies
DOK	2 Skill/concept
LEVEL OF EXPECTATION	Mastery level – 80%

Student Resources	Teacher Resources
The Practice of Statistics	The Practice of Statistics; Yates, Moore, Starnes; Freeman; ©2003; ISBN # 0-7167-4773-1

Identity Equity and Readiness

Gender Equity		Technology Skills	
Racial/Ethnic Equity		Research/Information	
Disability Equity		Workplace/Job Prep	

Content Area: Mathematics	**Course: AP Statistics**	**Strand: Data and Probability 8**
Learner Objectives: Students will anticipate patterns by exploring random phenomena using probability and simulation.		

Concepts: C: Interpret the normal distribution

Students Should Know	Students Should Be Able to
● Properties of the normal distribution	● Use tables of normal distribution (IIIC2) ● Use the normal distribution as a model for measurements (IIIC3) ● Calculate probabilities and cut-off points in normal distributions

Instructional Support

Student Essential Vocabulary					
Data	Sample	Statistic(s)	Variability	Random	Random Selection
Transformations	Mean	68-95-99.7 Rule	Degrees of Freedom	Standard Deviation	Variance
Central Limit Theorem	Spread	Discrete Random Variable	Complement	Law of Large Numbers	Probability
Assumptions and Conditions	Model	Continuous Random Variable		Probability Distribution	Probability Model
Conditions and Assumptions		Transforming Data	Parameter	Population	Extrapolation
Statistically Significant	Dotplot	Proportion	Distribution	Histogram	Normal
Outliers	Shape	Skew	Stem-and-Leaf	Symmetric	Z-Score
Quantitative Variable	Percentile	Standardized Value	Quartile	Standard Normal Model	
Normal Probability Plot	Tails				

Sample Learning Activities	Sample Assessments
Learning Activity #1 : Men's heights are normally distributed with a mean of 69 in. and a standard deviation of 2.8 in. Early missions for NASA required their astronauts to be between 65 in and 68 inches tall. What proportion of men would meet the height requirement?	**Assessment #1:** A grain elevator and shipper has a fleet of tractor trailer trucks that carry corn that are normally distributed with a mean of 46,000 lbs and a standard deviation of 2800 lbs. ➢ If you must purchase special permits for any load over 50,000 lbs, what is the proportion of trucks that would require such permits?

Solution:

$$P(65 < x < 68) = P\left(\frac{65-69}{2.8} < z < \frac{68-69}{2.8}\right) = P(-1.43 < z < -.36$$

By calculator: *normalcdf(65,68,69,2.8)* = 0.2839
or *normalcdf(-1.43,-.36,0,1)* = 0.2839

Activity's Alignment		
CONTENT	MA3	Data analysis
PROCESS	1.10	Apply information, ideas and skills
DOK	2	Skill/concept

Solution:

$$P(X > 50000) = P\left(z > \frac{50000-46000}{2800}\right) = 0.0766$$

One of their customers cannot take full shipments daily because of storage concerns. They take shipments that are normally distributed with a mean of 30,000 lbs and a standard deviation of 4200. A clause in the contract with the customer described above states that an unusual order for corn is the top 1% of their orders. Any shipment that is **not** unusual and is **not** filled by the elevator is considered a breach of contract and allows the customer to void the contract.

- What is the probability of them being able to void the contract based on any one shipment?

Solution: The z-score for the top 1% would be $\approx$ 2.326 from either the calculator or a chart. Therefore,

$$2.326 < \frac{X-30000}{4200}$$
$$X < 39769.2lbs.$$

Therefore an order of less than approximately 39769.2 lbs. that is not met would be grounds for voiding the contract. So

$$P(X_{delivery} < 39769.2) = P\left(Z < \frac{39769.2-46000}{2800}\right) \approx 0.0130$$

Assessment's Alignment		
CONTENT	MA3	Data analysis
PROCESS	1.10	Apply information, ideas and skills

INSTRUCTIONAL STRATEGIES	Homework and practice

Learning Activity #2:

ACT scores are normally distributed with a mean of 20.8 and a standard deviation of 5.6. A certain elite university wishes to consider only the top 2% of students based on ACT scores.

What score must you achieve in order to be considered at this university?

Solution: z-score for the top 2% is 2.054. Therefore:

$$2.054 = \frac{x - 20.8}{5.6} \Rightarrow x = 20.8 + 2.054(5.6) = 32.3024$$

This indicates you would need to score at least 33 on the ACT to be considered by this university.

By calculator: *invnorm(.98,20.8,5.6)* ≈ 32.3009

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.10 Apply information, ideas and skills

DOK	3 Strategic Thinking
LEVEL OF EXPECTATION	Mastery level – 75%

Assessment #2:

ACT scores are normally distributed with a mean of 20.8 and a standard deviation of 5.6.

- What is the probability of someone randomly selected getting at least a 24 on the ACT?

Solution: $$P(x > 24) = P\left(\frac{24 - 20.8}{5.6}\right) = 0.2839$$

- If a university were to accept only the top 1% of its applicants based upon their ACT scores, what score would be required in order to be considered for admission into this university?

Solution: The z-score for the top 1% is 2.326.Therefore

$$\frac{x - 20.8}{5.6} = 2.326$$

so $x = 2.326 * 5.6 + 20.8 = 33.8256$

You would need to get a 34 or higher to be considered at this university for admission based on your ACT score.

Assessment's Alignment	
CONTENT	MA3 Data analysis

DOK	2 Skill/concept
INSTRUCTIONAL STRATEGIES	Identifying similarities and differences

PROCESS	1.10 Apply information, ideas and skills
DOK	2 Skill/concept
LEVEL OF EXPECTATION	Mastery level – 80%

Student Resources	Teacher Resources
The Practice of Statistics	The Practice of Statistics; Yates, Moore, Starnes; Freeman; ©2003; ISBN # 0-7167-4773-1

Identity Equity and Readiness			
Gender Equity		Technology Skills	
Racial/Ethnic Equity		Research/Information	
Disability Equity		Workplace/Job Prep	

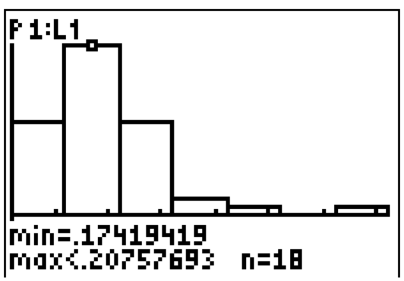
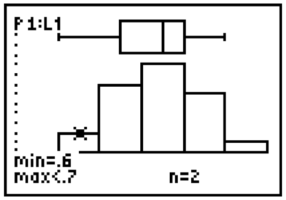
Content Area: Mathematics	Course: AP Statistics	Strand: Data and Probability 9
Learner Objectives: Students will anticipate patterns by exploring random phenomena using probability and simulation.		

Concepts: D: Interpret sampling distributions

Students Should Know	Students Should Be Able to
● Central Limit Theorem ● Properties of t-distribution ● Properties of Chi-square distribution	● Describe sampling distribution of a sample proportion (IIID1) ● Describe sampling distribution of a sample mean (IIID2) ● Apply Central Limit Theorem (IIID3) ● Describe sampling distribution of a difference between two independent sample proportions (IIID4) ● Describe sampling distribution of a difference between two independent sample means (IIID5) ● Perform a simulation of sampling distributions (IIID6) ● Use tables of t-distribution and Chi-square (IIID7, IIID8) ● Calculate probabilities and cut-offs points in t-distributions and Chi-square distributions (IIID7, IIID8)

Instructional Support

Student Essential Vocabulary					
Variablity	Data	Transformations	68-95-99.7 Rule	Degrees of Freedom	Standard Deviation
Random Selection	Sample	Variance	Model	Central Limit Theorem	Spread
Assumptions and Conditions	Statistic(s)	Complement	Mean	Discrete Random Variable	Probability
Conditions and Assumptions		Continuous Random Variable		Law of Large Numbers	Probability Distribution
Normal Approximation to the Binomial		Probability Model	Independence	Transforming Data	Parameter
Statistically Significant	Population	Extrapolation	Proportion	Distribution	Dotplot
Histogram	Normal	Outliers	Shape	Skew	Stem-and-Leaf
Symmetric	Z Score	Quantitative Variable	Percentile	Standardized Value	Quartile
Standard Normal Model	Tails	Simulation	Bias	Chi-square Distribution	Component Values
Goodness of Fit	Cell	Interval	Pooling	Sample Mean Model	Sampling Distribution
Sample Proportion Model	Test	Sampling Variability	T Distribution	Test of Homogeneity	Test of Independence

Sample Learning Activities	Sample Assessments
Learning Activity #1 : From a large bin of multi-colored beads, there are 18% that are red. Students are asked to scoop out a cup-full of beads and calculate out the proportion that is red. While the individual proportions were all different, the distribution for my two classes is shown below. The distribution is at least somewhat "normal" (a bit skewed right with an outlier to the right). The "peak" (or center) does appear around 18% however. We also talked about how the outlier may have occurred by not properly shaking the bin after a sample was taken and therefore the reds may have occurred at a higher rate since they may not have been randomly distributed throughout the bin.  ***See Appendix***	**Assessment #1:** Suppose that you and your lab partner flip a coin 20 times and you calculate the proportion of tails to be 0.7. Your partner seems surprised at these results and suspects that the coin is not fair. ➢ Write a brief statement that describes why you either agree or disagree with him. Solution: Answers may vary. I would disagree with him. Because of sampling variability, 0.7 is definitely possible. Using the binomial probability, the chances of getting 0.7 tails is 0.0370 ➢ This problem extends the previous problem. You flip the coin 20 more times and calculate the proportion of tails. You repea the process again and again until you have calculated 25 proportions $\hat{p}$. Your lab partner then plots a histogram of the results on his TI-83 and overlays a box plot for the same data shown. What would these results suggest? Explain clearly. 

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.10 Apply information, ideas and skills
DOK	2 Skill/concepts
INSTRUCTIONAL STRATEGIES	Nonlinguistic representation

Learning Activity #2:

German Tanks

Follow the script for the lesson:

Draw 7 numbers from the bag and return them when completed. Decide an estimate of N, the number of "tanks" in the population, and also a clear statement of the method you used. The method should be applicable to any set of seven randomly selected numbers.

Groups that work quickly should be encouraged to come up with a second or even third method and estimate.

Write estimate on the board, along with a short formula describing how they arrived at the estimate. If different groups come up with the same method, list the method only once with the different estimates it produces listed beside it.

Say, "The true value of N is 342. So which of these methods is best?"

Answer: the method that happened to produce an estimate closest to 342, but some students are likely to pick up on the fact that the method may have just

Solution**:** Because the process is repeated many times, the distribution of samples indicates that the "mean" proportion of tails is more likely to be 0.8 or higher. Therefore, it does appear that the coin is unfair.

Assessment's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.10 Apply information, ideas and skills
DOK	2 Skill/concepts
LEVEL OF EXPECTATION	Mastery level – 80%

Assessment #2:
A survey asks a random sample of 1500 adults in Ohio if they support an increase in the state sales tax from 5% to 6%, with the additional revenue going to education. Let *p* denote the proportion in the sample that says they support the increase. Suppose that 40% of all adults in Ohio support the increase.

- If $\hat{p}$ is the proportion of the sample who support the increase what is the mean of $\hat{p}$?

Solution**:** $\mu_{\hat{p}} = p = 0.4$

- What is the standard deviation of $\hat{p}$?

Solution**:** $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.4*0.6}{1500}} \approx 0.0126$

gotten "lucky" with a particular sample. In fact, the question above contains an undefined word that should be discussed with the class: best. What do we mean by the "best" method? The one that's exactly right most often? The one that's within 50 tanks of being right most often?

(Many statisticians think of the "best" estimator as the one that, among all unbiased methods, has the smallest standard deviation.)

Lead a discussion to get at these points:
• You can't judge a method ("estimator") by how it performs on one random sample only. You have to judge it by how it performs over many random samples. In other words, you have to look at the distribution of its estimates over many random samples. Simulation (especially with technology) is useful for looking at that distribution.
• It is desirable that an estimator be unbiased. That is, you would like for the mean of its estimates over many random samples to equal the parameter (N) being estimated.
• It is desirable that an estimator have low variability, perhaps measured by standard deviation. That is, you would like the estimates it produces over many random samples to be relatively close to one another.
• The combination of unbiasedness and low variability makes an estimator that comes close to the desired target for a large number of possible random samples.

Remember that the ultimate goal of the estimator is to estimate a parameter given a single random sample. Since they only get one shot, they want their method to be one that works well for many possible random samples, thus minimizing their risk of being far from correct. The ability of an estimator to work well for many different random samples is sometimes referred to as its robustness.

Students will likely want to know what the British mathematicians did. They used the statistic that has the minimal variance among all unbiased estimators, which for this activity would be (8/7) max, where max is the largest value in the sample. For this one, point out that distribution need not

- Explain why you can use the formula for the standard deviation of $\hat{p}$ in this setting.

Solution**:** The population of Ohio is more than 15000, or ten times our sample size.

- Check that you can use the normal approximation for the distribution of $\hat{p}$.

Solution:
$$n*\hat{p} = 1500*0.4 = 600 \geq 10$$
$$and$$
$$n*\hat{q} = 1500*0.6 = 900 \geq 10$$

- Find the probability that $\hat{p}$ takes a value between 0.37 and 0.43.

Solution:

$$P\left(0.37 < \hat{p} < .43\right) = P\left(\frac{0.37-.4}{\sqrt{\frac{0.4*0.6}{1500}}} < z < \frac{0.43-.4}{\sqrt{\frac{0.4*0.6}{1500}}}\right) = 0.9823$$

be normal or even symmetric: all that is desirable is that they come close to the target for many samples.

Students trade lists, draw new samples, or use randInt(1,342), and try their method again on a new list to see if their method estimates well with a different sample.

Run the Fathom Simulation based on the formulas presented by the class to show the distribution of each described statistic over many simulations. Each simulation represents the computation of the statistic when a sample of 7 observations is taken from the integers 1 to 342. The vertical line in each case is placed at 342—the true value that the statistic is shooting for.

Discuss the SOCS for each distribution.

What makes an estimator "unbiased"? Compare distributions for each estimator. Which one would the Allied forces prefer? Why?

Sample Solutions:
See Appendix

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.6 Discover/evaluate relationships
DOK	4 Extended thinking
INSTRUCTIONAL STRATEGIES	Nonlinguistic representations

Assessment's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.6 Discover/evaluate relationships
DOK	3 Strategic thinking
LEVEL OF EXPECTATION	Mastery level – 75%

Student Resources	Teacher Resources
The Practice of Statistics	The Practice of Statistics; Yates, Moore, Starnes; Freeman; ©2003; ISBN # 0-7167-4773-1

Identity Equity and Readiness			
Gender Equity		Technology Skills	
Racial/Ethnic Equity		Research/Information	
Disability Equity		Workplace/Job Prep	

Content Area: Mathematics	**Course: AP Statistics**	**Strand: Data and Probability 10**
Learner Objectives: Students will make statistical inferences by estimating population parameters and testing hypotheses.		

Concepts: A: Determine estimation

Students Should Know	Students Should Be Able to
• Use of properties of point estimators, including unbiasedness and variability • Logic of confidence intervals, meaning of confidence level and confidence intervals, and properties of confidence intervals • Relationship between confidence interval and two-sided alternative	• Estimate population parameters using margins of error (IVA1) • Construct and interpret large sample confidence interval for a proportion (IVA4) • Construct and interpret large sample confidence interval for a difference between two proportions (IVA5) • Construct and interpret confidence interval for a mean (IVA6) • Construct and interpret confidence interval for a difference between two means (unpaired and paired) (IVA7)

	• Construct and interpret confidence interval for the slope of a least-squares regression line (IVA8)

Instructional Support

Student Essential Vocabulary					
Data	Sample	Statistic(s)	Variability	Random	Random Selection
Transformations	68-95-99.7 Rule	Degrees of Freedom	Mean	Spread	Standard Deviation
Variance	Model	Central Limit Theorem	Assumptions and Conditions		Complement
Conditions and Assumptions		Continuous Random Variable		Discrete Random Variable	
Law of Large Numbers		Normal Approximation to the Binomial		Probability	Probability Distribution
Probability Model	Transforming Data	Independence	Parameter	Population	Statistically Significant
Extrapolation	Proportion	Distribution	Dotplot	Histogram	Normal
Outlier(s)	Shape	Skew	Stem-and-Leaf	Symmetric	Quantitative Variable
Standardized Value	Z Score	Standard Normal Model		Tails	Chi-square Distribution
Interval	Pooling	Sample Mean Model	Sample Proportion Model		Sampling Distribution
Sampling Distribution Models		Sampling Variability	T Distribution	Test	Causation
Experiment	Matched Pairs	Conditional Probability	Confidence Level	Boxplot	Alpha
Confidence Interval	Critical Value	Hypothesis	Margin of Error	Null	Reject
Retain	Significant Level	Standard Error	Two-sided	1 Proportion Z-interval	2 Proportion Z-interval
5 Number Summary	One Sample T-interval		One Sample Z-interval		Paired T-interval
Two Sample T-interval		*Normal Probability Plot*			

Sample Learning Activities	Sample Assessments
Learning Activity #1 : A researcher wants to estimate the mean weight of children of a particular age. Assume the distributions of weights of children at a specific age are roughly normal with a standard deviation of 3 pounds. The researcher selects a SRS of 25 children and finds their mean weight to be 62 pounds. What is this researcher's 90% confidence level estimate?	**Assessment #1:** National Fuelsaver Corporation manufactures the Platinum Gasaver, a device they claim "may increase gas mileage by 22%." Here are the percent changes in gas mileage for 15 identical vehicles, as presented in one of the company's advertisements: 48.3 46.9 46.8 44.6 40.2 38.5 34.6 33.7 28.7 28.7 24.8 10.8 10.4 6.9 -12.4

Solution:

P Population
μ = Mean weight of children of this age

A There are surely more than 10 (25) = 250 children of some age
A SRS is stated
We are told σ = 3
Since population weight are ≈ normal, we can use Z scores/normal approximations

N Use a Z interval $\bar{x} \pm z^* \bullet \frac{\sigma}{\sqrt{n}}$

I Z interval

$62 \pm 1.645 \bullet \frac{3}{\sqrt{25}}$

x = 62
n = 25

C I'm 90% confident that the mean weight of children of this age is between 61.013 lbs and 62.987 lbs

- Construct a 90% confidence interval to estimate the mean fuel savings in the population of all such vehicles. Follow the Inference Toolbox.

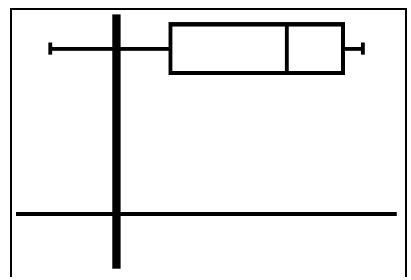
Solution: Let be the true population mean percent change in gas mileage.

Conditions:

1. SRS (assumed although probably should be questioned given that it is an advertisement for the company in question)

2. It is reasonable to assume that more than 150 cars will use the device.

3. The distribution seems only slightly skewed left, so the sample size of 15 should be sufficient to overcome the slight skew.



$\bar{x}$ =28.77&s = 17.766

df=14

Therefore $\bar{x} \pm t_{90,14} \frac{s}{\sqrt{n}} = 28.77 \pm 1.753 * \frac{17.766}{\sqrt{15}} = (20.687, 36.846)$

Conclusion: If the data can be trusted, then we are 90% confident the true mean fuel savings for all vehicles would be somewhere between 20.687% and 36.846%.

Activity's Alignment		
CONTENT	MA3	Data analysis
PROCESS	1.2	Conduct research
DOK	3	Strategic thinking
INSTRUCTIONAL STRATEGIES	Generating and testing hypotheses	

Learning Activity #2:

Citing the problem of childhood obesity, a researcher believes the mean weight of children of a particular age has changed. The long held belief is that children this age have a mean weight of 58 pounds. Assume the distributions of weights of children at a specific age are roughly normal with a standard deviation of 3 pounds. The researcher selects a SRS of 25 children and finds their mean weight to be 62 pounds. Is there evidence that the mean weight has changed?

P Population: children of some specific age
μ = mean weight of children that age

H Null $H_0 : \mu = 58$

Alternative $H_a : \mu \neq 58$

A There are over 10(25) = 250 children of any given age
The researcher took a SRS

- Explain what "90% confidence" means in this setting.

Solution: 90% of all such confidence intervals produced in the same manner would capture the true mean.

Assessment's Alignment		
CONTENT	MA3	Data analysis
PROCESS	3.5	Reason logically
DOK	3	Strategic thinking
LEVEL OF EXPECTATION	Mastery level – 75%	

Assessment #2:

A study of chromosome abnormalities and criminality examined data from 4,124 males born in Copenhagen. Each man was classified as having a criminal record or not, using the registers maintained in the local police offices. Each was also classified as having the normal male XY chromosome pair or one of the abnormalities XYY or XXY. Of the 4,096 men with normal chromosomes 381 had criminal records, while 8 of the 28 men with abnormal chromosomes had criminal records. Some experts believe chromosome abnormalities are associated with increased criminality.

- Construct and interpret a 95% confidence interval for the difference in proportions.

Solution: Let p_1 be the true population proportion of men with normal chromosomes who have criminal records and p_2 be the true population proportion of men with abnormal chromosomes who have criminal records. Therefore $p_1 - p_2$ will represent the true difference in population proportions of

	We are told σ = 3, and that these weights are approximately normally distributed
N	Z test
	$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
T	$z = \frac{62 - 58}{\frac{2}{\sqrt{25}}} = 6.67$
O	P value P(z = 6.67 or greater assuming H_0 is true) = 2.6 x 10^{-11}
M	Because this P value is so small (<σ =0.05) I'll reject H_0 .
S	There is significant evidence that the mean weight of children this age has changed

Activity's Alignment		
CONTENT	MA3	Data analysis
PROCESS	1.2	Conduct research
DOK	3	Strategic thinking
INSTRUCTIONAL STRATEGIES	Generating and testing hypotheses	

those with criminal records between those with normal and abnormal chromosomes.

Conditions:

1. SRS (assumed)
2. The population of men in Denmark > 40960
3. $n_1 * \hat{p}_1 = 4096 * 0.093 = 381 \geq 5$
   $n_1 * \hat{q}_1 = 4096 * 0.907 = 3715 \geq 5$
   $n_2 * \hat{p}_2 = 28 * 0.286 = 8 \geq 5$
   $n_2 * \hat{q}_2 = 28 * 0.714 = 20 \geq 5$

$$(0.093 - 0.286) \pm 1.960 * \sqrt{\frac{(0.093)(0.917)}{4096} + \frac{(0.286)(0.714)}{28}} = (-0.3603, -0.0251)$$

Conclusion: We are 95% confident that the true difference in population proportions of those with criminal records between those with normal and abnormal chromosomes are from -0.3603 & -0.0251. This means that men with normal chromosomes are about 2.5% to 36% LESS LIKELY to have criminal records than those with abnormal chromosomes. Since 0 is NOT in the interval, that suggests that there is a significant difference in the proportion of men with criminal records with respect to normal and abnormal chromosomes

Assessment's Alignment		
CONTENT	MA3	Data analysis
PROCESS	3.5	Reason logically
DOK	3	Strategic thinking

	LEVEL OF EXPECTATION: Mastery level – 75%

Student Resources	Teacher Resources
The Practice of Statistics	The Practice of Statistics; Yates, Moore, Starnes; Freeman; ©2003; ISBN # 0-7167-4773-1

Identity Equity and Readiness			
Gender Equity		Technology Skills	
Racial/Ethnic Equity		Research/Information	
Disability Equity		Workplace/Job Prep	

Content Area: Mathematics	Course: AP Statistics	Strand: Data and Probability 11
Learner Objectives: Students will make statistical inferences by estimating population parameters and testing hypotheses.		

Concepts: B: Determine tests of significance

Students Should Know	Students Should Be Able to
• Logic of significance testing, null and alternative hypotheses • Relationship between p-value and test statistic • Difference between one- and two-sided tests • Relationship among Type I and Type II errors and power • Relationship between confidence interval and two-sided alternative	• Perform and interpret large sample test for a proportion (IVB2) • Perform and interpret large sample test for a difference between two proportions (IVB3) • Perform and interpret test for a mean (IVB4) • Perform and interpret test for a difference between two means (unpaired and paired) (IVB5) • Perform and interpret Chi-square test for goodness of fit, homogeneity of proportions, and independence (one- and two-way tables) (IVB6) • Perform and interpret test for the slope of a least-squares regression line (IVB7)

Instructional Support

Student Essential Vocabulary					
Data	Sample	Statistic(s)	Variability	Random	Random Selection
Transformations	68-95-99.7 Rule	Degrees of Freedom	Mean	Spread	Standard Deviation
Variance	Model	Central Limit Theorem	Assumptions and Conditions		Complement

Conditions and Assumptions		Continuous Random Variable		Discrete Random Variable	
Law of Large Numbers		Normal Approximation to the Binomial		Probability	Probability Distribution
Probability Model	Transforming Data	Independence	Parameter	Population	Statistically Significant
Extrapolation	Proportion	Distribution	Dotplot	Histogram	Normal
Outlier(s)	Shape	Skew	Stem-and-Leaf	Symmetric	Quantitative Variable
Standardized Value	Z Score	Standard Normal Model		Tails	Chi-square Distribution
Interval	Pooling	Sample Mean Model	Sample Proportion Model		Sampling Distribution
Sampling Distribution Models		Sampling Variability	T Distribution	Test	Causation
Experiment	Matched Pairs	Conditional Probability	Confidence Level	Boxplot	Alpha
Confidence Interval	Critical Value	Hypothesis	Margin of Error	Null	Reject
Retain	Significant Level	Standard Error	Two-sided	Cell	Component Values
Goodness of Fit	Test of Homogeneity		Test of Independence		Association
Explanatory	Lurking Variables	Prediction	Response	Response Variable	Expected Value
Coefficient of Determination (R^2)		Correlation Coefficient (r)		Direction	Exponential Model
Form	Intercept	Least Squares Regression (LSRL)		Line of Best Fit	Linear Model
Power Model	Predicted Value	Regression	Residual	Regression Outliers	Residual Plot
Scatterplot	Slope (Rate of Change)		Strength	Categorical Variable	Conditional Distribution
Contingency Table	Frequency Table	1 Proportion Z-test	Alternative	2 Proportion Z-test	One Sample T-test
One Sample Z-test	P Value	Paired T-test	Power	Residual Standard Deviation	
Standard Error of the Slope		Test Statistic	T-test for Regression Slope		Two Sample T-test
Type I Error	Type II Error	*Normal Probability Plot*		*Ladder of Powers*	*Logarithmic Model*
Monotonicity	*Subset*				

Sample Learning Activities	Sample Assessments
Learning Activity #1 : Study of Iron Deficiency A study of iron deficiency among infants compared samples of infants following different feeding regimens. One group contained breast-fed infants, while the children in another group were fed a standard baby formula without any iron supplements. The results on blood hemoglobin levels at 12 months of age showed that 37 breast-fed children had a mean blood hemoglobin level $\bar{x}$ = 13.3 with s = 1.7. The 41 formula-fed children had a mean blood hemoglobin level $\bar{x} = 1\square.\square$ with s = 1.8.	**Assessment #1:** Mars Inc., makers of M&M candies, claims that they produce plain M&Ms with the following distribution: Brown: 30% Red: 20% Yellow: 20% Orange: 10% Green: 10% Blue: 10% A bag of plain M&Ms was selected randomly from the grocery store shelf, and the color counts were as follows: Brown: 16 Red: 11 Yellow: 19 Orange: 5 Green: 7 Blue: 3 ➢ You want to conduct an appropriate test of a consumer's claim that the proportion of yellow M&Ms is lower than Mars candies claims. Perform an appropriate significance test.

Where your ID number is □□□□□□. Is there significant statistical evidence that the mean hemoglobin level is higher among breast-fed babies? Show all "PHANTOMS" steps. Solution: P Population – infants Parameter: M^b = mean blood hemoglobin level of breast-fed children M^f = mean blood hemoglobin level of formula-fed children H Hypotheses Ho: $M^b = M^f$ Ha: $M^b > M^f$ A Assumptions and Conditions We need 2 independent SRSs We assume that the sampling distributions are normal, even though it is not specifically states There are more than 10(37) = 370 and 10(41) = 410 breast-fed and formula-fed infants, respectively σ^b and σ^f are unknown N Name of and formula for test 2 sample T test	Solution: Let *p* be the true population proportion of yellow M&Ms. $H_0: p = 0.30$ $H_a: p < 0.30$ Conditions: 1. SRS selection (assumed) 2. There are more than (10)(61) = 610 M&Ms 3. 61*0.30 or 18.3 ≥ 10 61*.0.70 or 42.7 ≥ 10 $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.311 - .3}{\sqrt{\frac{(0.3)(0.7)}{61}}} = 0.196$ $P(z < 0.196) = 0.578$ Conclusion: If there is indeed no change in the proportion of yellow M&Ms being 30%, then results like ours would occur approximately 57.8% of the time by chance. Therefore, we conclude there is virtually no evidence to reject that there is no change in the yellow proportion of M&Ms leaving open the possibility that it could be 30%. Our evidence in fact shows that the proportion is higher than 30% so we should not conclude that there has been a decrease in yellow M&Ms.

$$t = \frac{\left(\bar{x}_b - \bar{x}_f\right) - \left(\mu_b - \mu_f\right)}{\sqrt{\frac{s_b^2}{n_b} + \frac{s_f^2}{n_f}}}$$

T Test statistic or CI calculated

$$t = \frac{(13.3 - 11.5) - (0)}{\sqrt{\frac{1.7^2}{37} + \frac{1.8^2}{41}}} = 4.541$$

O Obtain a p value

P = 1.036E – S

M Make a decision

Since p = 1.036E – S < 0.05, I'll reject Ho

S State a conclusion

There is evidence that the mean blood hemoglobin level of breast-fed children is higher than that of formula-fed children.

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.2 Conduct research
DOK	3 Strategic thinking

Assessment's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.10 Apply information, ideas and skills
DOK	3 Strategic thinking
LEVEL OF EXPECTATION	Mastery level – 75%

Assessment #2:

In a study of heart surgery, one issue was the effect of drugs called beta-blockers on the pulse rate of patients during surgery. The available

INSTRUCTIONAL STRATEGIES	Generating and testing hypotheses

Learning Activity #2:

Record your distribution of colors below for your peanut M&Ms

Red	Orange	Yellow	Green	Blue	Brown

Give the class totals from the plain M&M activity before and the class totals for the peanut M&Ms today. Record those below.

	Red	Orange	Yellow	Green	Blue	Brown	Totals
Plain							
Peanut							

Mars Inc. claims the distribution of peanut M&Ms is 12% Red, 23% Orange, 15% Yellow, 15% Green, 23% Blue, and 12% Brown.

1. Perform a $\chi 2$ Goodness of Fit test on the class data to test their claim.
2. Why would you not conduct this test on your individual bags?
3. Perform an appropriate test on the class data to see if the distributions between plain chocolate and peanut M&Ms are different.
4. Is the test for #3 above a $\chi 2$ Test for Homogeneity or a $\chi 2$ Test for Independence (no association)?

Sample solution:

subjects were divided at random into two groups of 30 patients each. One group received a beta-blocker; the other group received a placebo. The pulse rate of each patient at a critical point during the operation was recorded. The treatment group had mean 65.2 and standard deviation 7.8. For the control group, the mean was 70.3 and the standard deviation was 8.3.

- Perform an appropriate significance test to see if beta-blockers reduce the pulse rate. Follow the Inference Toolbox.

Solution: Let μ_1 be the true population mean pulse rate of those on beta blockers in the treatment group and μ_2 be the true population mean pulse rate of those who are not on beta blockers, so $\mu_1 - \mu_2$ will be the difference in the pulse rates of the two groups.

$$H_0 : \mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 < \mu_2 \text{ or } \mu_1 - \mu_2 < 0$$

Conditions:

1. SRS selection (assumed)
2. Potential heart surgery patients at least 300.
3. Although the distribution is unknown, 30 patients per group is sufficiently large to overcome any potential skew or outliers.

$$t = \frac{\left(\overline{x_1} - \overline{x_2}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{\left(\overline{x_1} - \overline{x_2}\right)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(65.2 - 70.3)}{\sqrt{\frac{7.8^2}{30} + \frac{8.3^2}{30}}} = -2.453$$

	Red	Orange	Yellow	Green	Blue	Brown	Totals
Plain	76	104	90	89	109	61	529
Peanut	130	301	223	263	329	174	1420

1. Since the distributions for peanut M&Ms is 12% Red, 23% Orange, 15% Yellow, 15% Green, 23% Blue, and 12% Brown, the expected values for the peanut M&Ms listed above in line one are 63.48 each for Red and Brown, 121.67 each for Orange and Blue, and 79.35 each for Yellow and Green.

H_0 : All proportions are the same as reported by Mars.

H_a : At least one proportion is significantly different.

$$\chi^2 = \sum \frac{(observed - \exp ected)^2}{\exp ected}$$

$$= \frac{(76-63.48)^2}{63.48} + ... + \frac{(61-63.48)^2}{63.48} = 9.0547$$

and

Therefore: $P(\chi^2 > 9.0547) = 0.0597$

Conclusion: If there is no difference from the stated proportions by Mars, then results like ours would occur 5.97% of the time by chance. Therefore, there is some evidence (significant at 10% but not at 5%) to reject Mars claim. There does appear to be at least one color that is significantly different. A look at the χ^2 component values indicates Orange with a component value of 2.5662 is lower than it should be.

3. We would not conduct this test on individual bags because individual bags would fail the conditions necessary to perform this test.

H_0 : There is no difference in the distribution of plain and peanut M&Ms

$P(t < -2.453) = 0.0086$

Conclusion: If there is indeed no difference in heart rates between the two groups (i.e. $\mu_1 = \mu_2$), then results like ours would occur only about 9 times in 1000 by chance. Therefore, we conclude there is significantly strong evidence (significant at =0.01) to reject that there is no difference in favor of the idea that beta-blockers do indeed significantly reduce pulse rates during heart surgery.

H_a: There is at least one value in the distributions between plain and peanut M&Ms that is different.

The observed and expected matrices are given below.

$$O = \begin{bmatrix} 76 & 104 & 90 & 89 & 109 & 61 \\ 130 & 301 & 223 & 263 & 329 & 174 \end{bmatrix}$$

$$E = \begin{bmatrix} 55.913 & 109.926 & 84.955 & 95.540 & 118.883 & 63.784 \\ 150.087 & 295.074 & 228.045 & 256.460 & 319.117 & 171.216 \end{bmatrix}$$

$$\chi^2 = \sum \frac{(observed - \exp ected)^2}{\exp ected}$$

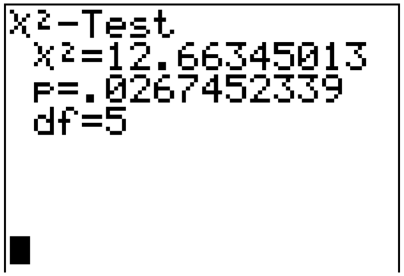
$$= \frac{(76-55.913)^2}{55.913} + \ldots + \frac{(174-171.216)^2}{171.216} = 12.663$$

and

$$P(\chi^2 > 12.663) = 0.0267$$

Conclusion: If there is no difference in the distribution of plain and peanut M&Ms, then our results would occur about 2.67% of the time by random chance. Therefore, there is fairly strong evidence (significant @ 5%, not significant @ 1%) to reject that they have the same distributions. Based upon the χ^2 component value, the red peanut M&Ms appear to be overrepresented if the distributions were the same.

Assessment's Alignment



4. The test between plain and peanut M&Ms is one of Homogeneity.

Activity's Alignment	
CONTENT	MA3 Data analysis
PROCESS	1.2 Conduct research
DOK	3 Strategic thinking
INSTRUCTIONAL STRATEGIES	Generating and testing hypotheses

CONTENT	MA3 Data analysis
PROCESS	1.2 Conduct research
DOK	3 Strategic thinking
LEVEL OF EXPECTATION	Mastery level – 70%

Student Resources	Teacher Resources
The Practice of Statistics	The Practice of Statistics; Yates, Moore, Starnes; Freeman; ©2003; ISBN # 0-7167-4773-1

Identity Equity and Readiness			
Gender Equity		Technology Skills	
Racial/Ethnic Equity		Research/Information	
Disability Equity		Workplace/Job Prep	