Probability & Statistics
Modular Learning Exercises

Teacher Edition

A Curriculum Resource for
Your Accelerated Math Students

Think like an Actuary! Produce real world data from storm statistics.

THE ACTUARIAL FOUNDATION®

Preparing for tomorrow's possibilities® *through education*

About The Actuarial Foundation

The Actuarial Foundation, a 501(c)(3) nonprofit organization, develops, funds and executes education, scholarship and research programs that serve the public and the profession by harnessing the talents and resources of actuaries. Through an array of educational initiatives, the Foundation is *dedicated* to minimizing risk through maximizing education by providing the tools to help improve math and financial knowledge for all. We are proud to add *Probability & Statistics Modular Learning Exercises*, a curriculum resource for your accelerated math students to our library of math resources. Please visit the Foundation's web site at: www.actuarialfoundation.org for additional educational materials.

What is an Actuary? According to BeAnActuary.org, Actuaries are the leading professionals in finding ways to manage risk. It takes a combination of strong math and analytical skills, business knowledge and understanding of human behavior to design and manage programs that control risk. Careercast.com reported that "Actuary" is ranked as the 2nd best job out of 200 of the best and worst jobs. To learn more about the profession, go to: www.BeAnActuary.org.

Some of the activities in this book reference specific Web pages. While active at the time of publication, it is possible that some of these Online Resource links may be renamed or removed by their hosts. Note that these links were provided simply as a convenience; a quick search should reveal some of the many other online resources that can be used to complete these activities. Facts and opinions contained are the sole responsibility of the organizations expressing them and should not be attributed to The Actuarial Foundation and/or its sponsor(s).

Teacher Introduction - Background Information

Purpose

The purpose of these modules is to provide an introduction to the world of probability and statistics to accelerated mathematics students at the high school level. The modules also introduce students to real world math concepts and problems that property and casualty actuaries come across in their work. They are designed to be used by teachers and students of AP courses (primarily AB and BC Calculus) after the AP exam, but of course they can be used at any time.

Statistical thinking is critical in today's society. Data are everywhere and students will see and recognize misinterpretations all the time. For one prominent mathematician's view on why learning statistics is important, please view the following TEDTalks video: http://www.ted.com/talks/arthur_benjamin_s_formula_for_changing_math_education.html.

Project Description

The materials are centered on the fictional town of Happy Shores, a coastal community which is at risk for hurricanes. (For more information, see the Student Background materials). Actuaries at an insurance company figure out the risks and possible losses associated with hurricanes and tropical storms in Happy Shores. Students use historical data to create models about possible damages if hurricanes of certain strengths hit.

This project introduces statistics and probability concepts to students but is not meant to provide enough material or practice for students to master the material. A full-year or a semester statistics course would do that. The modules assume little to no statistical knowledge or exposure to advanced probability. Connections are made to calculus where relevant, though an understanding of calculus is not required in order to benefit from these materials.

The idea is to whet students' appetite with a different type of mathematics, one that is much more applied and related to the real world.

The Modules

This series includes four units:

- **Module 1 - Basic Statistics Concepts**. Focuses on graphical and numerical displays of data. Time required: 1–2 class periods.
- **Module 2 - The Normal Distribution**. Introduces a specific model for distribution of data. Time required: 1–2 class periods.
- **Module 3 - Expected Value and Standard Deviation of Discrete Probability Distributions**. Covers graphical and numerical displays of data. Time required: 1–3 class periods.
- **Module 4 - Correlation and Regression**. Teaches how to create models for relationships between two quantitative variables. Time required: 1–3 class periods.

Module 1: Basic Statistics Concepts

MODULE 1

In this module, students will learn about basic statistical concepts, including graphical summaries (histograms, dot plots) and numerical summaries (median, mean, standard deviation), and use them to analyze and interpret univariate data.

Content Learning Objectives

Through the analysis of actual historical hurricane data and data from the fictional town of Happy Shores, students will be able to do the following:

- Represent and analyze distributions of one quantitative variable by looking at dot plots and histograms.
- Compute and interpret summary statistics describing the center of a distribution (mean and median).
- Compute and interpret summary statistics describing the variability of a distribution (standard deviation).

Contextual Learning Objectives

Using the content knowledge, students will be able to do the following:

- Analyze the history of hurricanes by looking at how many storms occur in the U.S. each year.
- Gain an understanding of how a community like Happy Shores can be damaged based on the category of storm that hits.

Common Core State Standards for Mathematics

Interpreting Categorical and Quantitative Data (S-ID)

- Summarize, represent, and interpret data on a single count or measurement variable.
    1. *Represent data with plots on the real number line (dot plots, histograms, and box plots).*
    2. *Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.*
    3. *Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).*

Materials Needed

Much of the work in this module can be accomplished through the student handouts. Presentation format can vary. The lesson can be teacher directed or students can work in small groups. Although not required, students may wish to use the statistical capabilities of graphing calculator technology (TI-83/84). Students can also use statistical functions in spreadsheet programs such as Microsoft Excel. Guidelines and general instructions for using these tools are provided in the student modules in the Technology Connections section.

Timing

This module is meant to be completed in 1–2 class periods. Exercises can be done in class or for homework. Teachers can spend more time if necessary and interested.

Definitions

- **Statistics**—a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data
- **Data**—facts, statistics, or items of information
- **Distribution**—the values a variable takes and how often it takes those values
- **Histogram**—a type of bar graph that looks at the distribution of one quantitative variable and that may group values of the variable together
- **Dot plot**—a graph that looks at the distribution of one quantitative variable by plotting every data value as a dot above its value on a number line
- **Median**—the midpoint of a distribution where half the observations are smaller and the other half are larger
- **Mean**—the numerical average of a distribution
- **Mode**—the value in a range of values that has the highest frequency
- **Unimodal**—a description of shape for a distribution with a single mode (either a single value or range of values)
- **Bimodal**—a description of the shape of a distribution with two modes (either a single value or range of values)
- **Standard deviation**—a measure of how spread out the observations are from the mean in a distribution
- **Variability**—the spread of a variable or distribution
- **Outlier**—a data point in a sample that is widely separated from the main cluster of data points in that sample

Framing the Scenario

Actuaries at an insurance company have asked the class to help assess the risk and potential loss due to hurricanes in the coastal town of Happy Shores. Happy Shores is a small oceanfront community of approximately 200 households. Some homes are located on the beach while others are farther away, but all are within a few miles of the ocean. The actuaries need the students' help in understanding the characteristics of the community and what losses might occur due to hurricanes.

Teaching Tips

- Consider assigning the "Student Introduction" materials as reading homework before you begin working on the modules. This way, students will have a basic understanding about the motivation behind the modules. Then at the beginning of the first class period, you can begin by framing the scenario. As a class, brainstorm what data you might need in order to evaluate the risk and potential loss in the community. Students may come up with some of the following:
    - *Location, value and construction material of homes in Happy Shores*
    - *Historical information about hurricanes in Happy Shores and the geographic region*
    - *Historical insurance claims in the area due to hurricanes, along with historical claims in other areas*
    - *Data about hurricane frequency in the U.S.*
- There is no correct list of data. The idea is to brainstorm with students about what kind of data they could review.
- Define "Statistics"—a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.
- Look at historical data about hurricanes and tropical storms.
- It is much more important for students to understand the concept of standard deviation than the details of its computation. Stress the meaning of standard deviation as a measure of spread that gives us an indication of how data in a distribution vary from the mean.

Discussion Questions

Q1: What do the histogram and dot plot tell us about the distribution of hurricanes and tropical storms since 1932? (p. 8)

A1: We can see that the basic shape of both distributions is unimodal and skewed to the right. There appears to be an outlier year in which there are 15 hurricanes and 28 tropical storms (although we cannot necessarily assume that this is the same year from the graphs alone—we have to look at the raw data). We can see that the number of hurricanes spans from 2 to 15, and the number of tropical storms spans from 4 to 28 (spread). It appears that the middle of the hurricane distribution is around 6 and the middle of the tropical storm distribution is around 10 (center).

Q2: What do the histogram and dot plot NOT show that might be important? (p. 8)

A2: The histogram and dot plot do not show the year. We cannot see the changes in the number of hurricanes over time. This information would allow us to see the trend of hurricane/storm occurrences over time. It may be useful to know if it appears that the propensity of storms is increasing or decreasing. We also cannot gauge the intensity of storms; perhaps hurricane frequency and severity are related.

Q3: What if there was one year that had 30 hurricanes? How would this affect the median, mean, and standard deviation of the data? (p. 9)

A3: This would not adversely affect the median of the data. Median is resistant to the effect of outliers. This is a very important property of the median. Even if the maximum value is one million, the median is still the middle data value. Mean and standard deviation, however, are affected by outliers and would change if there were a year with 30 hurricanes. The further away the outlier, the more it will affect the mean and standard deviation. Mean and standard deviation are measures which are not resistant to outliers.

Practice Exercise (p. 9)

1. **Describe the distribution.**
The distribution appears unimodal and somewhat symmetrical. The spread is from about $60,000 to $150,000 and the center is around $100,000. We can say that the typical claim is around $100,000 with a minimum claim of $60,000 and maximum of $150,000. There don't appear to be any major outliers although the maximum claim of $150,000 is the highest by about $30,000.

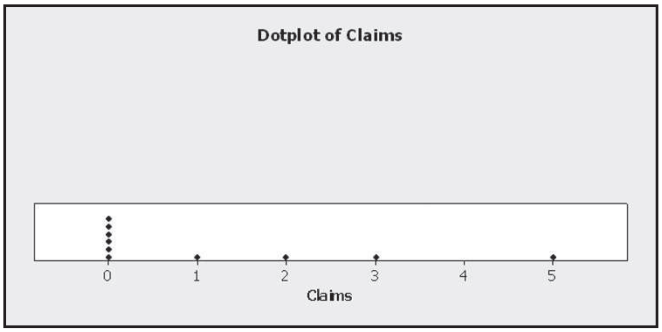
2. **Compute the median, mean, and standard deviation. The table below may help in computing the standard deviation.**
The mean of the distribution is 98.8 and the standard deviation is 21.33. The table that will aid students in computing the standard deviation is shown below:

Claims	Claim - Mean	(Claim - Mean)^2
112	13.2	174.24
92	-6.8	46.24
99	0.2	0.04
90	-8.8	77.44
117	18.2	331.24
79	-19.8	392.04
141	42.2	1780.84
66	-32.8	1075.84
86	-12.8	163.84
106	7.2	51.84
SUM		4093.6

The sum divided by 9 is 454.84 and the square root of this is the standard deviation of 21.33.

3. **What percentage of claims in the data set are (a) within 1 standard deviation of the mean (that is, from the mean minus the standard deviation to the mean plus the standard deviation); (b) within 2 standard deviations of the mean; and (c) within 3 standard deviations of the mean?**
a. 1 std dev is from 77.5 to 120.1 = 80%, b. 2 std dev is from 56.3 to 141.4 = 100%, c. 3 std dev = 100%

4. **Create a dot plot showing these claims.**



5. **Describe the distribution.**
Mean = 1.1; std dev = 1.73

6. **What percentage of claims in the data set are (a) within 1 standard deviation of the mean; (b) within 2 standard deviations of the mean; (c) within 3 standard deviations of the mean?**
80% are within 1 std dev, 90% are within 2 std devs, and 100% are within 3 std devs

7. **What are the main differences between the distributions of claims from homes farther down the beach and the one of claims from homes right on the beach?**
Clearly, claims are much lower for these 10 homes. Variability is also much smaller. The shape of this distribution is skewed right. 7 out of the 10 homes do not even have any claims. This tells us that homes on the beach are at much greater risk for damage in a Category 3 hurricane.

Module 2: The Normal Model

MODULE 2

In this module students will learn about standardizing a distribution and about a particularly useful distribution, the normal model.

Content Learning Objectives

Through the analysis of actual historical hurricane data and data from the fictional town of Happy Shores, students will be able to do the following:

- Compute z-scores (number of standard deviations from mean) in order to standardize values from a distribution.
- Understand when it is appropriate (and not appropriate) to use the normal model to represent real world data.
- Recognize connections between calculus and finding area under normal curves.
- Use a normal model to compute probabilities of outcomes.
- Use a normal model to estimate percentiles.

Contextual Learning Objectives

Using the content, students will be able to do the following:

- Estimate probabilities of the insurance company receiving different value claims when a certain category of hurricane hits the town.

Common Core State Standards for Mathematics

Interpreting Categorical and Quantitative Data (S-ID)

- Summarize, represent, and interpret data on a single count or measurement variable.

    4. *Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.*

Materials Needed

Much of the work in this module can be accomplished through the student handouts. Presentation format can vary. The lesson can be teacher directed or students can work in small groups. Although not required, students may wish to use the statistical capabilities of graphing calculator technology (TI-83/84). Students can also use statistical functions in spreadsheet programs such as Microsoft Excel. Guidelines and general instructions for using these tools are provided in the student modules in the Technology Connections section.

Timing

The module is meant to be completed in 1–2 class periods. Exercises can be done in class or for homework. Teachers can spend more time if necessary and interested.

Definitions

- **Standardized values**—values that can be compared between distributions by looking at the number of standard deviations from the mean
- **Z-scores**—a common name for standardized values
- **Model**—the description of a distribution using a mathematical curve that approximately fits the histogram of the data
- **Normal model**—a distribution that is symmetrical, bell-shaped and unimodal
- **Parameters**—the mean and standard deviation of a model
- **Percentile**—the value in a distribution below which a certain percent of observations fall

Discussion Questions

Q1: Compute the number of standard deviations from the mean for all the observations (p. 15)

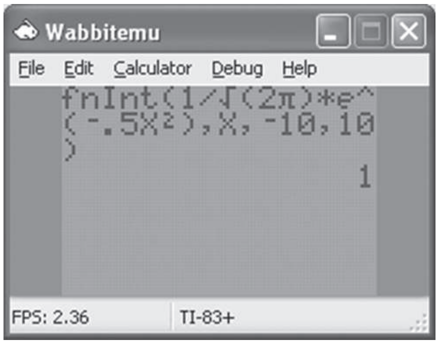
A1:

Claims	sst dev
112	0.61904762
92	-0.33333333
99	0
90	-0.42857143
117	0.85714286
79	-0.95238095
141	2
66	-1.57142857
86	-0.61904762
106	0.33333333

Calculus Connection (p. 17)

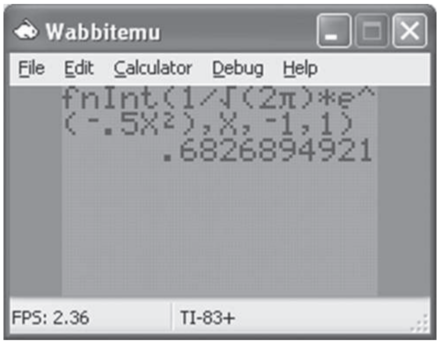
Can you compute the area under the standard normal model?

1

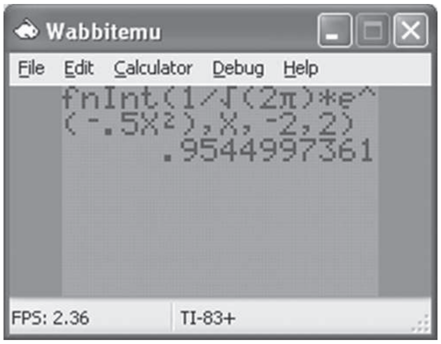


Using the same calculus techniques, find the following areas under the standard normal model:

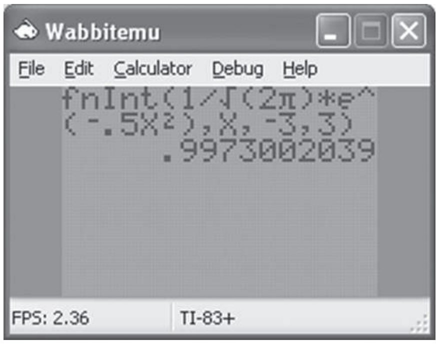
- Area between -1 and 1



- Area between -2 and 2



- Area between -3 and 3

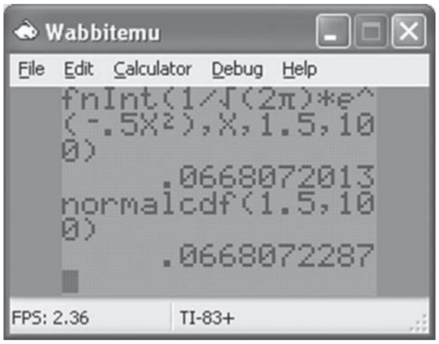


Practice Exercise 1 (p. 17)

If SAT scores can be modeled with a normal distribution, and the mean score is 500 with a standard deviation of 100, then find the following probabilities:

1. **The probability that someone scores between 400 and 600**
   About 0.68. Note that you can use the answer to the Calculus Connection question for the percent of observations between z-scores of -1 and 1

2. **The probability that someone scores over 600**
   About 0.16. Since 68% is in the middle, this leaves 32% in the 2 tails. We want the upper tail which is exactly half of 32% since the normal model is symmetrical



3. **The probability that someone scores over 650**
   Z-score = 1.5

   Answer: 0.0668

   2 methods using calculator are shown to the right

4. **The probability that someone scores between 450 and 600**
   Between z-scores of -0.5 and 1

   Answer: 0.532

5. **The probability that someone scores less than 420**
   Z-score of -0.8

   Answer: 0.222

Practice Exercise 2 (p. 20)

Suppose the data looking at insurance claims of oceanfront homes due to a category 3 hurricane are appropriately modeled by a normal curve with a mean of 99 (thousand) and a standard deviation of 21 (thousand).

1. **If a category 3 hurricane hits, what is the probability that a particular household files a claim for more than $110,000?**
   Z-score = 11/21 = 0.5238

   Answer: about 0.3

2. **What is the probability that a particular household files a claim for more than $150,000?**
Z-score = 2.43

Answer = 0.00758

3. **What is the probability that a particular household files a claim for less than $90,000?**
Z-score = -0.429

Answer = 0.334

4. **What claim would represent the 90th percentile?**
Z-score = -1.28

Answer: about $125, 913

5. **Approximately 5% of all claims would be below what amount?**
Z-score = -1.645

Answer: about $64,458

Module 3: Discrete Probability Distributions

MODULE 3

In this module, students will learn about discrete probability distributions. They will compute value and standard deviation of a probability distribution and use this information to understand how much an insurance company might need to pay out if a hurricane hits Happy Shores.

Content Learning Objectives

Through the use of historical data about hurricanes and Happy Shores, students will be able to do the following:

- Estimate probabilities based on historical empirical data.
- Construct a probability distribution (probability model) for a discrete situation.
- Compute and interpret the expected value of a discrete probability distribution.
- Compute and interpret the standard deviation of a discrete probability distribution.

Contextual Learning Objectives

Using the content, students will be able to do the following:

- Estimate probabilities of different level storms hitting Happy Shores based on historical data.
- Find the expected damage a hurricane may cause along with the standard deviation. This will be done for each category of storm and neighborhood within Happy Shores.

Common Core State Standards for Mathematics

Using Probability to Make Decisions (S-MD)

- Calculate expected values and use them to solve problems.

    1. *Define a random variable for a quantity of interest by assigning a numerical value to each event in a sample space; graph the corresponding probability distribution using graphical displays as for data distributions.*
    2. *Calculate the expected value of a random variable; interpret it as the mean of a probability distribution.*
    3. *Develop a probability distribution for a random variable defined for a sample space in which theoretical probabilities can be calculated; find the expected value.*
    4. *Develop a probability distribution for a random variable defined for a sample space in which probabilities are assigned empirically; find the expected value.*

Materials Needed

Much of the work in this module can be accomplished through the student handouts. Presentation format can vary. The lesson can be teacher directed or students can work in small groups. Although not required, students may wish to use the statistical capabilities of graphing calculator technology (TI-83/84). Students can also use statistical functions in spreadsheet programs such as Microsoft Excel.

Timing

This module is designed to take 1–3 class periods depending on the depth desired.

Definitions

- **Random phenomena**—completely unpredictable outcomes in the short term
- **Trial**—each occasion in which a random phenomenon is observed
- **Outcome**—the value of the random phenomenon at each trial
- **Sample space**—all possible outcomes of the random phenomenon
- **Probability**—the likelihood or chance of a certain outcome occurring
- **Probability distribution (probability model)**—a table of outcomes and probabilities
- **Discrete probability model**—a distribution where the outcomes only take certain values
- **Continuous**—a distribution where the outcomes can take on any value in a given interval
- **Expected value**—the mean of the probability distribution
- **Standard deviation of a random variable**—a measure of the variation from the mean in a probability distribution

Discussion Questions

Q1: Fill in the following table based on historical data: (p. 21)

A1:

Outcome	Probability
No storms	0.7
Tropical Storm	0.14
Category 1 Hurricane	0.08
Category 2 Hurricane	0.05
Category 3 Hurricane	0.01
Category 4 Hurricane	0.01
Category 5 Hurricane	0.01

Q2: What would the probability distribution for this insurance company be: (p. 22)

A2:

Policyholder Outcome	Payout (x)	Probability P(X = x)
Death	10,000	1/1000
Disability	5,000	2/1000
Neither	0	997/1000

Practice Exercise 1 (p. 23)

1. **Create a probability distribution for the possible claim amounts in a given year. Fill out the following table:**

Results	Cat 5	Cat 4	Cat 3	Cat 2	Cat 1	TS	NONE	
Claim Amt	500	350	100	50	25	5	0	
Probability	0.01	0.01	0.01	0.05	0.08	0.14	0.7	
Amt * Prob	5	3.5	1	2.5	2	0.7	0	**14.7**

2. **Find the expected amount of the claim for these homes.**
Expected Value = $14,700

3. **Based on these numbers, what do you think is a reasonable amount for the insurance company to charge as its premium for hurricane insurance for these homes? (Remember, the insurance company needs to make a profit!)**
Answers may vary. Certainly the insurance company should charge more than $14,700 per year for a house on the beach.

Practice Exercise 2 (p. 25)

1. **We found that the average claim for a home in Happy Shores on the beach in a given year is $14,700. Compute the standard deviation for the claim amount in a given year.**

Results	Cat 5	Cat 4	Cat 3	Cat 2	Cat 1	TS	NONE
Claim Amt	500	350	100	50	25	5	0
Probability	0.01	0.01	0.01	0.05	0.08	0.14	0.7
Amt - Mean	485.3	335.3	85.3	35.3	10.3	-9.7	-14.7
(Amt - Mean)^2	235516.09	112426.1	7276.09	1246.09	106.09	94.09	216.09
Prob*(Amt - Mean)^2	2355.1609	1124.261	72.7609	62.3045	8.4872	13.1726	151.263

3787.41	Variance
61.54193692	Standard Deviation

For the houses in Neighborhood A, we would expect to pay out $14,700 with a standard deviation of $62,000.

Additional Exercises (p. 25)

NEIGHBOORHOD B

Expected Value 8.225

	Home Value ($ thousands)	Cat 5	Cat 4	Cat 3	Cat 2	Cat 1	TS	None
Probability		0.01	0.01	0.01	0.05	0.08	0.14	0.7
Damage %	250	80%	50%	15%	10%	8%	5%	0%
Claim Amount		200	125	37.5	25	20	12.5	0
(Claim - Mean)		191.78	116.78	29.275	16.775	11.775	4.275	-8.225
(Claim - Mean)^2		36778	13636	857.03	281.4	138.65	18.276	67.651
Prob*(Claim - Mean)^2		367.78	136.36	8.5703	14.07	11.092	2.5586	47.355

Variance 587.78688

Std. Dev 24.244316

NEIGHBORHOOD C

Expected Value 11.515

	Home Value ($ thousands)	Cat 5	Cat 4	Cat 3	Cat 2	Cat 1	TS	None
Probability		0.01	0.01	0.01	0.05	0.08	0.14	0.7
Damage %	350	80%	50%	15%	10%	8%	5%	0%
Claim Amount		280	175	52.5	35	28	17.5	0
(Claim - Mean)		268.49	163.49	40.985	23.485	16.485	5.985	-11.52
(Claim - Mean)^2		72084	26727	1679.8	551.55	271.76	35.82	132.6
Prob*(Claim - Mean)^2		720.84	267.27	16.798	27.577	21.74	5.0148	92.817

Variance 1152.0623

Std. Dev 33.942043

NEIGHBORHOOD D

Expected Value 6.58

	Home Value ($ thousands)	Cat 5	Cat 4	Cat 3	Cat 2	Cat 1	TS	None
Probability		0.01	0.01	0.01	0.05	0.08	0.14	0.7
Damage %	200	80%	50%	15%	10%	8%	5%	0%
Claim Amount		160	100	30	20	16	10	0
(Claim - Mean)		153.42	93.42	23.42	13.42	9.42	3.42	-6.58
(Claim - Mean)^2		23538	8727.3	548.5	180.1	88.736	11.696	43.296
Prob*(Claim - Mean)^2		235.38	87.273	5.485	9.0048	7.0989	1.6375	30.307

Variance 376.1836

Std. Dev 19.395453

NEIGHBORHOOD E

Expected Value 8.28

	Home Value ($ thousands)	Cat 5	Cat 4	Cat 3	Cat 2	Cat 1	TS	None
Probability		0.01	0.01	0.01	0.05	0.08	0.14	0.7
Damage %	400	60%	30%	10%	5%	5%	3%	0%
Claim Amount		240	120	40	20	20	12	0
(Claim - Mean)		231.72	111.72	31.72	11.72	11.72	3.72	-8.28
(Claim - Mean)^2		53694	12481	1006.2	137.36	137.36	13.838	68.558
Prob*(Claim - Mean)^2		536.94	124.81	10.062	6.8679	10.989	1.9374	47.991

Variance 739.6016

Std. Dev 27.195617

NEIGHBORHOOD F

Expected Value 2.805

	Home Value ($ thousands)	Cat 5	Cat 4	Cat 3	Cat 2	Cat 1	TS	None
Probability		0.01	0.01	0.01	0.05	0.08	0.14	0.7
Damage %	150	50%	20%	10%	5%	5%	3%	0%
Claim Amount		75	30	15	7.5	7.5	4.5	0
(Claim - Mean)		72.195	27.195	12.195	4.695	4.695	1.695	-2.805
(Claim - Mean)^2		5212.1	739.57	148.72	22.043	22.043	2.873	7.868
Prob*(Claim - Mean)^2		52.121	7.3957	1.4872	1.1022	1.7634	0.4022	5.5076

Variance 69.779475

Std. Dev 8.353411

NEIGHBORHOOD G

Expected Value 1

	Home Value ($ thousands)	Cat 5	Cat 4	Cat 3	Cat 2	Cat 1	TS	None
Probability		0.01	0.01	0.01	0.05	0.08	0.14	0.7
Damage %	100	25%	10%	10%	5%	2%	1%	0%
Claim Amount		25	10	10	5	2	1	0
(Claim - Mean)		24	9	9	4	1	0	-1
(Claim - Mean)^2		576	81	81	16	1	0	1
Prob*(Claim - Mean)^2		5.76	0.81	0.81	0.8	0.08	0	0.7

Variance 8.96

Std. Dev 2.9933259

Module 4: Correlation and Regression

MODULE 4

In this module, students will learn about creating models to describe relationships between 2 quantitative variables. This will allow students to understand the relationship between hurricane damage amounts for homes based on distance from the ocean and other factors.

Content Learning Objectives

Through the use of historical data about hurricanes and Happy Shores, students will be able to do the following:

- Create scatterplots to describe 2-variable data.
- Interpret the Pearson Correlation Coefficient describing the strength and direction of the linear relationship between variables.
- Compute linear regression models and interpret them.
- Interpret correlation tables.

Contextual Learning Objectives

Using the content, students will be able to do the following:

- Create models for the relationship between the insurance claim amount for a household based on its distance from the ocean and the category of storm that hits.
- Analyze correlation tables to determine key variables related to the magnitude of insurance claims (square footage, material home is made of, elevation, etc.).
- Analyze recent trends in tropical storms and hurricanes in the U.S.

Common Core State Standards for Mathematics

Interpreting Categorical and Quantitative Data (S-ID)

- Summarize, represent and interpret data on two categorical and quantitative variables.
    - Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

        6. *Fit a function to the data; use functions fitted to data to solve problems in the context of data. Use given functions or choose by the context. Emphasize linear, quadratic and exponential models.*

            a. *Informally assess the fit of a function by plotting and analyzing residuals.*

            b. *Fit a linear function for a scatterplot that suggests a linear association.*

- Interpret linear models

    7. *Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.*

    8. *Compute (using technology) and interpret the correlation coefficient of linear fit.*

    9. *Distinguish between correlation and causation.*

Materials Needed

Much of the work in this module can be accomplished through the student handouts. Presentation format can vary. The lesson can be teacher directed or students can work in small groups. Although not required, students may wish to use the statistical capabilities of graphing calculator technology (TI-83/84). Students can also use statistical functions in spreadsheet programs such as Microsoft Excel. Guidelines and general instructions for using these tools are provided in the student modules in the Technology Connections section.

Timing

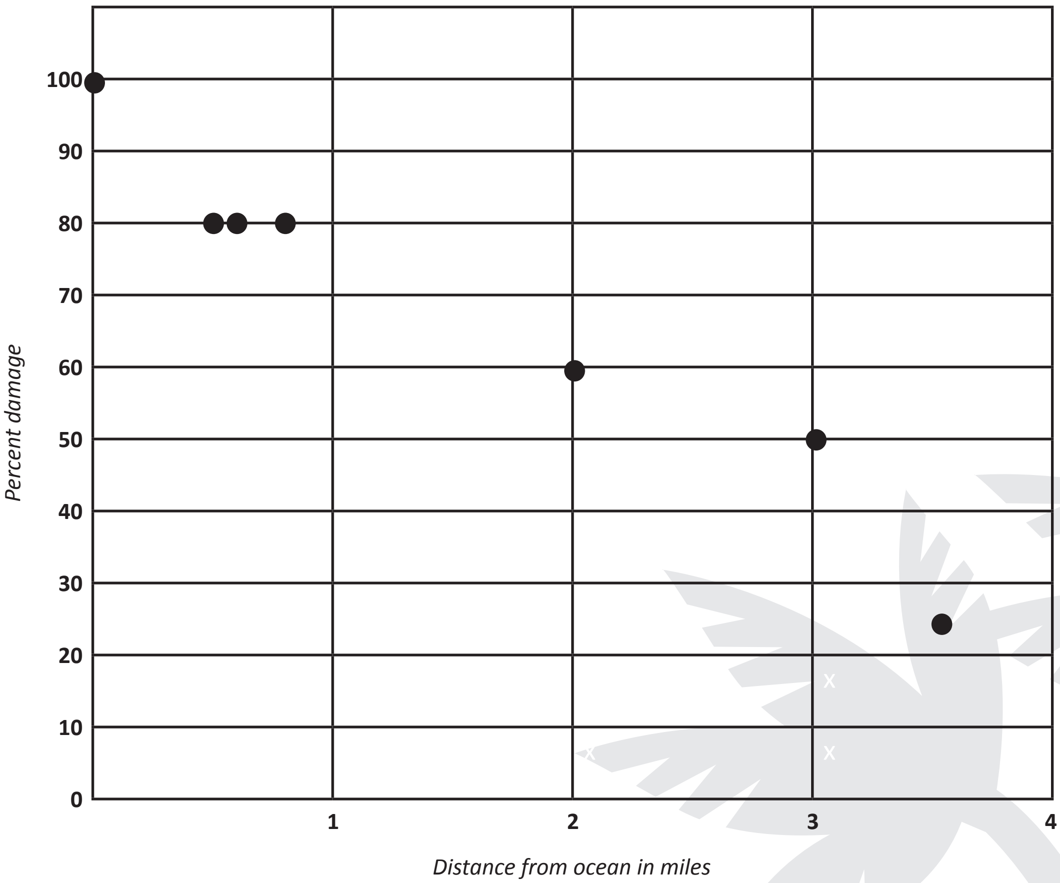
This module is designed to take 1–3 class periods depending on the depth desired.

Definitions

- **Scatterplot**—the most common graph for looking at the relationship between two quantitative variables
- **Response variable**—the y-axis on a scatterplot
- **Explanatory variable**—the x-axis on a scatterplot
- **Correlation coefficient**—a measure of the strength and direction of the linear relationship between two quantitative variables
- **Linear regression**—a predictive model that creates a line of best fit for a set of data points
- **Correlation table**—a table showing the correlations between every pair of variables in a collection of variables

Discussion Question

Plot each of these points [as found in the student guide] on this graph: (p. 26)



Discussion Questions (p. 30)

Q1: What is the y-intercept in the model? What does it represent?

A1: The y-intercept is 94.24. This represents the predicted damage percentage for homes that are right on the beach.

Q2: What is the slope in the model? What does it represent? (p. 30)

A4: The slope is -17.76. This means that for each mile a home is away from the ocean, we predict about 17.76% less damage.

Q3: Use the model to predict the percent damage of a home that is 1.5 miles from the ocean.

A4: Plug in 1.5 into equation: $94.24 – 17.76(1.5) = 67.6\%$

Q4: Use the model to predict the percent damage of a home that is 5 miles from the ocean.

A4: $94.24 – 17.76(5) = 5.44\%$

Practice Exercise 1 (p. 31)

We only looked at damages due to a Category 5 hurricane. What about other storm categories? Pick a storm category and create a regression model relating the percent damage to the miles from the ocean. Also report the correlation. Is the relationship stronger or weaker than the one we looked at for a category 5 hurricane?
Cat 4 % Damage = 0.623 - 0.150 Dist; r = -0.97942

Cat 3 % Damage = 0.173 - 0.0250 Dist; r = -0.89826

Cat 2 % Damage = 0.106 - 0.0183 Dist; r = -0.92895

Cat 1 % Damage = 0.0770 - 0.0124 Dist; r = -0.74275

TS % Damage = 0.0412 - 0.00559 Dist; r = -0.42122

The relationship between the damage and the distance from the ocean seems to weaken for Category 3, 2, 1 and tropical storms. The relationship is still quite significant though.

Practice Exercise 2 (p. 32)

What seems to be correlated with percent damage to the home? Explain each variable and the strength and direction of the correlation.
Distance seems to have a strong negative association with damage. Homes closer to the ocean are damaged more. A weak positive association exists between square footage and percent damage. As homes are bigger they tend to get damaged more, but the relationship is not that strong. Elevation has a moderate negative relationship. Homes that are lower in elevation tend to get damaged more. There is a strong positive association between the amount of wood construction in the house and the damage.

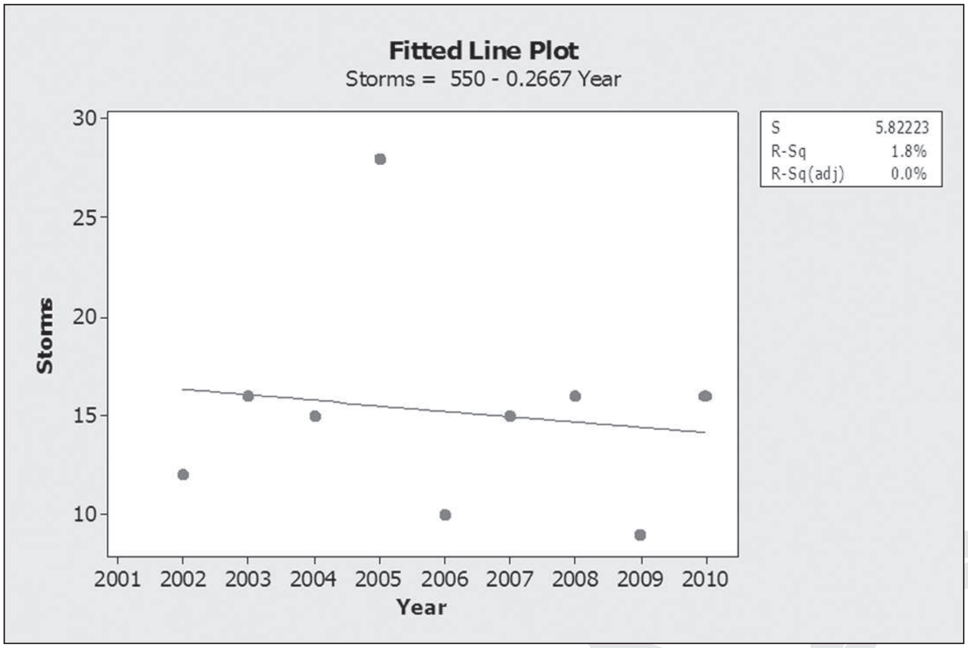
What is NOT correlated strongly with percent damage to the home?
Little association exists between damage and the number of inhabitants in the home.

Describe any other patterns you may see.
Answers can vary.

How could an insurance company use this information when trying to decide what to charge different households for hurricane insurance?
An insurance company would know to charge higher premiums to those homes located closer to the ocean, made of wood, and having lower elevations.

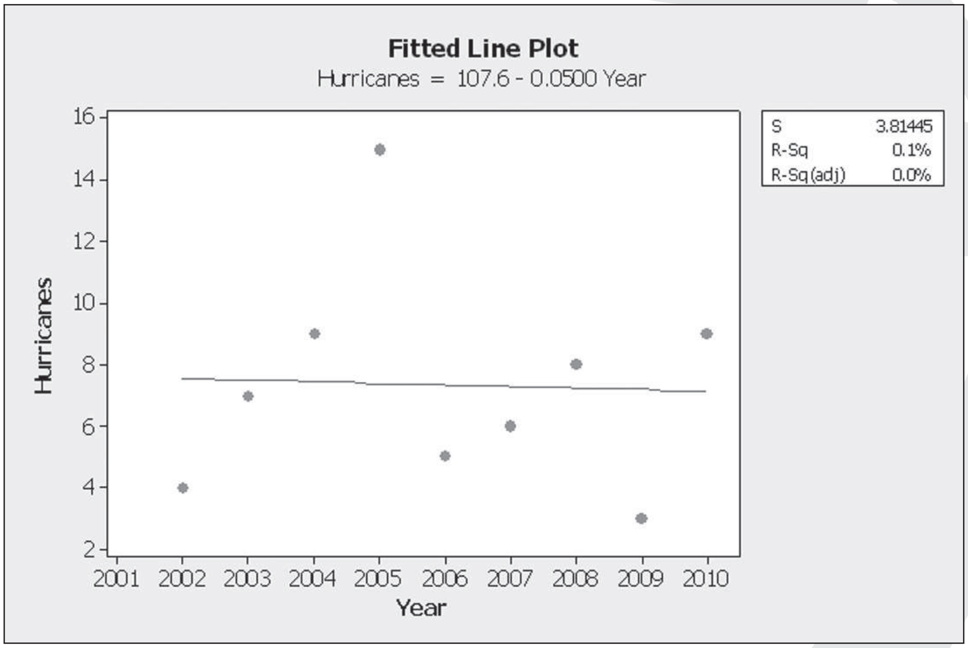
Practice Exercise 3 (p. 33)

1. Number of Hurricanes vs. Year



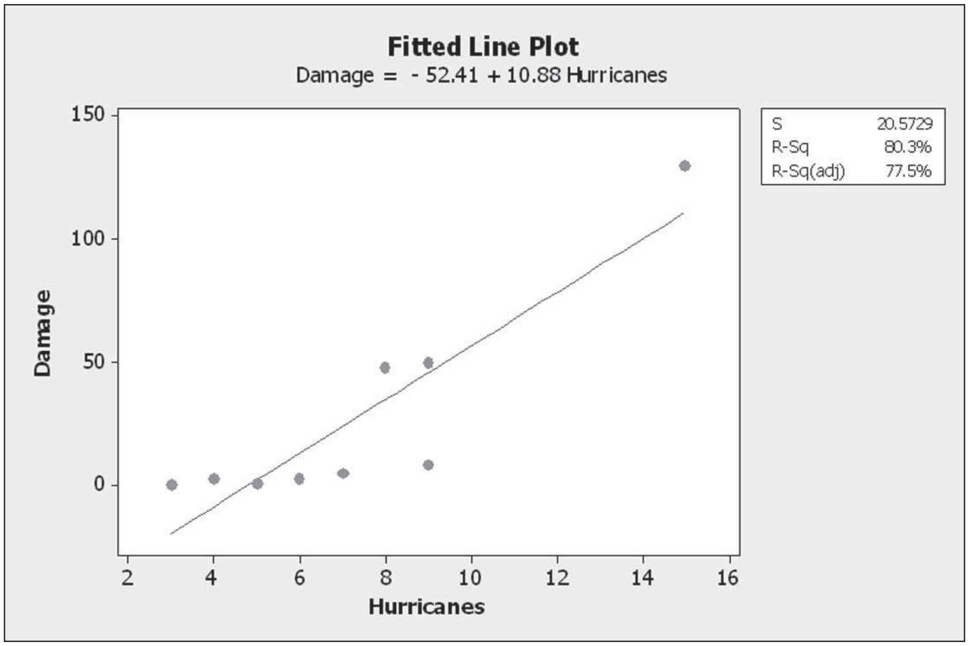
There does not seem to be a trend.

2. Number of Total Storms vs. Year



There is a weak downward trend in the number of storms per year.

3. Damage vs. Number of Hurricanes



As the number of hurricanes increases, the extent of the damage increases.

Looking for More?

We hope that you and your students have enjoyed the activities in this series. Many students with an aptitude for math may find the actuarial field to be a natural fit with their strengths and professional interests; while this unit offers a good, hands-on introduction to the field, there are other resources that can provide them with more challenges and more information on the profession.

If you or your students are interested in trying out a few actuarial problems to see if this career is for them, have them go to: www.beanactuary.org/why/?fa=solve-real-problems.

Or for a group project, have the class break out into teams and try to answer questions from real actuarial exams that have been given out in the past, which can be found at: www.soa.org/education/exam-req/syllabus-study-materials/edu-multiple-choice-exam.aspx.

Contact The Actuarial Foundation at programs@actfnd.org to see if an Actuary is available to present some of the materials in this book or to work with your students on some of the exam questions mentioned above and to talk about the profession.

And of course, they can learn more about the profession itself by visiting: www.BeAnActuary.org.

Thanks for sharing these materials with your students – we look forward to supplying you with even more resources in the future!

- The Actuarial Foundation

THE ACTUARIAL FOUNDATION®

Preparing for tomorrow's possibilities® *through education*

475 North Martingale Road, Suite 600 • Schaumburg, IL 60173-2226

Ph 847-706-3535 • Fx 847-706-3599

www.actuarialfoundation.org