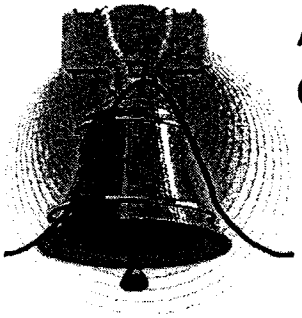
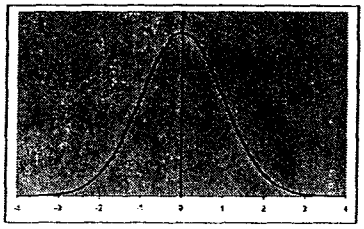
Statistics Part II: Understanding the Bell Curve



Administrative Staff Analyst

OSA Training



Course Summary

- Using data about a population to draw graphs
- Frequency distribution and variability within populations
- Bell Curves: What are they and where do we see them?
- Normal distribution
- Interpreting bell curves by their mean, variance, and standard deviation
- Understanding and calculating Z scores
- Proportion: Calculating the area under the curve
- Skewness in Curves
- Correlation: What is the relationship between two variables?

2

Using data to draw graphs about a population

- A statistic is a way to represent or organize information in a way that helps you understand it better than simply looking at a series of numbers.
- You can use a set of data to draw a picture that will help you to understand and interpret that data.

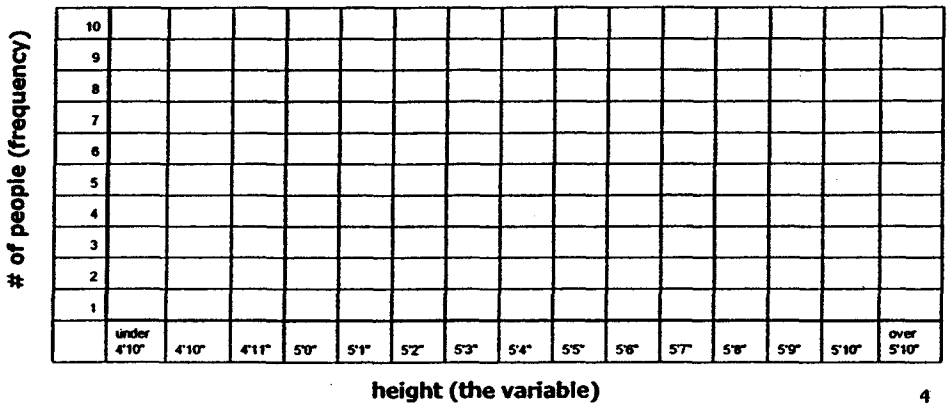
3

Using Data to Draw Graphs:

In-Class Exercise of Height Frequency Distribution

Instructions: Fill in the graph according to the results in class.

Frequency Distribution of Women's Height



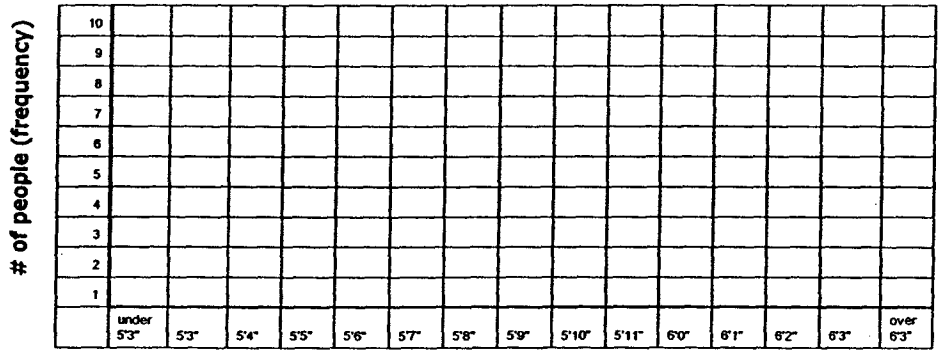
4

Using Data to Draw Graphs:

In-Class Exercise of Height Frequency Distribution

Instructions: Fill in the graph according to the results in class.

Frequency Distribution of Men's Height

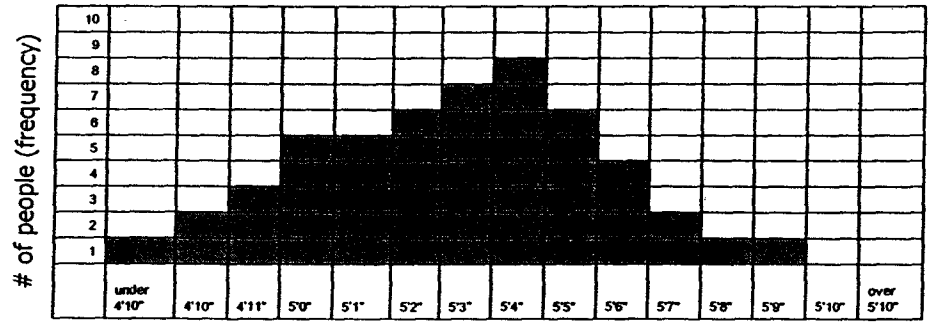


5

Example: Height distribution among a group of 55 women

- The X axis (horizontal) refers to the variable, or the observation value that you are looking at in a population.
- The Y axis (vertical) reflects the frequency, or the number of times a particular value of X appears in a population.

Example of the Frequency Distribution of Women's Height



6

Properties of Populations

Definition of a Population

- A population is any group whose characteristics you look at. A population is different from a sample, which is a small portion of the population used to generalize about the whole population.

Central Tendency

- Large populations often tend to cluster towards their middle, or average, which is also known as the **mean.**

Variability

- In large populations, there is often a lot of diversity. For example, people come in a variety of heights and weights.

7

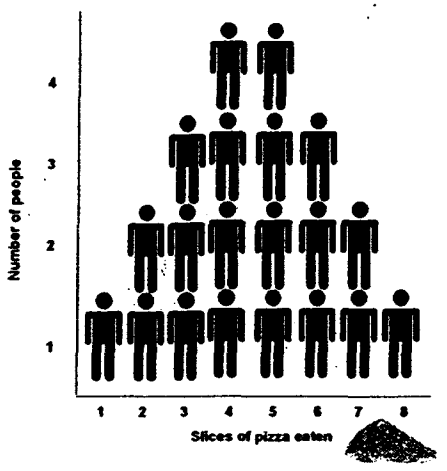
Example: The Hungry Softball Team

Situation

A softball team has just won a game. All 20 players on the team – the population – have gone to eat pizza.

Graph

A simple graph shows how many slices each of the 20 team members ate. For example, four people ate 5 slices of pizza, while only one person ate 8 slices.

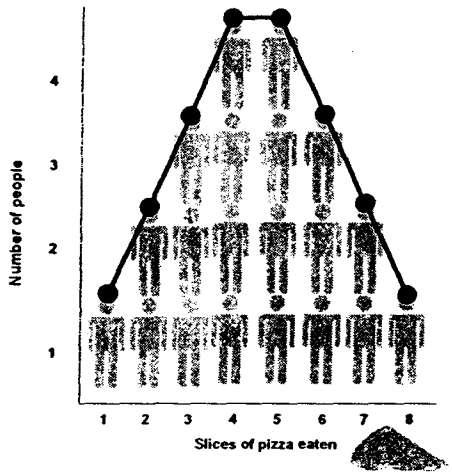


8

Example:
The Hungry Softball Team

Graph

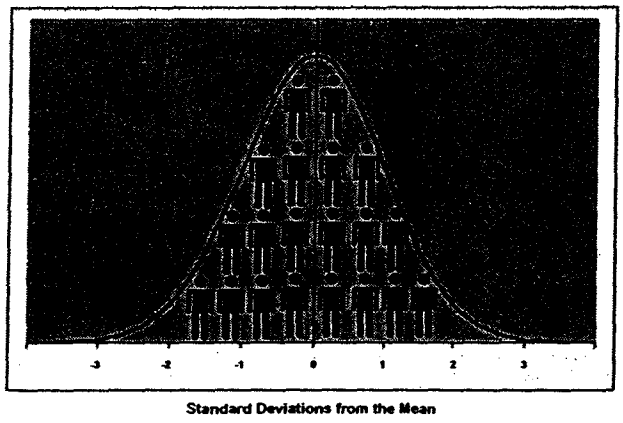
A line shows how you could draw a simple graph using the tops of the heads of each group of players.



Example:
The Hungry Softball Team

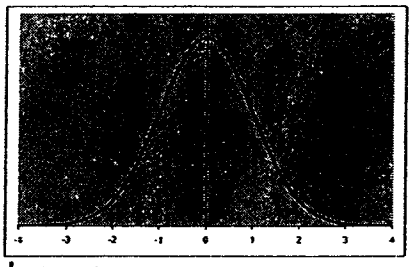
Graph

This graph is a simplification of how you could graph pizza slices eaten into a bell curve.



10

Bell Curves: What are they?



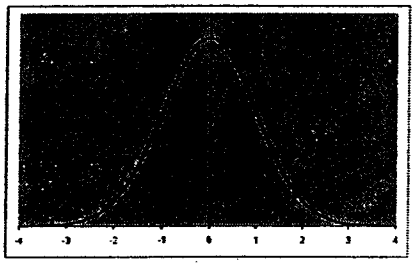
Basic Properties

- A bell curve is a very special kind of curve with unique properties.
- It is shaped like a bell.
- Also called a "normal curve" or "normal distribution," it shows how frequently different values recur in a population.
- It is symmetric and has a single peak at its mean.
- Its unique properties make it very useful in making statistical calculations.

11

Bell Curves: Where do we see them?

- Normal distributions occur often, especially when a large group of data is concerned.
- Examples:
  - Height
  - Weight
  - SAT scores
  - IQ



12

Bell Curves: Where do we see them?

- Example: fish size
- This diagram illustrates how MOST fish in a given species fall pretty close to the average
- Very small or large fish – called **outliers** because of their uncommon size – are much more rare and show up on one end of the bell curve.

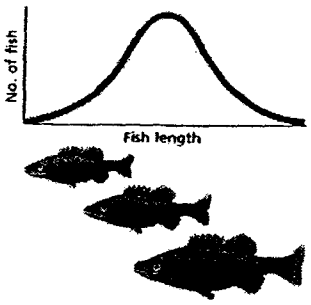


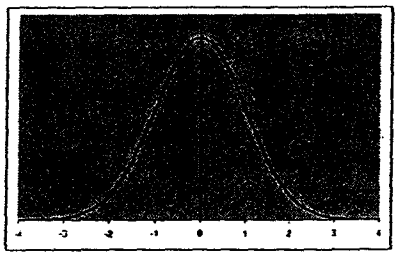
Figure 16-2 A bell curve illustrates how most members of a population are grouped in an average range for a given trait while only a few are at the extreme ends of the range.

13

Bell Curves: Mean

Mean

- The mean shows the average of all the values in a population.



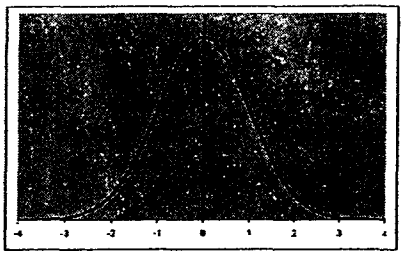
Mean = Addition of all the observations together
# of observations

14

Bell Curves: Variance

Variance

- A measure of the variability of the population described by a bell curve.
- Calculated by adding together the square of the difference between EACH observation and the mean



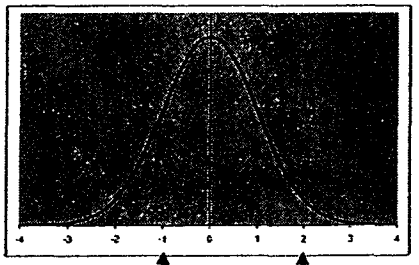
Variance = Sum of (each observation − mean)²
/ # of all observations

15

Bell Curves: Standard Deviation

Standard Deviation

- "A rough measure of the average amount by which observations deviate on either side of their means" (Witte & Witte, 2001)
- It's a way of measuring how far any observation is from the mean.
- In precise terms, it's the square root of the variance.



One standard deviation from the mean; Z = -1

Two standard deviations from the mean; Z = 2

16

Bell Curves: What are they?

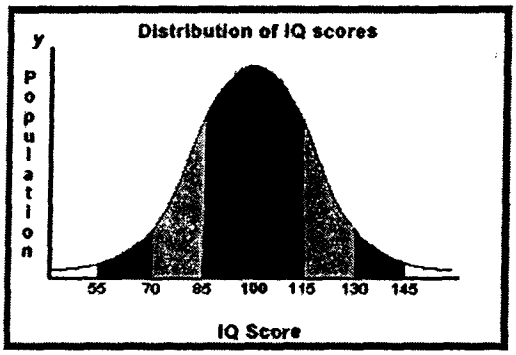
Advanced Properties

- They extend approximately 3 standard deviations above and below the mean.
- They have a total area under the curve of 1.00 (100%).
- The mean, median, and mode of a normal distribution are identical and fall exactly in the center of the curve.

17

Z Scores: What are they?

- Z scores are a way to convert real data in the world into a form that fits on a bell curve.
- This only works if you have a normal distribution to begin with.
- IQ is a very standard example of a normal distribution that can be easily converted to Z scores.



18

Z Scores: What are they?

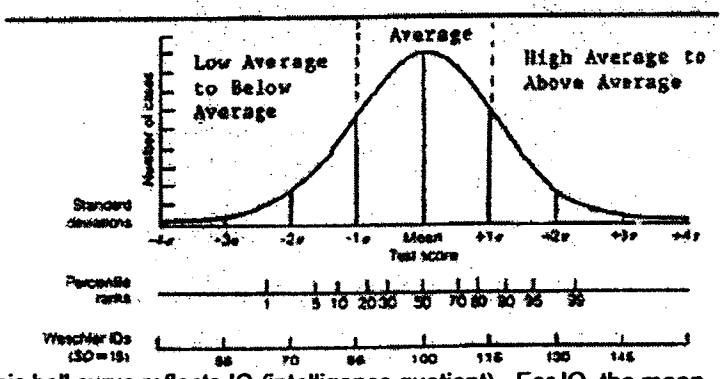
Mean and Standard Deviation

- The mean always has a Z score of 0.
- Other scores are converted to Z scores by their distance from the mean – how many standard deviations they are from the mean.
- The standard deviation is always equal to 1.
  - Example 1: Z score of -2 means that the value of an observation is two standard deviations from the mean.
  - Example 2: If a person's height is one standard deviation above the mean, the Z score for his or her height is equal to one.

19

Z Scores: How to Calculate Them

$$\text{Z score} = \frac{\text{Mean} - \text{observation value}}{\text{Standard deviation}}$$

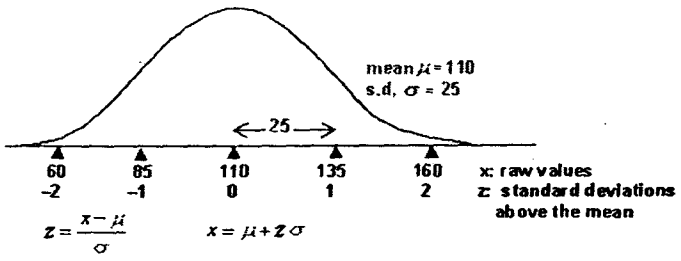


This bell curve reflects IQ (intelligence quotient). For IQ, the mean equals 100 and the standard deviation equals 15.

20

"Real" Data Compared with Z Scores: Example

- The diagram below illustrates how you convert "real" numbers or "raw values" into Z scores.
  - This example has a mean of 110 and standard deviation of 25.
  - Again, when you have a normal ("bell") curve, you can always convert the numbers so that the mean is 0 and a standard deviation is equal to 1.



21

Proportion

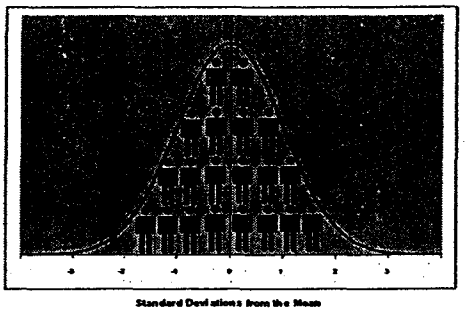
- The area under a bell curve tells you what percentage of ALL observations fall within that area.
- The total area under a bell curve is always equal to 1, or 100%.

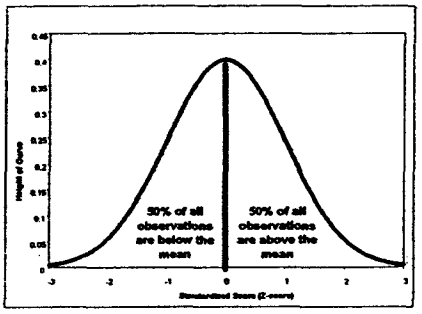
22

Proportion Example: The Hungry Softball Team Redux

Definition of Proportion

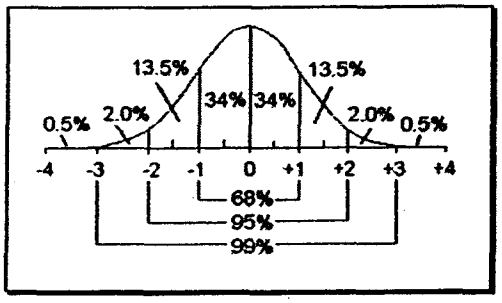
Think of proportion as counting the number of observations that would fit under a certain part of the curve.





23

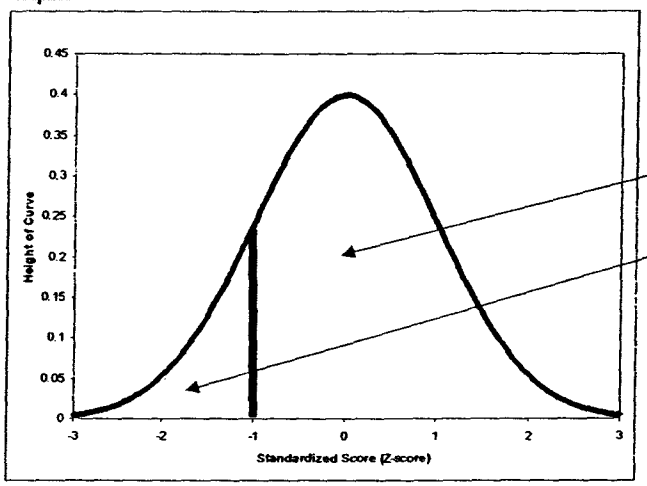
Proportion: Properties of All Normal Distributions



- 68% of observations fall within 1 standard deviation of the mean (34% on either side)
- 95% of observations fall within 2 standard deviations of the mean (47.5% on either side)
- 99% of observations fall within 3 standard deviations of the means (49.5% on either side)

24

Proportion: Example



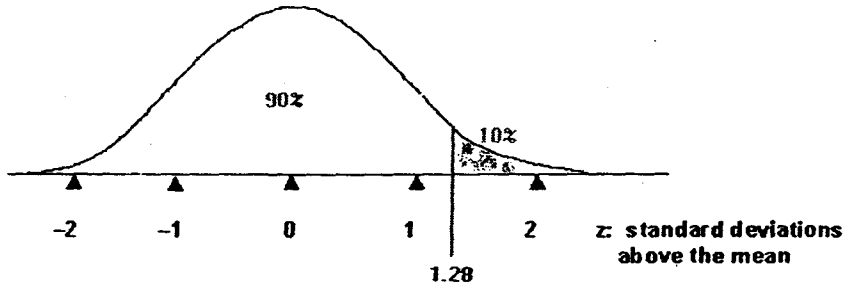
84% of observations fall to the right of the Z score of -1.

16% fall to the left of it.

25

Proportion

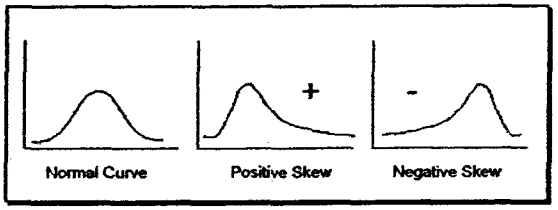
- Except at the mean, percentages are not "pretty" numbers (an even multiple of 10) when you have a whole number Z score
- Similarly, the Z score is usually not a "pretty" whole number when you have a "pretty" percentage.



26

Skew in Bell Curves

- The skew of a distribution refers to how the curve leans.
- When a curve has extreme scores on the right hand side of the distribution, it is said to be positively skewed. In other words, when high numbers are added to an otherwise normal distribution, the curve gets pulled in an upward or positive direction.
- When the curve is pulled downward by extreme low scores, it is said to be negatively skewed. The more skewed a distribution is, the more difficult it is to interpret.[1]



[1] Text from: http://allpsych.com/researchmethods/distributions.html.

27

Correlation: What is the relationship between two variables?

Overview

- Up to this point, the discussion has been focused on bell curves. Bell curves only really measure the distribution of one variable within a population.
- Correlation, by contrast, refers to the relationship between TWO variables within a population.

28

Correlation: What is the relationship between two variables?

Direction

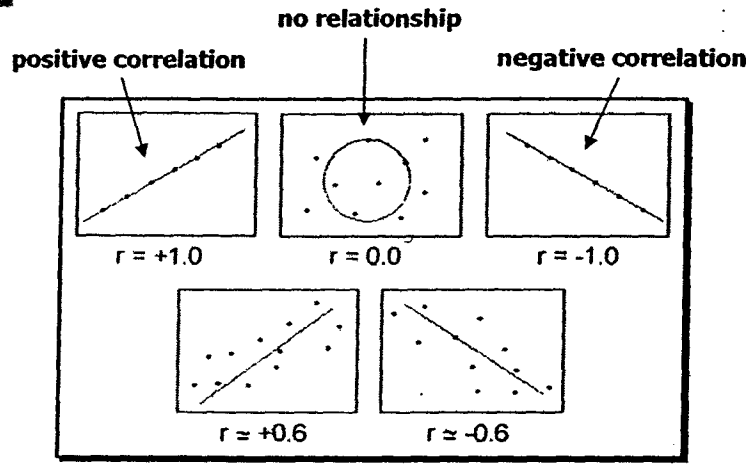
- Positive correlation: When you see an **increase** in one variable, you also tend to see an **increase** in the other variable.
  - Example: Income and SAT scores. As income rises, so, too, do SAT scores tend to rise for students.
- Negative correlation: When you see an **increase** in one variable, you tend to see a **decrease** in the other variable.
  - Example: alcohol consumption and manual dexterity. As the number of drinks someone has rises, his or her score on a manual dexterity test will tend to fall.
- No relationship: The two variables do not affect each other at all.
  - Example: Ice cream consumption and shark attacks.

Intensity ("r")

- How strong is the relationship between two variables?
- Values of r = 1 or r = -1 are the strongest, while r= 0 is the weakest.

29

Types of Correlation



30

Correlation vs. Causation: Being Cautious with Conclusions

- One common mistake is made by people interpreting a correlation as meaning that one thing **causes** another thing. When we see that depression and self-esteem are negatively correlated, we often surmise that depression must therefore cause the decrease in self-esteem. When contemplating this, consider the following correlations that have been found in research:
  - Positive correlation between ice cream consumption and drownings
  - Positive correlation between ice cream consumption and murder
  - Positive correlation between ice cream consumption and boating accidents
  - Positive correlation between ice cream consumption and shark attacks
- If we were to assume that every correlation represents a causal relationship then ice cream would most certainly be banned due to the devastating effects it has on society. Does ice-cream consumption cause people to drown? Does ice cream lead to murder? The truth is that often two variables are related only because of a third variable that is not accounted for within the statistic. In this case, the weather is this third variable because as the weather gets warmer, people tend to consume more ice cream. Warmer weather also results in an increase in swimming and boating and therefore increased drownings, boating accidents, and shark attacks.

31

Correlation vs. Causation: Conclusions

- So looking back at the positive correlation between depression and self-esteem, it could be that depression causes self-esteem to go down, or that low self-esteem results in depression, or that a third variable causes the change in both.
- When looking at a correlation, be sure to recognize that the variables may be related but that it in no way implies that the change in one **causes** the change in the other.[2]

[2] Correlation notes taken from from the following web site: http://allpsych.com/researchmethods/correlation.html

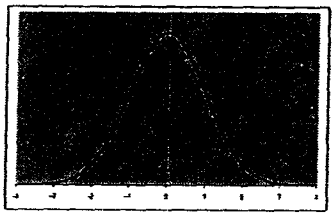
32

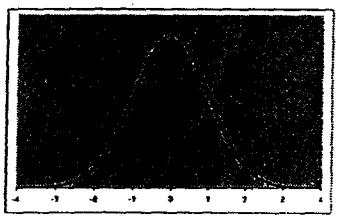
Sources/Additional Resources

- Basic explanation of bell curves: http://allpsych.com/researchmethods/distributions.html
- Understanding Proportions: http://www.utah.edu/stat/bots/game7/Game7.html
- Basic explanation and proportions: http://www1.hollins.edu/faculty/clarkjm/Stat140/normalcurves.htm

33

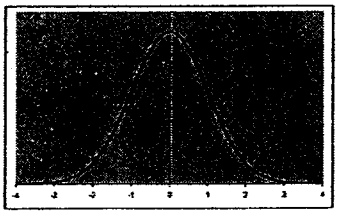
Notes

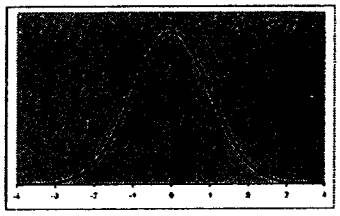




34

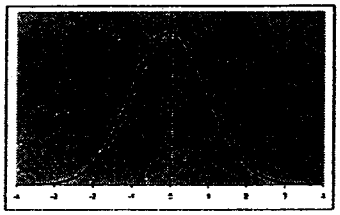
Notes

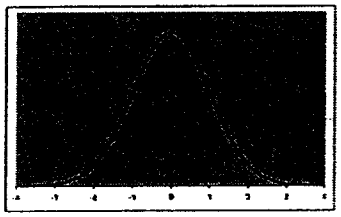




35

Notes





36