Module 4: Z-scores, Normal Curve, and Simple Probability

Jon Starkweather, PhD
jonathan.starkweather@unt.edu
Consultant
Research and **S**tatistical **S**upport

UNT UNIVERSITY OF NORTH TEXAS
Discover the power of ideas.

Introduction to Statistics for the Social Sciences

RSS
Research and Statistical Support

The RSS short courses

The Research and Statistical Support (RSS) office at the University of North Texas hosts a number of "Short Courses". A list of them is available at:

http://www.unt.edu/rss/Instructional.htm

Outline

Outline

RSS
Research and Statistical Support

Outline

RSS
Research and Statistical Support

Outline

Outline

Describing a score in relation to a distribution

Describing a score in relation to a distribution

- Knowing a score give little information without knowing where it is placed in the distribution.

Describing a score in relation to a distribution

- Knowing a score give little information without knowing where it is placed in the distribution.
- Knowing the mean of the distribution allows us to tell if a score is above or below the average.

Describing a score in relation to a distribution

- Knowing a score give little information without knowing where it is placed in the distribution.
- Knowing the mean of the distribution allows us to tell if a score is above or below the average.
- An individual scores 26 on a leadership test.

RSS Research and Statistical Support

Describing a score in relation to a distribution

- Knowing a score give little information without knowing where it is placed in the distribution.
- Knowing the mean of the distribution allows us to tell if a score is above or below the average.
- An individual scores 26 on a leadership test.
    - What does that mean? Are they a good leader or a bad leader?

RSS
Research and Statistical Support

Describing a score in relation to a distribution

- Knowing a score give little information without knowing where it is placed in the distribution.
- Knowing the mean of the distribution allows us to tell if a score is above or below the average.
- An individual scores 26 on a leadership test.
  - What does that mean? Are they a good leader or a bad leader?
  - If the mean of the test is 20, then we know that the person is above average in leadership (compared to others who have taken the test).

RSS
Research and Statistical Support

Example 1

Example 1

- Knowing the standard deviation allows us to tell how much above or below the mean that score is in relations to the spread of the distribution.

Example 1

- Knowing the standard deviation allows us to tell how much above or below the mean that score is in relations to the spread of the distribution.
- Leadership test: $\overline{X} = 20, S = 3$

RSS
Research and Statistical Support

Example 1

- Knowing the standard deviation allows us to tell how much above or below the mean that score is in relations to the spread of the distribution.
- Leadership test: $\overline{X} = 20$, $S = 3$
- The person with the score of 26 is two standard deviations above the mean.

Example 2

Example 2

- Planning Ability

Example 2

- Planning Ability
- A person scores 84 on a test of planning ability.

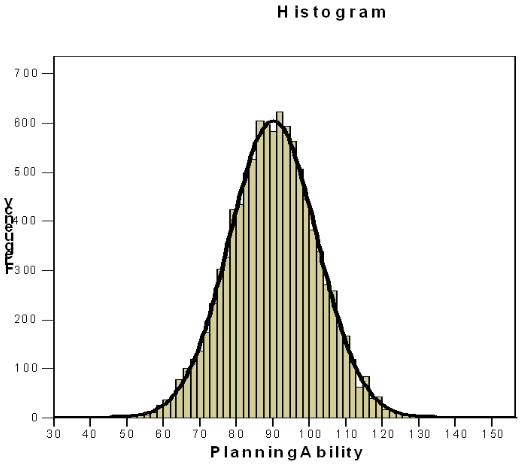
Example 2

- Planning Ability
- A person scores 84 on a test of planning ability.
- If $\overline{X} = 90$, $S = 12$, this person's score is 0.5 standard deviations below the mean.

Example 2

- Planning Ability
- A person scores 84 on a test of planning ability.
- If $\overline{X} = 90$, $S = 12$, this person's score is 0.5 standard deviations below the mean.
- Thus the person is below average, but not by a lot.

Planning ability: $\overline{X} = 90, S = 12$



We can visualize that the score of 84 would be slightly lower than the mean.

Examples 3 and 4

- A person scores 114 on the test of planning ability.

Examples 3 and 4

- A person scores 114 on the test of planning ability.
  - If $\overline{X} = 90$, $S = 12$, this person's score is 2 standard deviations above the mean.

Examples 3 and 4

- A person scores 114 on the test of planning ability.
  - If $\overline{X} = 90$, $S = 12$, this person's score is 2 standard deviations above the mean.
- Thus the person is above average by quite a bit.

RSS Research and Statistical Support

Examples 3 and 4

- A person scores 114 on the test of planning ability.
  - If $\overline{X} = 90, S = 12$, this person's score is 2 standard deviations above the mean.
- Thus the person is above average by quite a bit.
- What if a person scored 78?

RSS
Research and Statistical Support

Examples 3 and 4

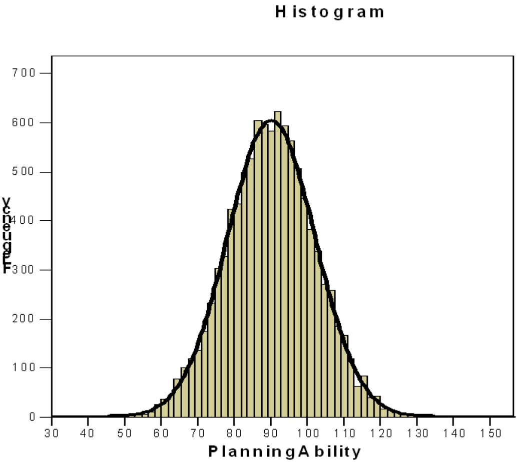
- A person scores 114 on the test of planning ability.
  - If $\overline{X} = 90, S = 12$, this person's score is 2 standard deviations above the mean.
- Thus the person is above average by quite a bit.
- What if a person scored 78?
  - Then the person is below the average by about the average amount that scores deviated from the mean.

Examples 3 and 4

- A person scores 114 on the test of planning ability.
  - If $\overline{X} = 90$, $S = 12$, this person's score is 2 standard deviations above the mean.
- Thus the person is above average by quite a bit.
- What if a person scored 78?
  - Then the person is below the average by about the average amount that scores deviated from the mean.
- Recall, standard deviation *is* the average amount the scores *deviated* from the mean.

RSS
Research and Statistical Support

Planning ability: $\overline{X} = 90, S = 12$



Visualize the score of 114 as far above the mean, while the score of 78 would be slightly below the mean.

Z-score

- Recall from the last module, standard deviation is the average amount of deviation around the mean scores have for a particular distribution.

Z-score

- Recall from the last module, standard deviation is the average amount of deviation around the mean scores have for a particular distribution.
  - So, if $S = 12$, then the average deviation for scores from our Planning Ability test mean is 12.

Z-score

- Recall from the last module, standard deviation is the average amount of deviation around the mean scores have for a particular distribution.
  - So, if $S = 12$, then the average deviation for scores from our Planning Ability test mean is 12.
- The number of standard deviations a particular score is above or below the mean is called its **Z-score**.

RSS
Research and Statistical Support

Z-score

- Recall from the last module, standard deviation is the average amount of deviation around the mean scores have for a particular distribution.
  - So, if $S = 12$, then the average deviation for scores from our Planning Ability test mean is 12.
- The number of standard deviations a particular score is above or below the mean is called its **Z-score**.
  - If a score is below the mean, the Z-score is negative.
  - If a score is above the mean, the Z-score is positive.

RSS
Research and Statistical Support

Z-score

- Recall from the last module, standard deviation is the average amount of deviation around the mean scores have for a particular distribution.
  - So, if $S = 12$, then the average deviation for scores from our Planning Ability test mean is 12.
- The number of standard deviations a particular score is above or below the mean is called its **Z-score**.
  - If a score is below the mean, the Z-score is negative.
  - If a score is above the mean, the Z-score is positive.
- The standard deviation now becomes a kind of yardstick, a unit of measure in its own right.

RSS
Research and Statistical Support

Rulers Analogy

The two scales are something like a ruler with inches lined up on the one side and centimeters on the other.

Rulers Analogy

The two scales are something like a ruler with inches lined up on the one side and centimeters on the other.

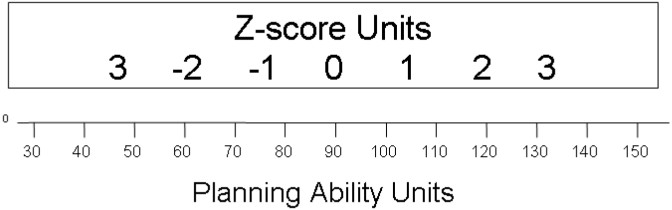
- A **raw** score of 78 is 1 standard deviation below the mean and so, it has a **Z-score** of -1.

RSS
Research and Statistical Support

Rulers Analogy

The two scales are something like a ruler with inches lined up on the one side and centimeters on the other.

- A **raw** score of 78 is 1 standard deviation below the mean and so, it has a **Z-score** of -1.
- **Raw scores**: $\overline{X} = 90$, $S = 12$ and **Z-scores**: $\overline{X} = 0$, $S = 1$.



RSS Research and Statistical Support

Z-scores' usefulness

Z-scores' usefulness

- Z-scores provide a helpful way to compare scores on measures that are on completely different scales.

Z-scores' usefulness

- Z-scores provide a helpful way to compare scores on measures that are on completely different scales.
- If a person scored:

Z-scores' usefulness

- Z-scores provide a helpful way to compare scores on measures that are on completely different scales.
- If a person scored:
  - 26 on Leadership (example 1: $\overline{X} = 20, S = 3$) and
  - 84 on Planning Ability (example 2: $\overline{X} = 90, S = 12$)

RSS
Research and Statistical Support

Z-scores' usefulness

- Z-scores provide a helpful way to compare scores on measures that are on completely different scales.
- If a person scored:
  - 26 on Leadership (example 1: $\overline{X} = 20, S = 3$) and
  - 84 on Planning Ability (example 2: $\overline{X} = 90, S = 12$)
- We can say that person's scores are much higher than average on Leadership and slightly lower than average on Planning Ability.

RSS
Research and Statistical Support

Converting from Raw score to Z-score units.

- Formula for Z-score: $Z = \frac{X - \overline{X}}{S}$

RSS
Research and Statistical Support

Converting from Raw score to Z-score units.

- Formula for Z-score: $Z = \frac{X - \overline{X}}{S}$
- Example 2: $Z = \frac{84 - 90}{12} = \frac{-6}{12} = -0.5$

RSS
Research and Statistical Support

Converting from Raw score to Z-score units.

- Formula for Z-score: $Z = \frac{X - \overline{X}}{S}$
- Example 2: $Z = \frac{84-90}{12} = \frac{-6}{12} = -0.5$
- Example 3: $Z = \frac{114-90}{12} = \frac{24}{12} = 2.0$

RSS
Research and Statistical Support

Converting from Raw score to Z-score units.

- Formula for Z-score: $Z = \frac{X - \overline{X}}{S}$
- Example 2: $Z = \frac{84-90}{12} = \frac{-6}{12} = -0.5$
- Example 3: $Z = \frac{114-90}{12} = \frac{24}{12} = 2.0$
- Example 4: $Z = \frac{78-90}{12} = \frac{-12}{12} = -1.0$

Converting from Z-score to Raw score units.

- Formula for raw score: $(Z)(S) + \overline{X} = X$

Converting from Z-score to Raw score units.

- Formula for raw score: $(Z)(S) + \overline{X} = X$
- Example 2: $(-0.5)(12) + 90 = -6 + 90 = 84$

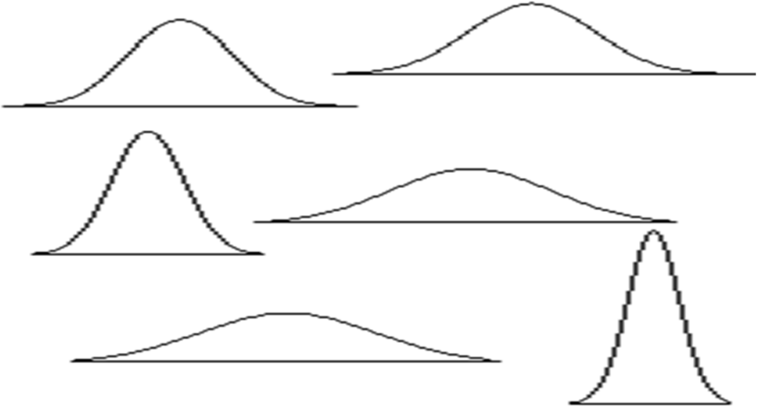
Converting from Z-score to Raw score units.

- Formula for raw score: $(Z)(S) + \overline{X} = X$
- Example 2: $(-0.5)(12) + 90 = -6 + 90 = 84$
- Example 3: $(2)(12) + 90 = 24 + 90 = 114$

Converting from Z-score to Raw score units.

- Formula for raw score: $(Z)(S) + \overline{X} = X$
- Example 2: $(-0.5)(12) + 90 = -6 + 90 = 84$
- Example 3: $(2)(12) + 90 = 24 + 90 = 114$
- Example 4: $(-1)(12) + 90 = -12 + 90 = 78$

Normal Curves



Characteristics of Normal Curves

- Most scores fall near the center, fewer at the extremes.

Characteristics of Normal Curves

- Most scores fall near the center, fewer at the extremes.
- Symmetrical.

Characteristics of Normal Curves

- Most scores fall near the center, fewer at the extremes.
- Symmetrical.
  - Skewness = zero

Characteristics of Normal Curves

- Most scores fall near the center, fewer at the extremes.
- Symmetrical.
  - Skewness = zero
- Mean, Median, Mode are all the same.

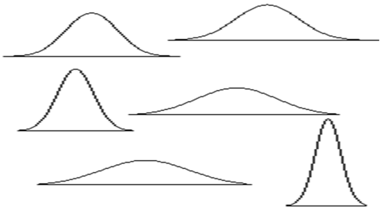
Characteristics of Normal Curves

- Most scores fall near the center, fewer at the extremes.
- Symmetrical.
  - Skewness = zero
- Mean, Median, Mode are all the same.
  - At the center.

RSS
Research and Statistical Support

Characteristics of Normal Curves

- Most scores fall near the center, fewer at the extremes.
- Symmetrical.
  - Skewness = zero
- Mean, Median, Mode are all the same.
  - At the center.
- Dispersion can change.



RSS Research and Statistical Support

Standard Normal Curve

The **Standard** Normal Curve is a special case of the family of normal curves.

Standard Normal Curve

The **Standard** Normal Curve is a special case of the family of normal curves.

- Theoretical distribution (rather than Empirical distribution).

Standard Normal Curve

The **Standard** Normal Curve is a special case of the family of normal curves.

- Theoretical distribution (rather than Empirical distribution).
  - Meaning, it exists in theory and we assume it represents actual variables.

Standard Normal Curve

The **Standard** Normal Curve is a special case of the family of normal curves.

- Theoretical distribution (rather than Empirical distribution).
  - Meaning, it exists in theory and we assume it represents actual variables.
- Precisely defined by:

Standard Normal Curve

The **Standard** Normal Curve is a special case of the family of normal curves.

- Theoretical distribution (rather than Empirical distribution).
  - Meaning, it exists in theory and we assume it represents actual variables.
- Precisely defined by:
  - $\overline{X} = 0$

RSS
Research and Statistical Support

Standard Normal Curve

The **Standard** Normal Curve is a special case of the family of normal curves.

- Theoretical distribution (rather than Empirical distribution).
  - Meaning, it exists in theory and we assume it represents actual variables.
- Precisely defined by:
  - $\overline{X} = 0$
  - $S = 1$

Standard Normal Curve

The **Standard** Normal Curve is a special case of the family of normal curves.

- Theoretical distribution (rather than Empirical distribution).
  - Meaning, it exists in theory and we assume it represents actual variables.
- Precisely defined by:
  - $\overline{X} = 0$
  - $S = 1$
- So, the shape of the Standard Normal Curve is always the same (i.e., standardized).

Standard Normal Curve

- Does not **exactly** correspond to any distribution in nature.

Standard Normal Curve

- Does not **exactly** correspond to any distribution in nature.
  - However, most distributions in nature are assumed to closely resemble a normal curve.

Standard Normal Curve

- Does not **exactly** correspond to any distribution in nature.
  - However, most distributions in nature are assumed to closely resemble a normal curve.
- The scale (x-axis) has no limit above or below the mean (Re: Theoretical Distribution).

Standard Normal Curve

- Does not **exactly** correspond to any distribution in nature.
    - However, most distributions in nature are assumed to closely resemble a normal curve.
- The scale (x-axis) has no limit above or below the mean (Re: Theoretical Distribution).
- It is always a ‘smooth’ curve because as either a histogram or frequency polygon, the ‘intervals’ of the scale are infinitesimal.

RSS
Research and Statistical Support

Why is it so common?

- Each raw score is influenced by many things, each of which is essentially random.

Why is it so common?

- Each raw score is influenced by many things, each of which is essentially random.
- The combination of these random influences is likely to be a middle score in most cases.

Research and Statistical Support

Why is it so common?

- Each raw score is influenced by many things, each of which is essentially random.
- The combination of these random influences is likely to be a middle score in most cases.
  - At or near the mean; unimodal.

Why is it so common?

- Each raw score is influenced by many things, each of which is essentially random.
- The combination of these random influences is likely to be a middle score in most cases.
    - At or near the mean; unimodal.
- When not a middle score, there are equal chances of an imbalance of the random influences being in either direction (+ or -).

Why is it so common?

- Each raw score is influenced by many things, each of which is essentially random.
- The combination of these random influences is likely to be a middle score in most cases.
  - At or near the mean; unimodal.
- When not a middle score, there are equal chances of an imbalance of the random influences being in either direction (+ or -).
  - Symmetry; each extreme score in magnitude is as likely as one at the opposite end of the distribution.

Importance of the normal curve concept.

1. Many of the variables we deal with are commonly assumed to be normally distributed in nature.

Importance of the normal curve concept.

1. Many of the variables we deal with are commonly assumed to be normally distributed in nature.
2. If we can assume a variable is distributed approximately normally, then the techniques we discuss here allow us to make inferences about the phenomena we are studying (i.e., going from our study's sample data to the naturally occurring phenomena in the population using 'traditional' statistical procedures).

RSS
Research and Statistical Support

Importance of the normal curve concept.

1. Many of the variables we deal with are commonly assumed to be normally distributed in nature.
2. If we can assume a variable is distributed approximately normally, then the techniques we discuss here allow us to make inferences about the phenomena we are studying (i.e., going from our study's sample data to the naturally occurring phenomena in the population using 'traditional' statistical procedures).
    - Assuming normality allows us to assign probabilities.

RSS Research and Statistical Support

Suspicion of Normality

- Central Limit Theorem (more on this later).

Suspicion of Normality

- Central Limit Theorem (more on this later).
  - Mathematical Proof that an infinite number of random samples **will** create a normal distribution.

Suspicion of Normality

- Central Limit Theorem (more on this later).
  - Mathematical Proof that an infinite number of random samples **will** create a normal distribution.
- Most traditional inferential statistical techniques have an assumption of normality which must be met in order to have confidence in the conclusions we come to, based on our sample results.

RSS
Research and Statistical Support

Suspicion of Normality

- Central Limit Theorem (more on this later).
  - Mathematical Proof that an infinite number of random samples **will** create a normal distribution.
- Most traditional inferential statistical techniques have an assumption of normality which must be met in order to have confidence in the conclusions we come to, based on our sample results.
  - We must always consider this assumption.

Suspicion of Normality

- Central Limit Theorem (more on this later).
  - Mathematical Proof that an infinite number of random samples **will** create a normal distribution.
- Most traditional inferential statistical techniques have an assumption of normality which must be met in order to have confidence in the conclusions we come to, based on our sample results.
  - We must always consider this assumption.
  - Often we can test this assumption with objective criteria.

Suspicion of Normality

- Central Limit Theorem (more on this later).
  - Mathematical Proof that an infinite number of random samples **will** create a normal distribution.
- Most traditional inferential statistical techniques have an assumption of normality which must be met in order to have confidence in the conclusions we come to, based on our sample results.
  - We must always consider this assumption.
  - Often we can test this assumption with objective criteria.
  - If we find the assumption violated, there are often modern (robust) analysis alternatives.

RSS
Research and Statistical Support

Central Limit Theorem

- If we draw an infinite number of repeated random samples (each with replacement), then we will end up with a normal distribution of sample means.

Central Limit Theorem

- If we draw an infinite number of repeated random samples (each with replacement), then we will end up with a normal distribution of sample means.
- The mean of that distribution (of means) will be **the** mean of the population.

http://onlinestatbook.com/simulations/CLT/clt.html

Research and Statistical Support

Why some distributions are not near normal.

- Limitations on minimum or maximum score possible.

Why some distributions are not near normal.

- Limitations on minimum or maximum score possible.
  - **Ceiling Effects** such as percentage correct (max. = 100).

Why some distributions are not near normal.

- Limitations on minimum or maximum score possible.
  - **Ceiling Effects** such as percentage correct (max. = 100).
  - **Floor Effects** such as SAT scores of college students (only those with decent to high scores are *in* college).

RSS
Research and Statistical Support

Why some distributions are not near normal.

- Limitations on minimum or maximum score possible.
  - **Ceiling Effects** such as percentage correct (max. = 100).
  - **Floor Effects** such as SAT scores of college students (only those with decent to high scores are *in* college).
- Limited number of categories.

Why some distributions are not near normal.

- Limitations on minimum or maximum score possible.
  - **Ceiling Effects** such as percentage correct (max. = 100).
  - **Floor Effects** such as SAT scores of college students (only those with decent to high scores are *in* college).
- Limited number of categories.
  - Non-continuous variable; number of children in a family.

RSS
Research and Statistical Support

Why some distributions are not near normal.

- Limitations on minimum or maximum score possible.
  - **Ceiling Effects** such as percentage correct (max. = 100).
  - **Floor Effects** such as SAT scores of college students (only those with decent to high scores are *in* college).
- Limited number of categories.
  - Non-continuous variable; number of children in a family.
- Growth or time effects: children's height...age.

Research and Statistical Support

Why some distributions are not near normal.

- Limitations on minimum or maximum score possible.
  - **Ceiling Effects** such as percentage correct (max. = 100).
  - **Floor Effects** such as SAT scores of college students (only those with decent to high scores are *in* college).
- Limited number of categories.
  - Non-continuous variable; number of children in a family.
- Growth or time effects: children's height...age.
  - Some things are in constant growth or recession.

Why some distributions are not near normal.

- Limitations on minimum or maximum score possible.
  - **Ceiling Effects** such as percentage correct (max. = 100).
  - **Floor Effects** such as SAT scores of college students (only those with decent to high scores are *in* college).
- Limited number of categories.
  - Non-continuous variable; number of children in a family.
- Growth or time effects: children's height...age.
  - Some things are in constant growth or recession.
- Direct random effects: roulette wheel.

RSS
Research and Statistical Support

Why some distributions are not near normal.

- Limitations on minimum or maximum score possible.
  - **Ceiling Effects** such as percentage correct (max. = 100).
  - **Floor Effects** such as SAT scores of college students (only those with decent to high scores are *in* college).
- Limited number of categories.
  - Non-continuous variable; number of children in a family.
- Growth or time effects: children's height...age.
  - Some things are in constant growth or recession.
- Direct random effects: roulette wheel.
  - This has to do with probability, more on this later.

RSS
Research and Statistical Support

Normal Curve & Z-scores

- Because the normal curve is *exactly* defined...

Normal Curve & Z-scores

- Because the normal curve is *exactly* defined...
- If a distribution is normal then there is an exact relation between any of its Z-scores and the percent of cases/scores above and below it.

Normal Curve & Z-scores

- Because the normal curve is *exactly* defined...
- If a distribution is normal then there is an exact relation between any of its Z-scores and the percent of cases/scores above and below it.
  - If you know a particular Z-score, you can determine (often with a table) the percentage of scores above or below it.

Normal Curve & Z-scores

- Because the normal curve is *exactly* defined...
- If a distribution is normal then there is an exact relation between any of its Z-scores and the percent of cases/scores above and below it.
    - If you know a particular Z-score, you can determine (often with a table) the percentage of scores above or below it.
    - If you know the percentages above and below, you can determine the Z-score.

RSS
Research and Statistical Support

Perfectly Symmetrical

- When dealing with the standard normal curve:

Perfectly Symmetrical

- When dealing with the standard normal curve:
  - If a person has a Z-score of 0 (right at the mean), then we know 50% of the scores are above and 50% of the scores are below that point.

Perfectly Symmetrical

- When dealing with the standard normal curve:
  - If a person has a Z-score of 0 (right at the mean), then we know 50% of the scores are above and 50% of the scores are below that point.
  - If a person has 50% of the scores above and 50% of the scores below, then we know that they have a Z-score of 0.

RSS
Research and Statistical Support

Perfectly Symmetrical

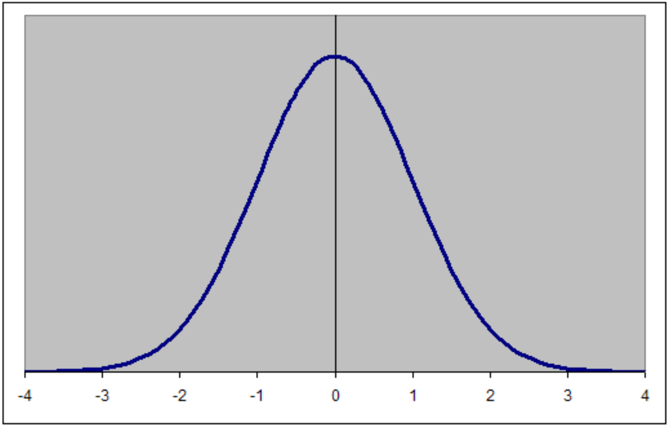
- When dealing with the standard normal curve:
  - If a person has a Z-score of 0 (right at the mean), then we know 50% of the scores are above and 50% of the scores are below that point.
  - If a person has 50% of the scores above and 50% of the scores below, then we know that they have a Z-score of 0.
- The score itself has no *area*, it is simply a boundary between, a marker if you prefer.

RSS
Research and Statistical Support

Perfectly Symmetrical

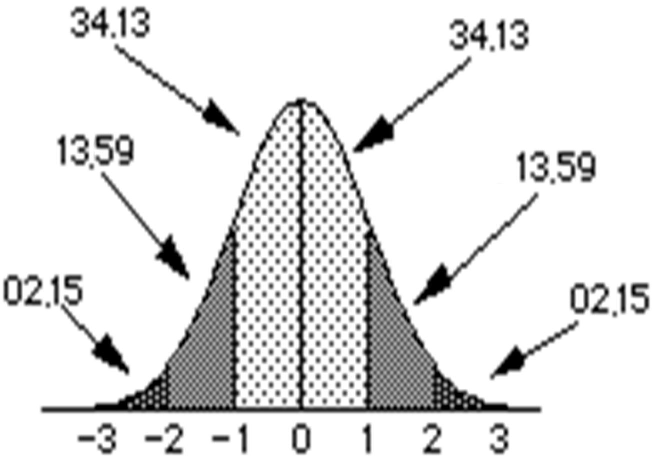
- When dealing with the standard normal curve:
  - If a person has a Z-score of 0 (right at the mean), then we know 50% of the scores are above and 50% of the scores are below that point.
  - If a person has 50% of the scores above and 50% of the scores below, then we know that they have a Z-score of 0.
- The score itself has no *area*, it is simply a boundary between, a marker if you prefer.
  - Like in geometry, point and line are identifiers.

RSS
Research and Statistical Support



Z-scores along the bottom, the percentage of scores between each can be figured mathematically but, we use tables instead:
http://www.mathsisfun.com/data/standard-normal-distribution-table.html

RSS Research and Statistical Support



Z-scores along the bottom, percentages around the top

Memorize this image...burn it into your brain!!

RSS
Research and Statistical Support

Percentages and Probabilities.

- The corresponding percentages from the previous slide can be expressed as probabilities.

Percentages and Probabilities.

- The corresponding percentages from the previous slide can be expressed as probabilities.
  - 34.13% = 34.13/100 = .3413
  - 13.59% = 13.59/100 = .1359
  - 02.15% = 02.15/100 = .0215

RSS
Research and Statistical Support

Percentages and Probabilities.

- The corresponding percentages from the previous slide can be expressed as probabilities.
  - 34.13% = 34.13/100 = .3413
  - 13.59% = 13.59/100 = .1359
  - 02.15% = 02.15/100 = .0215
- The proportion of scores between any two Z-scores is the same as the probability of selecting another score between those two scores.

Percentages and Probabilities.

- The corresponding percentages from the previous slide can be expressed as probabilities.
  - 34.13% = 34.13/100 = .3413
  - 13.59% = 13.59/100 = .1359
  - 02.15% = 02.15/100 = .0215
- The proportion of scores between any two Z-scores is the same as the probability of selecting another score between those two scores.
- Keep in mind, the complete table of z-scores and their associated probabilities is available at multiple sites online.

http://www.mathsisfun.com/data/standard-normal-distribution-table.html

The *BOTTOM LINE!*

Knowing the formula for converting a raw score to a z-score and knowing the percentages of the standard normal curve (or at least having access to a table):

RSS
Research and Statistical Support

The *BOTTOM LINE!*

Knowing the formula for converting a raw score to a z-score and knowing the percentages of the standard normal curve (or at least having access to a table):

- We can calculate the percentage of scores between any two points of the normal distribution.

The *BOTTOM LINE!*

Knowing the formula for converting a raw score to a z-score and knowing the percentages of the standard normal curve (or at least having access to a table):

- We can calculate the percentage of scores between any two points of the normal distribution.
- For instance, between a raw score and the mean or between any two Z-scores.

More on Percentages.

- When figuring percentage of scores between any two z-scores or between the mean and a Z-score, pay close attention to the column in the table you need.

Z	% Mean to Z	% in Tail
.00	.00	50.00
.01	.40	49.60
.	.	.
.	.	.
.	.	.
.50	19.15	30.85
1.00	34.13	15.87

http://davidmlane.com/hyperstat/z_table.html

RSS Research and Statistical Support

Percentiles = Directional Percentages

- Remember, when finding a percentile rank, the area under the normal curve becomes ranked (from 1 to 100 percentiles).

Percentiles = Directional Percentages

- Remember, when finding a percentile rank, the area under the normal curve becomes ranked (from 1 to 100 percentiles).
- For instance, we know that a Z-score of +2 represents 2 standard deviations above the mean. This same location, when converted to a percentile would be the 98th percentile.

Percentiles = Directional Percentages

- Remember, when finding a percentile rank, the area under the normal curve becomes ranked (from 1 to 100 percentiles).
- For instance, we know that a Z-score of +2 represents 2 standard deviations above the mean. This same location, when converted to a percentile would be the 98th percentile.
  - 50% of the scores are below the mean, (+) 34% of the scores between the mean and a z-score of +1, (+) the other 14% of scores between a Z-score of +1 and a Z-score of +2 = 98% of the scores, or the 98th percentile.

Research and Statistical Support

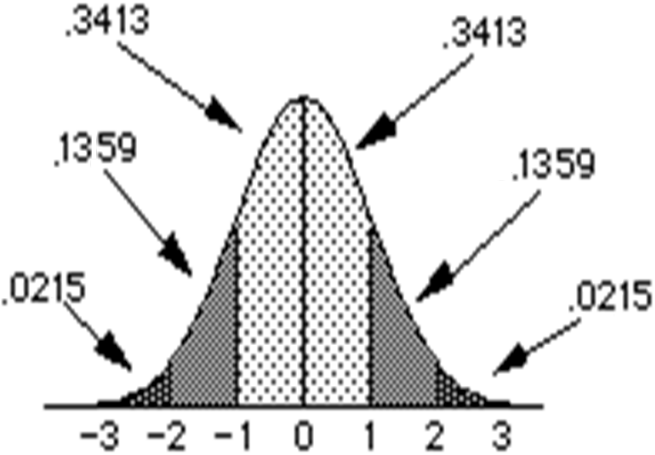
Percentiles

- Likewise, a Z-score of -1 which is one standard deviation below the mean would be expressed as the 16th percentile.

Percentiles

- Likewise, a Z-score of -1 which is one standard deviation below the mean would be expressed as the 16th percentile.
  - 2% of the scores are beyond 2 standard deviations below the mean, (+) 14% of the scores between 2 standard deviations below the mean and 1 standard deviation below the mean = 16% of the scores are below our Z-score of -1; a raw score with the Z-score of -1 is the 16th percentile.

RSS
Research and Statistical Support



Z-scores along the bottom, probabilities around the top

Memorize this image...burn it into your brain!!

RSS
Research and Statistical Support

Percentages & simple Probabilities

- The numbers on the outside of the standard normal curve from the previous slide can be expressed as percentages or probabilities.

Percentages & simple Probabilities

- The numbers on the outside of the standard normal curve from the previous slide can be expressed as percentages or probabilities.
  34% = .3413
  14% = .1359
  02% = .0215

Percentages & simple Probabilities

- The numbers on the outside of the standard normal curve from the previous slide can be expressed as percentages or probabilities.
  34% = .3413
  14% = .1359
  02% = .0215
- The proportion of scores between any two Z-scores is the same as the probability of selecting a case between those two Z-scores.

More on simple Probability

- The range of probabilities: Zero to One.

More on simple Probability

- The range of probabilities: Zero to One.
  - Range = $0 \rightarrow 1$

More on simple Probability

- The range of probabilities: Zero to One.
  - Range = $0 \rightarrow 1$
- Expressed as an italicized, lower-case *p*

More on simple Probability

- The range of probabilities: Zero to One.
  - Range = $0 \rightarrow 1$
- Expressed as an italicized, lower-case *p*
  - If there is no chance the event will occur, then $p = 0$

RSS
Research and Statistical Support

More on simple Probability

- The range of probabilities: Zero to One.
  - Range = $0 \rightarrow 1$
- Expressed as an italicized, lower-case *p*
  - If there is no chance the event will occur, then $p = 0$
  - If the event is certain to occur, then $p = 1$

More on simple Probability

- The range of probabilities: Zero to One.
  - Range = $0 \rightarrow 1$
- Expressed as an italicized, lower-case *p*
  - If there is no chance the event will occur, then $p = 0$
  - If the event is certain to occur, then $p = 1$
  - $p < .02$ means the probability of the event is less than .02

More on simple Probability

- The range of probabilities: Zero to One.
  - Range = $0 \rightarrow 1$
- Expressed as an italicized, lower-case *p*
  - If there is no chance the event will occur, then $p = 0$
  - If the event is certain to occur, then $p = 1$
  - $p < .02$ means the probability of the event is less than .02
- Calculating simple probability.

RSS
Research and Statistical Support

More on simple Probability

- The range of probabilities: Zero to One.
  - Range = $0 \rightarrow 1$
- Expressed as an italicized, lower-case *p*
  - If there is no chance the event will occur, then $p = 0$
  - If the event is certain to occur, then $p = 1$
  - $p < .02$ means the probability of the event is less than .02
- Calculating simple probability.
  - The number of possible successful or desirable outcomes divided by the number of **all** possible outcomes.

RSS
Research and Statistical Support

Interpretation of Probability

- Long-run relative-frequency interpretation.

Interpretation of Probability

- Long-run relative-frequency interpretation.
  - What you would expect to get, in the long run, if you were to repeat the experiment many times.

Interpretation of Probability

- Long-run relative-frequency interpretation.
  - What you would expect to get, in the long run, if you were to repeat the experiment many times.
  - The probability of rolling a 5 on a six-sided die is: $p = .166$, so if we roll the die 1000 times, we would expect to get a 5 roughly 166.6 times.

Interpretation of Probability

- Long-run relative-frequency interpretation.
    - What you would expect to get, in the long run, if you were to repeat the experiment many times.
    - The probability of rolling a 5 on a six-sided die is: $p = .166$, so if we roll the die 1000 times, we would expect to get a 5 roughly 166.6 times.
        - This is the interpretation we typically use.

RSS
Research and Statistical Support

Interpretation of Probability

- Long-run relative-frequency interpretation.
  - What you would expect to get, in the long run, if you were to repeat the experiment many times.
  - The probability of rolling a 5 on a six-sided die is: $p = .166$, so if we roll the die 1000 times, we would expect to get a 5 roughly 166.6 times.
    - This is the interpretation we typically use.
- Subjective interpretation of probability.

RSS
Research and Statistical Support

Interpretation of Probability

- Long-run relative-frequency interpretation.
  - What you would expect to get, in the long run, if you were to repeat the experiment many times.
  - The probability of rolling a 5 on a six-sided die is: $p = .166$, so if we roll the die 1000 times, we would expect to get a 5 roughly 166.6 times.
    - This is the interpretation we typically use.
- Subjective interpretation of probability.
  - How certain one is that a particular event will happen.

Interpretation of Probability

- Long-run relative-frequency interpretation.
  - What you would expect to get, in the long run, if you were to repeat the experiment many times.
  - The probability of rolling a 5 on a six-sided die is: $p = .166$, so if we roll the die 1000 times, we would expect to get a 5 roughly 166.6 times.
    - This is the interpretation we typically use.
- Subjective interpretation of probability.
  - How certain one is that a particular event will happen.
    - There is a 16.66% chance I'll roll a 5 on one independent roll of the die.

RSS
Research and Statistical Support

Example simple Probabilities

Remember, simple probability is the number of desired outcomes divided by the number of possible outcomes.

RSS
Research and Statistical Support

Example simple Probabilities

Remember, simple probability is the number of desired outcomes divided by the number of possible outcomes.

- Probability of getting a head or a tail on one toss of the coin?

Example simple Probabilities

Remember, simple probability is the number of desired outcomes divided by the number of possible outcomes.

- Probability of getting a head or a tail on one toss of the coin?
  - $p = 1$

RSS
Research and Statistical Support

Example simple Probabilities

Remember, simple probability is the number of desired outcomes divided by the number of possible outcomes.

- Probability of getting a head or a tail on one toss of the coin?
    - $p = 1$
- Probability of getting a head on one toss of the coin?

Example simple Probabilities

Remember, simple probability is the number of desired outcomes divided by the number of possible outcomes.

- Probability of getting a head or a tail on one toss of the coin?
  - $p = 1$
- Probability of getting a head on one toss of the coin?
  - $p = .50$

Example simple Probabilities

Remember, simple probability is the number of desired outcomes divided by the number of possible outcomes.

- Probability of getting a head or a tail on one toss of the coin?
  - $p = 1$
- Probability of getting a head on one toss of the coin?
  - $p = .50$
- Probability of NOT getting a head or a tail on one toss of the coin?

RSS
Research and Statistical Support

Example simple Probabilities

Remember, simple probability is the number of desired outcomes divided by the number of possible outcomes.

- Probability of getting a head or a tail on one toss of the coin?
  - $p = 1$
- Probability of getting a head on one toss of the coin?
  - $p = .50$
- Probability of NOT getting a head or a tail on one toss of the coin?
  - $p = 0$

Example simple Probabilities

Remember, simple probability is the number of desired outcomes divided by the number of possible outcomes.

- Probability of getting a head or a tail on one toss of the coin?
  - $p = 1$
- Probability of getting a head on one toss of the coin?
  - $p = .50$
- Probability of NOT getting a head or a tail on one toss of the coin?
  - $p = 0$
- Probability of getting a head on one toss of the coin after 2500 tosses of the coin?

RSS Research and Statistical Support

Example simple Probabilities

Remember, simple probability is the number of desired outcomes divided by the number of possible outcomes.

- Probability of getting a head or a tail on one toss of the coin?
    - $p = 1$
- Probability of getting a head on one toss of the coin?
    - $p = .50$
- Probability of NOT getting a head or a tail on one toss of the coin?
    - $p = 0$
- Probability of getting a head on one toss of the coin after 2500 tosses of the coin?
    - $p = .50$
    - Remember the old saying: "The coin has no memory."

RSS
Research and Statistical Support

Practice Examples

- The ‘Bear Cubs Problem’
- The ‘Balls of Two Colors’ problem

Available toward the bottom of the page at:

http://www.cut-the-knot.org/probability.shtml

Summary of Module 4

Module 4 covered the following topics:

- Describing the position of a score in a distribution.

RSS
Research and Statistical Support

Summary of Module 4

Module 4 covered the following topics:

- Describing the position of a score in a distribution.
- Z-scores; conversion from raw to Z and Z to raw scores.

RSS Research and Statistical Support

Summary of Module 4

Module 4 covered the following topics:

- Describing the position of a score in a distribution.
- Z-scores; conversion from raw to Z and Z to raw scores.
- Normal distributions and The Standard Normal Distribution.

Summary of Module 4

Module 4 covered the following topics:

- Describing the position of a score in a distribution.
- Z-scores; conversion from raw to Z and Z to raw scores.
- Normal distributions and The Standard Normal Distribution.
- Percentages and simple Probabilities.

Summary of Module 4

Module 4 covered the following topics:

- Describing the position of a score in a distribution.
- Z-scores; conversion from raw to Z and Z to raw scores.
- Normal distributions and The Standard Normal Distribution.
- Percentages and simple Probabilities.

All of these topics will be re-visited consistently in future modules.

This concludes Module 4

Next time Module 5.

- Next time we'll begin covering Hypothesis Testing.
- Until next time; have a nice day.

These slides initially created on: October 4, 2010
These slides last updated on: October 13, 2010

- The bottom date shown is the date this Adobe.pdf file was created; LaTeX[1] has a command for automatically inserting the date of a document's creation.

RSS
Research and Statistical Support

[1] This document was created in LaTeX using the Beamer package