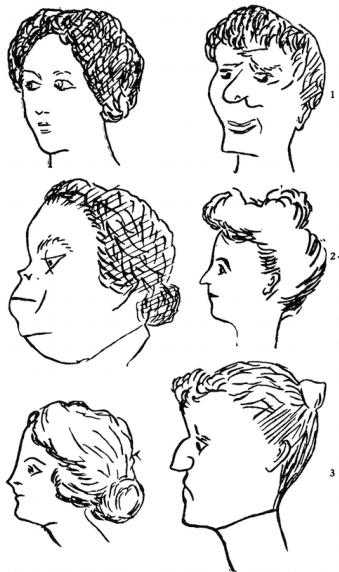


# E-411-PRMA

## Lecture 1

Christopher David Desjardins

17 August 2015



THE PSYCHOLOGICAL CLINIC is indebted for the loan of these cuts and those on p. 225 to the courtesy of Dr. Oliver F. Corman, Associate Superintendent of Schools of Philadelphia, and Chairman of Committee on Backward Children Investigation. See Report of Committee, Dec. 31, 1910, appendix.



*Extreme desespoir.  
Eufferfte Verzweiffung.*



*Colere meslée de Crainte.  
Zorn mit Furcht vermischet.*



## ► Topics

- Statistics, Classical Test Theory, Reliability, Validity, Item Response Theory, Generalizability Theory, Equating, and assessments/issues specific to various fields

## ► Assessments

- R computer assignments (30%)
- Item writing activity (5%)
- Midterm exam (25%)
- Final exam (50%)

The screenshot displays the RStudio environment with the following components:

- Script Editor (Top Left):** Contains R code for a script named `26_june_intro_R.R`. The code includes comments and commands like `precol`, `precol[20,2:3]`, `precol[19:21, c(1,3,6)]`, `str(precol)`, and `attr( "variable.labels" )`.
- Console (Bottom Left):** Shows the R startup message, copyright information, and a disclaimer: "R is free software and comes with ABSOLUTELY NO WARRANTY. You are welcome to redistribute it under certain conditions. Type 'license()' or 'licence()' for distribution details." It also mentions natural language support and collaborative project details.
- Environment (Top Right):** Displays the "Global Environment" and states "Environment is empty".
- Files, Plots, Packages, Help, Viewer (Middle Right):** The "Help" pane is active, showing the documentation for the `separate()` function from the `tidyr` package.

**separate (tidyr) R Documentation**

### Separate one column into multiple columns.

**Description**

Given either regular expression or a vector of character positions, `separate()` turns a single character column into multiple columns.

**Usage**

```
separate(data, col, into, sep = "[^a-zA-Z:]+", remove = TRUE,
  convert = FALSE, extra = "error", ...)
```

**Arguments**

- `data`: A data frame.
- `col`: Bare column name.
- `into`: Names of new variables to create as character vector.
- `sep`: Separator between columns.

R: <https://www.r-project.org>  
 RStudio: <https://www.rstudio.com>

# Why should I learn R?

- ▶ It's free and open-source

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses



# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill
- ▶ High quality graphics

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill
- ▶ High quality graphics
- ▶ Everyone is doing it

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill
- ▶ High quality graphics
- ▶ Everyone is doing it
- ▶ Steep learning curve

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill
- ▶ High quality graphics
- ▶ Everyone is doing it
- ▶ Steep learning curve
  - ▶ Will provide nearly all the code

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill
- ▶ High quality graphics
- ▶ Everyone is doing it
- ▶ Steep learning curve
  - ▶ Will provide nearly all the code
- ▶ No SPSS in this class



# Resources for R

- ▶ Icelandic resources

  - <http://kennslubanki.hi.is/search/efni/r>

  - <http://kennslubanki.hi.is/tolfraedi/myndbond/rrstudio-inngangur>

  - <http://kennslubanki.hi.is/tolfraedi/myndbond/rrstudio-fyrstu-skrefin>

- ▶ Please watch the last two videos before next class
- ▶ Please install R and RStudio before next class
- ▶ Next class will be an R workshop

What is **measurement**?

What is **measurement**?

What is a **test**?

What is **measurement**?

What is a **test**?

What is a **scale**?

What is **measurement**?

Assignment of numerical values based on a set of rules

What is a **test**?

What is a **scale**?

What is **measurement**?

Assignment of numerical values based on a set of rules

What is a **test**?

An instrument used to measure

What is a **scale**?

What is **measurement**?

Assignment of numerical values based on a set of rules

What is a **test**?

An instrument used to measure

What is a **scale**?

A set of numbers used to categorize or quantify variables  
("things")

What is **measurement**?

Assignment of numerical values based on a set of rules

What is a **test**?

An instrument used to measure

What is a **scale**?

A set of numbers used to categorize or quantify variables  
("things")

**Nominal**

**Ordinal**

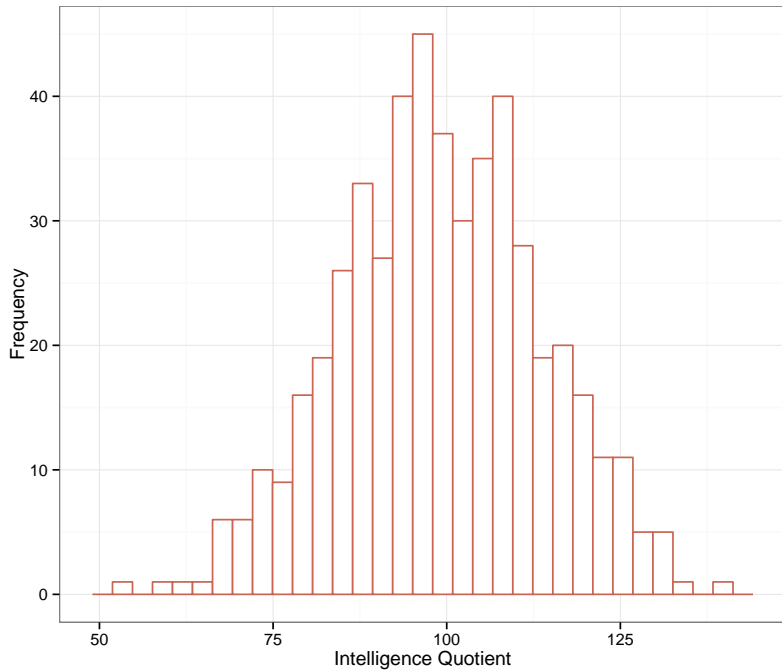
**Ratio**

**Interval**



# What kind of scales are these?

- ▶ Temperature
- ▶ Height
- ▶ Intelligence Quotient
- ▶ Color
- ▶ Ethnic group
- ▶ Likert-type items
- ▶ Job satisfaction

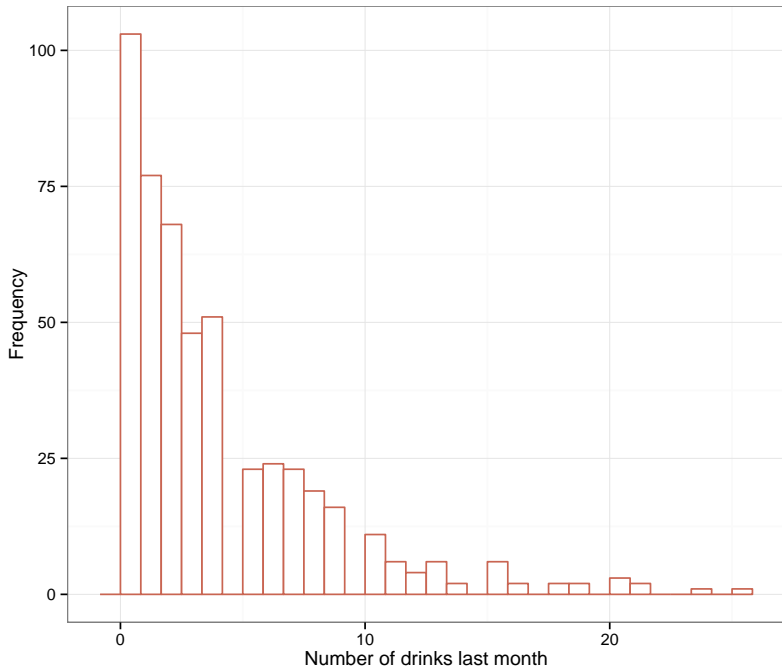


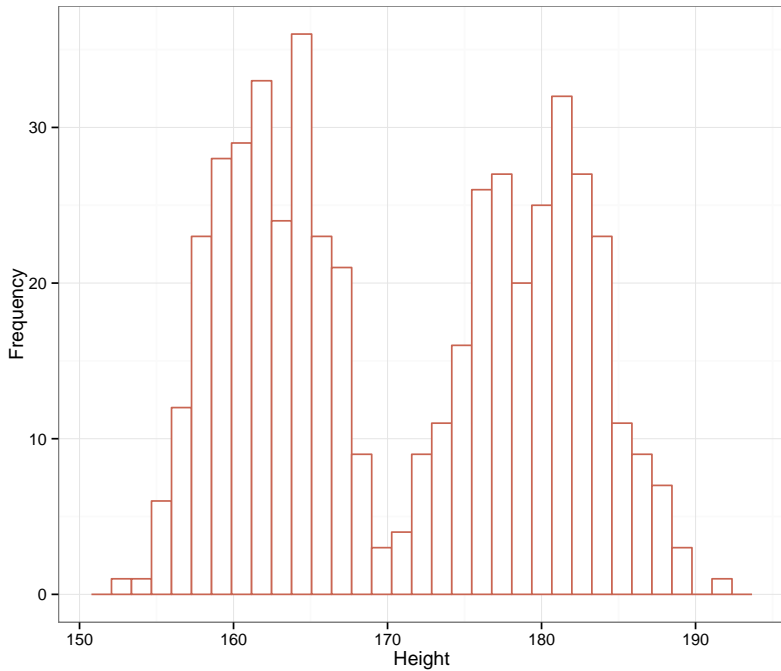
```
# Load the library
set.seed(101)
library("ggplot2")

# Set up the parameters
sample_size <- 500
mean <- 100
standard_deviation <- 15

# Generate random numbers
x <- rnorm(sample_size, mean, standard_deviation)

# Plot the data
qplot(x, fill = I("white"), color = I("#c96552")) +
  theme_bw() + xlab("Intelligence Quotient") +
  ylab("Frequency")
```





# Central Tendency Measures

## Mean

$$\bar{X} = \frac{\sum X_i}{n}$$

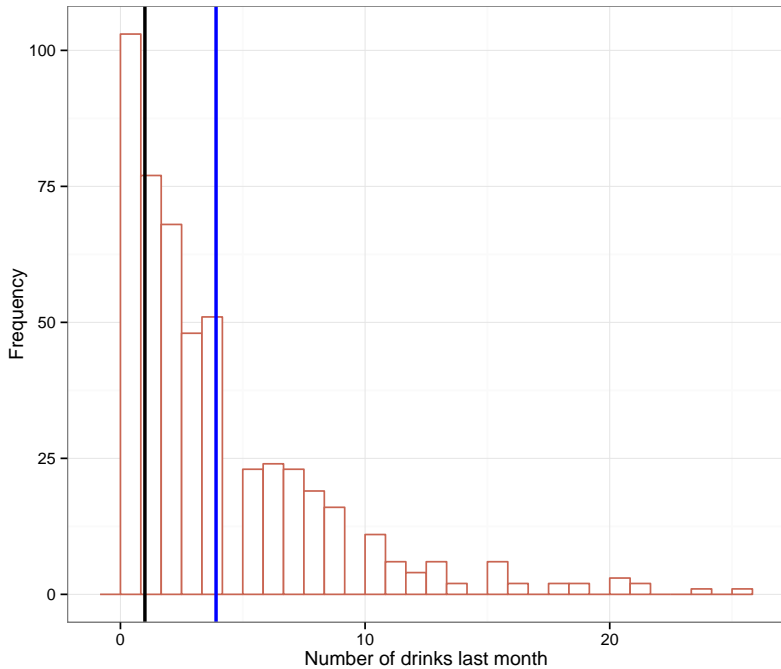
## Median

$$P(X \leq m) \geq \frac{1}{2} \text{ and } P(X \geq m) \geq \frac{1}{2}$$

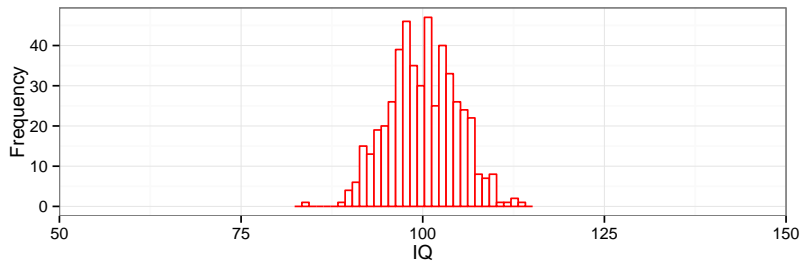
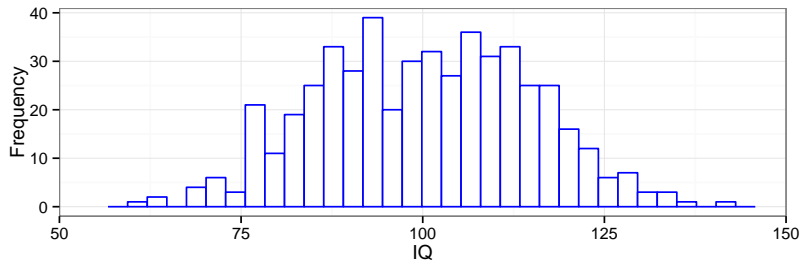
## Mode

The most frequently occurring value

Which of these statistics is most robust to outliers?



# Variability





# Measures of variability

Range

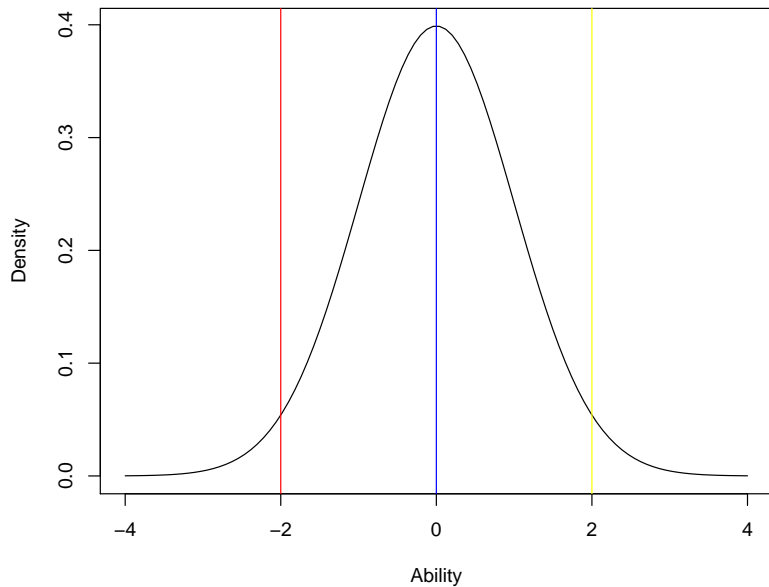
Interquartile range ( $Q_1$ ,  $Q_2$ ,  $Q_3$ )

Standard Deviation and Variance

$$s = \sqrt{\frac{\sum X_i - \bar{X}}{n - 1}}$$

$$s^2 = \frac{\sum X_i - \bar{X}}{n - 1}$$

## Normal Distribution

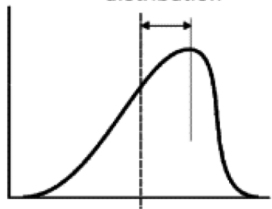


# Distributions, skewness, kurtosis

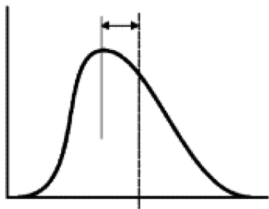
- ▶ What is a probability distribution
  - ▶ Assigns a probability, likeliness of occurrence, of a score of all possible scores
  - ▶ May be parametric or non-parametric
- ▶ What skew might you expect these outcomes to look like?
  - ▶ Reaction time in a psychological experiment
  - ▶ Number of children in a family
  - ▶ Scores on an easy test
  - ▶ Height in Iceland
- ▶ Platykurtic, mesokurtic, and leptokurtic
- ▶ Plot your data, rely less on statistics!

# Shapes of distributions

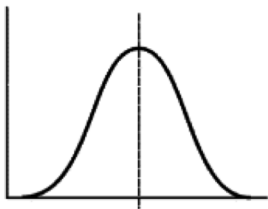
**a** Negatively skewed distribution



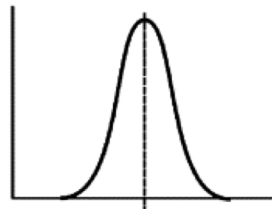
Positively skewed distribution



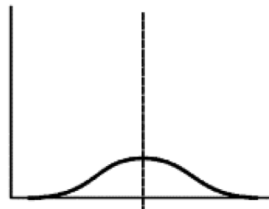
Normal distribution



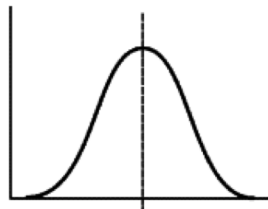
**b** Leptokurtic distribution



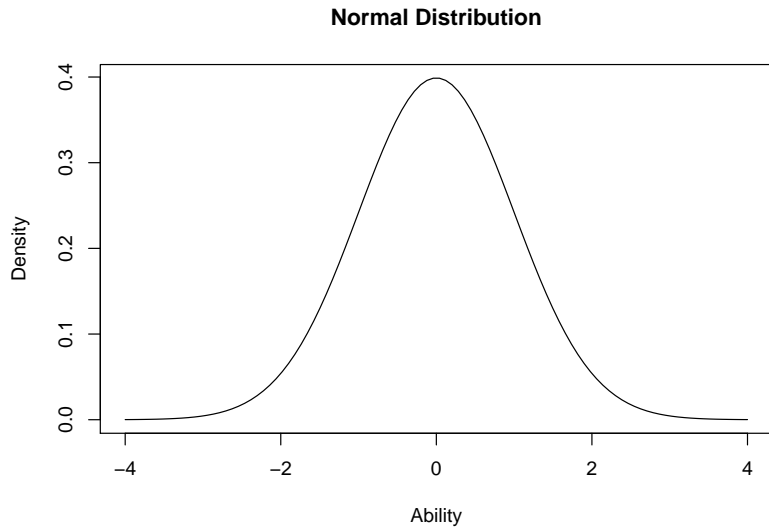
Platykurtic distribution



Normal distribution



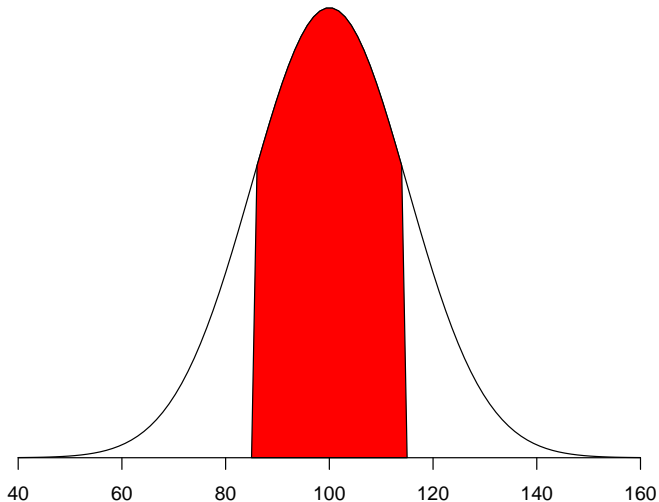
# Normal Distribution



# IQ - 1 Standard Deviation

## Normal Distribution

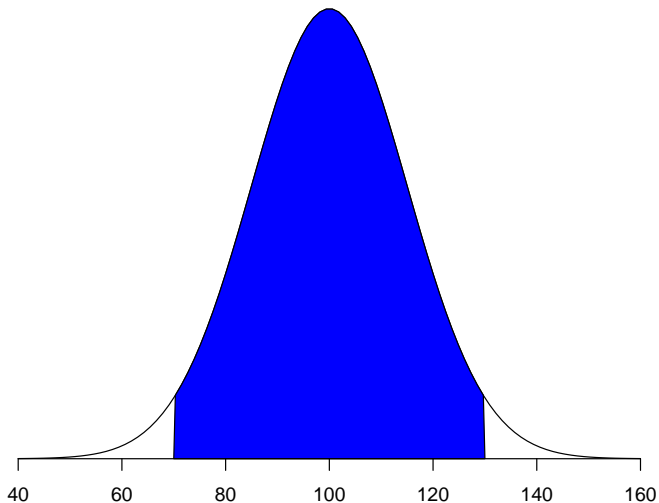
$$P(85 < IQ < 115) = 0.683$$



# IQ - 2 Standard Deviation

## Normal Distribution

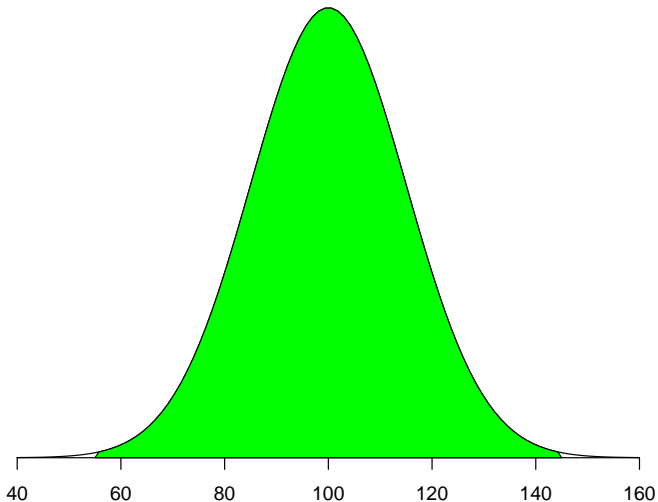
$$P(70 < IQ < 130) = 0.954$$



# IQ - 3 Standard Deviation

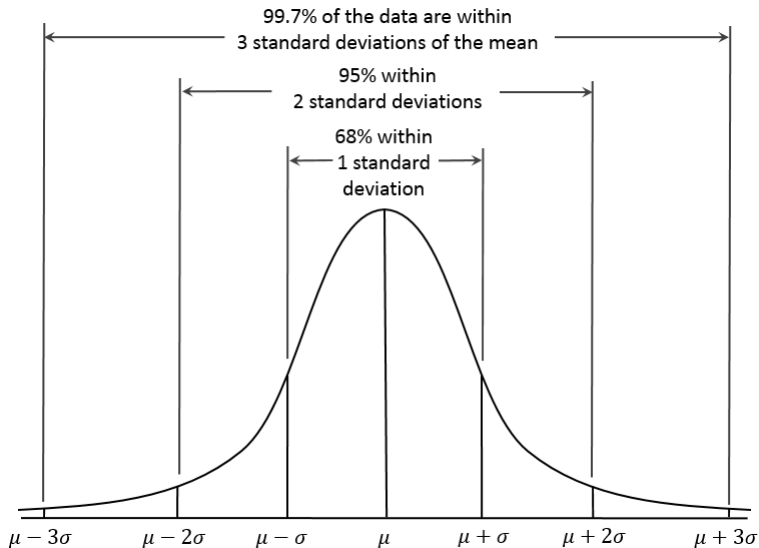
## Normal Distribution

$$P(55 < IQ < 145) = 0.997$$





# Characteristics of the Normal distribution



# Z-distribution

- ▶ The standard normal distribution is known as the z-distribution

# Z-distribution

- ▶ The standard normal distribution is known as the z-distribution
- ▶ What is the mean and the standard deviation of this distribution?

# Z-distribution

- ▶ The standard normal distribution is known as the z-distribution
- ▶ What is the mean and the standard deviation of this distribution?
- ▶ All normal distributions can be converted to the z-distribution

# Z-distribution

- ▶ The standard normal distribution is known as the z-distribution
- ▶ What is the mean and the standard deviation of this distribution?
- ▶ All normal distributions can be converted to the z-distribution
- ▶ A raw score can be converted to a z-score.

# Z-distribution

- ▶ The standard normal distribution is known as the z-distribution
- ▶ What is the mean and the standard deviation of this distribution?
- ▶ All normal distributions can be converted to the z-distribution
- ▶ A raw score can be converted to a z-score.

$$z = \frac{x - \mu}{\sigma}$$

# SAT

The SAT is an aptitude test that high schools take. It is one of the criteria that is used in a college's decision to admit a student. It is composed of a math and a verbal section. Each has a mean of 500 and a standard deviation of 110 and is normally distributed.

- ▶ What are the scores on the test that corresponds to 3, 2, 1, 0, -1, -2, -3 standard deviations?

# SAT

The SAT is an aptitude test that high schools take. It is one of the criteria that is used in a college's decision to admit a student. It is composed of a math and a verbal section. Each has a mean of 500 and a standard deviation of 110 and is normally distributed.

- ▶ What are the scores on the test that corresponds to 3, 2, 1, 0, -1, -2, -3 standard deviations?
- ▶ Assume 1000 people took the SAT,



# SAT

The SAT is an aptitude test that high schools take. It is one of the criteria that is used in a college's decision to admit a student. It is composed of a math and a verbal section. Each has a mean of 500 and a standard deviation of 110 and is normally distributed.

- ▶ What are the scores on the test that corresponds to 3, 2, 1, 0, -1, -2, -3 standard deviations?
- ▶ Assume 1000 people took the SAT,
  - ▶ If Jon got a 700 on the math section, how many people scored above him?

# SAT

The SAT is an aptitude test that high schools take. It is one of the criteria that is used in a college's decision to admit a student. It is composed of a math and a verbal section. Each has a mean of 500 and a standard deviation of 110 and is normally distributed.

- ▶ What are the scores on the test that corresponds to 3, 2, 1, 0, -1, -2, -3 standard deviations?
- ▶ Assume 1000 people took the SAT,
  - ▶ If Jon got a 700 on the math section, how many people scored above him?
  - ▶ If 300 people scored below Anna on the verbal section, what was Anna's score?

# SAT

The SAT is an aptitude test that high schools take. It is one of the criteria that is used in a college's decision to admit a student. It is composed of a math and a verbal section. Each has a mean of 500 and a standard deviation of 110 and is normally distributed.

- ▶ What are the scores on the test that corresponds to 3, 2, 1, 0, -1, -2, -3 standard deviations?
- ▶ Assume 1000 people took the SAT,
  - ▶ If Jon got a 700 on the math section, how many people scored above him?
  - ▶ If 300 people scored below Anna on the verbal section, what was Anna's score?
  - ▶ How many people got scores between 390 and 610?

# SAT

The SAT is an aptitude test that high schools take. It is one of the criteria that is used in a college's decision to admit a student. It is composed of a math and a verbal section. Each has a mean of 500 and a standard deviation of 110 and is normally distributed.

- ▶ What are the scores on the test that corresponds to 3, 2, 1, 0, -1, -2, -3 standard deviations?
- ▶ Assume 1000 people took the SAT,
  - ▶ If Jon got a 700 on the math section, how many people scored above him?
  - ▶ If 300 people scored below Anna on the verbal section, what was Anna's score?
  - ▶ How many people got scores between 390 and 610?
  - ▶ If Sigga got a 350 on the math section, how many people scored below her?

# SAT

The SAT is an aptitude test that high schools take. It is one of the criteria that is used in a college's decision to admit a student. It is composed of a math and a verbal section. Each has a mean of 500 and a standard deviation of 110 and is normally distributed.

- ▶ What are the scores on the test that corresponds to 3, 2, 1, 0, -1, -2, -3 standard deviations?
- ▶ Assume 1000 people took the SAT,
  - ▶ If Jon got a 700 on the math section, how many people scored above him?
  - ▶ If 300 people scored below Anna on the verbal section, what was Anna's score?
  - ▶ How many people got scores between 390 and 610?
  - ▶ If Sigga got a 350 on the math section, how many people scored below her?
  - ▶ If Einar was in the 98% percentile in math, what was Einar's score?

## Other standard scores

- ▶ T scores have a mean of 50 and a standard deviation of 10.

# Other standard scores

- ▶ **T scores** have a mean of 50 and a standard deviation of 10.
  - ▶ What would T scores of 30 and 70 be as z-scores?

# Other standard scores

- ▶ **T scores** have a mean of 50 and a standard deviation of 10.
  - ▶ What would T scores of 30 and 70 be as z-scores?
- ▶ **stanine**, range from 1 to 9, are centered at 5 with a standard deviation of 2. Each stanine, corresponds to  $1/2$  a standard deviation and the 5th stanine is at the mean.



# Other standard scores

- ▶ **T scores** have a mean of 50 and a standard deviation of 10.
  - ▶ What would T scores of 30 and 70 be as z-scores?
- ▶ **stanine**, range from 1 to 9, are centered at 5 with a standard deviation of 2. Each stanine, corresponds to  $1/2$  a standard deviation and the 5th stanine is at the mean.
  - ▶ If you were in the 3rd stanine, what would your z-score be?

# Other standard scores

- ▶ **T scores** have a mean of 50 and a standard deviation of 10.
  - ▶ What would T scores of 30 and 70 be as z-scores?
- ▶ **stanine**, range from 1 to 9, are centered at 5 with a standard deviation of 2. Each stanine, corresponds to  $1/2$  a standard deviation and the 5th stanine is at the mean.
  - ▶ If you were in the 3rd stanine, what would your z-score be?
  - ▶ How many people would be below you?

# Other standard scores

- ▶ **T scores** have a mean of 50 and a standard deviation of 10.
  - ▶ What would T scores of 30 and 70 be as z-scores?
- ▶ **stanine**, range from 1 to 9, are centered at 5 with a standard deviation of 2. Each stanine, corresponds to  $1/2$  a standard deviation and the 5th stanine is at the mean.
  - ▶ If you were in the 3rd stanine, what would your z-score be?
  - ▶ How many people would be below you?
  - ▶ What percent of the people are between the 3rd and the 6th stanines?

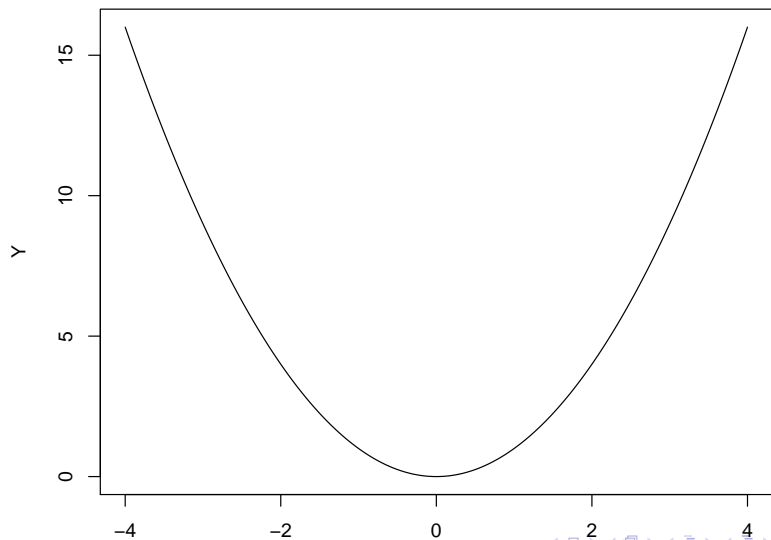
# Other standard scores

- ▶ **T scores** have a mean of 50 and a standard deviation of 10.
  - ▶ What would T scores of 30 and 70 be as z-scores?
- ▶ **stanine**, range from 1 to 9, are centered at 5 with a standard deviation of 2. Each stanine, corresponds to  $1/2$  a standard deviation and the 5th stanine is at the mean.
  - ▶ If you were in the 3rd stanine, what would your z-score be?
  - ▶ How many people would be below you?
  - ▶ What percent of the people are between the 3rd and the 6th stanines?
- ▶ Various linear and non-linear transformations are done to create scores and scores may be normalized.

# What is a correlation?

- ▶ Is it an association?
- ▶ Does it imply causation?
- ▶ Is a correlation necessary for causation?
- ▶ Does it need linearity?
- ▶ Is it affected by variability?
- ▶ Is it affected by outliers?
- ▶ Is it related to the simple linear regression?

# What is the correlation?



# Correlation

$$\frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum((Y - \bar{Y})^2)}}$$

# Calculating correlations

|      | X | Y |
|------|---|---|
|      | 5 | 6 |
|      | 3 | 0 |
|      | 1 | 0 |
| Mean | 3 | 2 |