Correlation & Linear Regression

Slides adopted from the Internet

Roadmap

- *Linear Correlation*
- Spearman's rho correlation
- Kendall's tau correlation
- Linear regression

Linear correlation

Recall: Covariance

$$\operatorname{cov}(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y})}{n-1}$$

Interpreting Covariance

cov(X,Y) > 0 $\longrightarrow$ X and Y are positively correlated

cov(X,Y) < 0 $\longrightarrow$ X and Y are inversely correlated

cov(X,Y) = 0 $\longrightarrow$ X and Y are independent

Correlation coefficient

- Pearson’ s Correlation Coefficient is standardized covariance (unitless):

$$r = \frac{\operatorname{cov}ariance(x, y)}{\sqrt{\operatorname{var} x}\sqrt{\operatorname{var} y}}$$

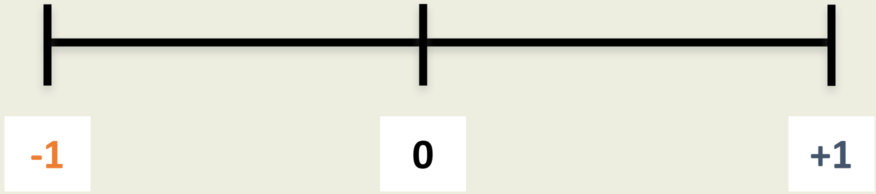
Calculating by hand…

$$\hat{r} = \frac{\operatorname{cov}ariance(x,y)}{\sqrt{\operatorname{var} x}\sqrt{\operatorname{var} y}} = \frac{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}\sqrt{\dfrac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}}$$

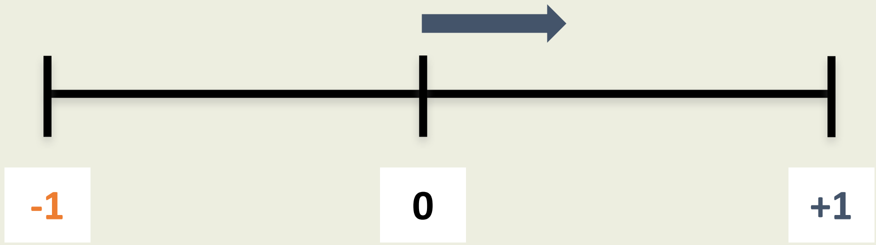
Correlation

- Measures the relative strength of the ***linear*** relationship between two variables
- Unit-less
- Ranges between –1 and 1
- The closer to –1, the stronger the negative linear relationship
- The closer to 1, the stronger the positive linear relationship
- The closer to 0, the weaker any positive linear relationship

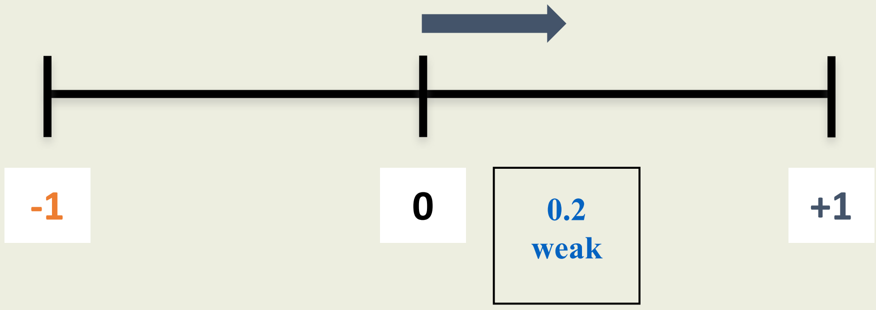
- The strength of the relationship depends on the **decimal value**.



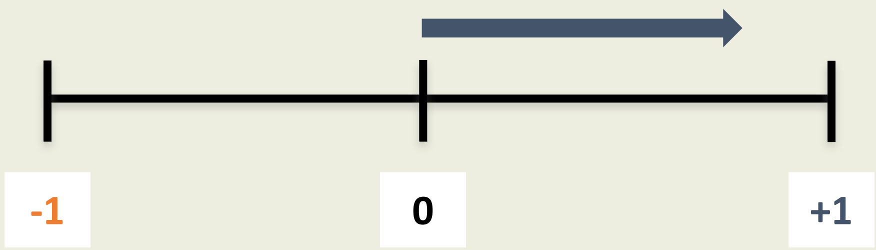
- The strength of the relationship depends on the **decimal value**.



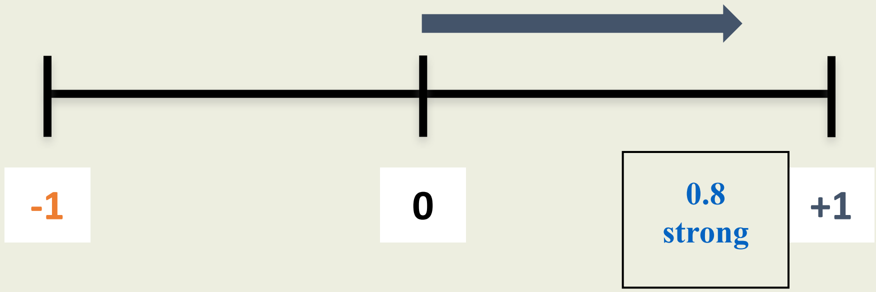
- The strength of the relationship depends on the **decimal value**.



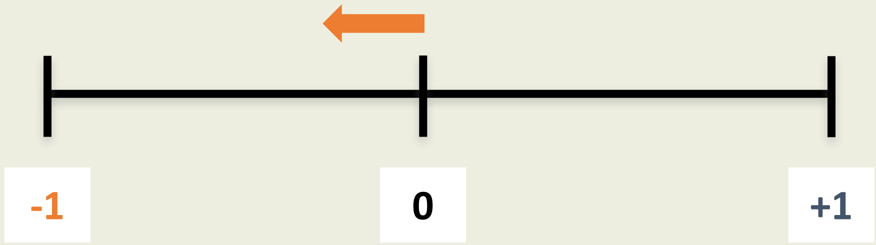
- The strength of the relationship depends on the **decimal value**.



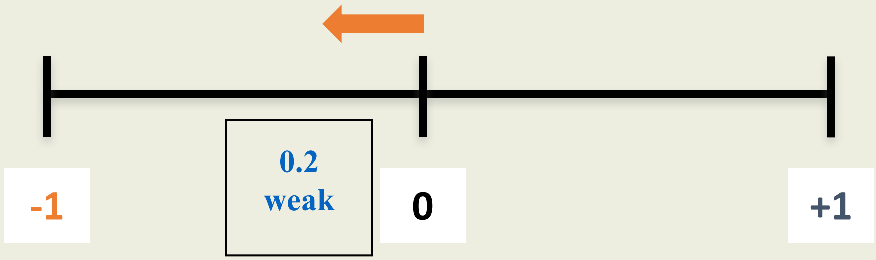
- The strength of the relationship depends on the **decimal value**.



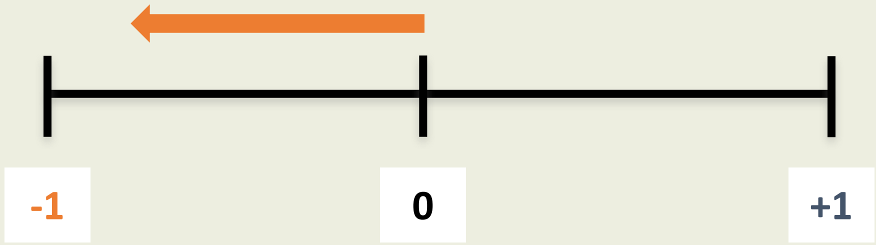
- The strength of the relationship depends on the **decimal value**.



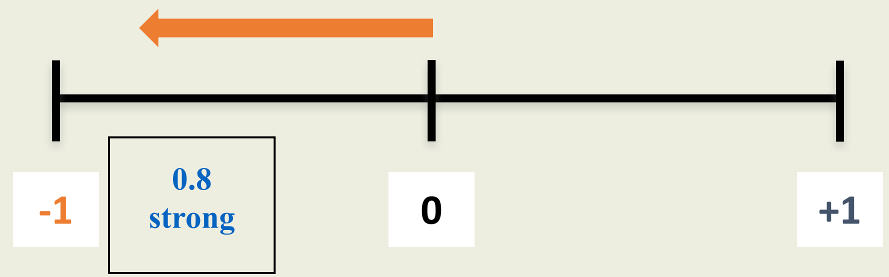
- The strength of the relationship depends on the **decimal value**.



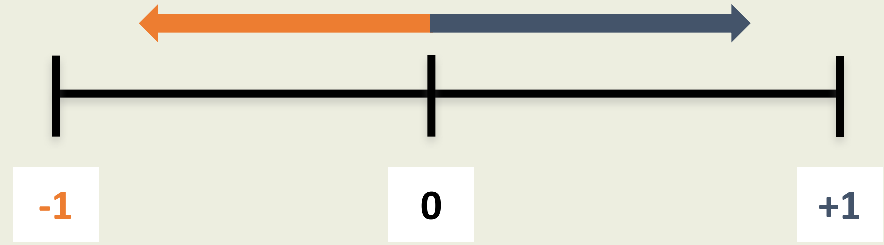
- The strength of the relationship depends on the **decimal value**.



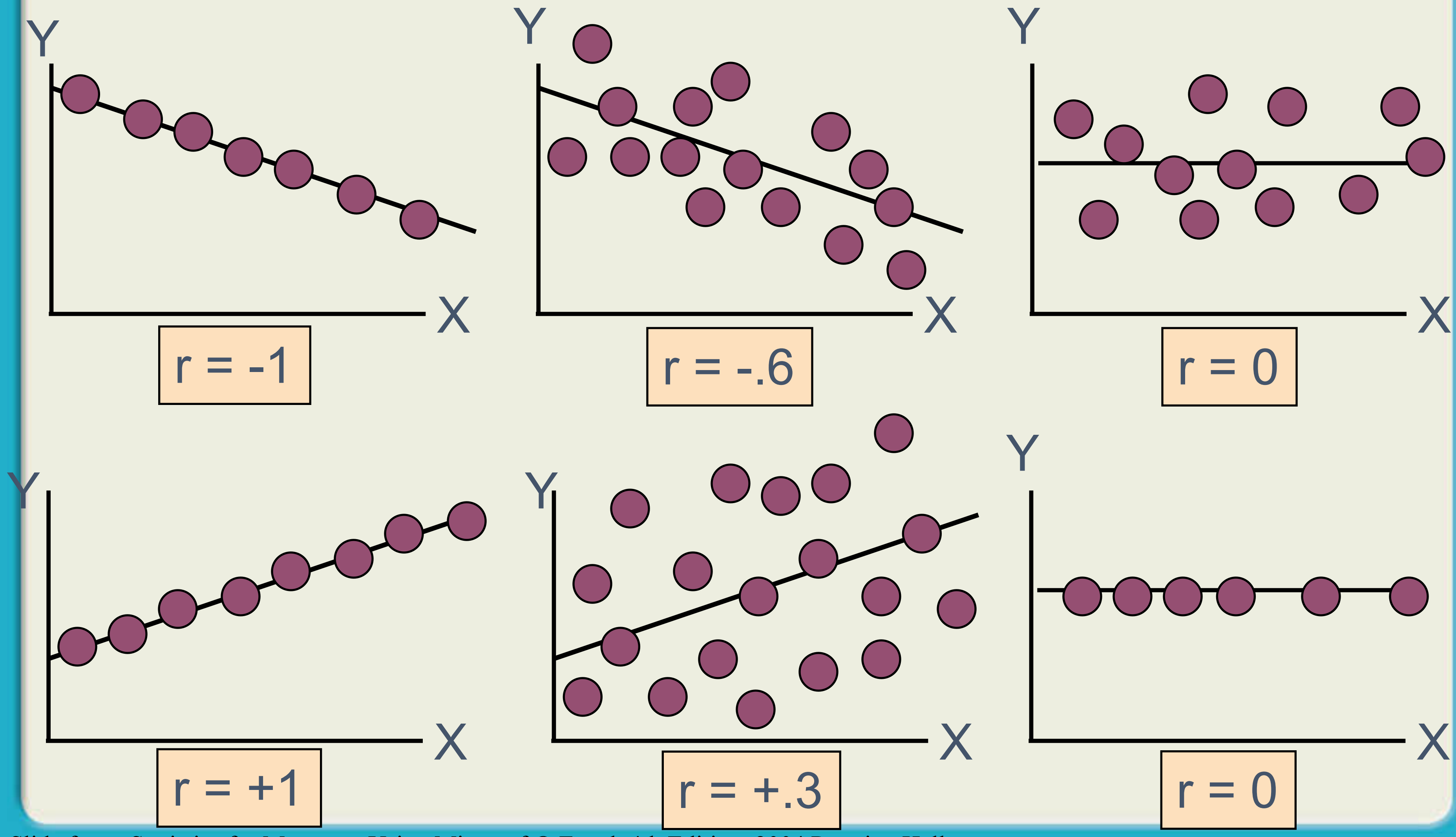
- The strength of the relationship depends on the **decimal value**.



- The strength of the relationship depends on the **decimal value**.

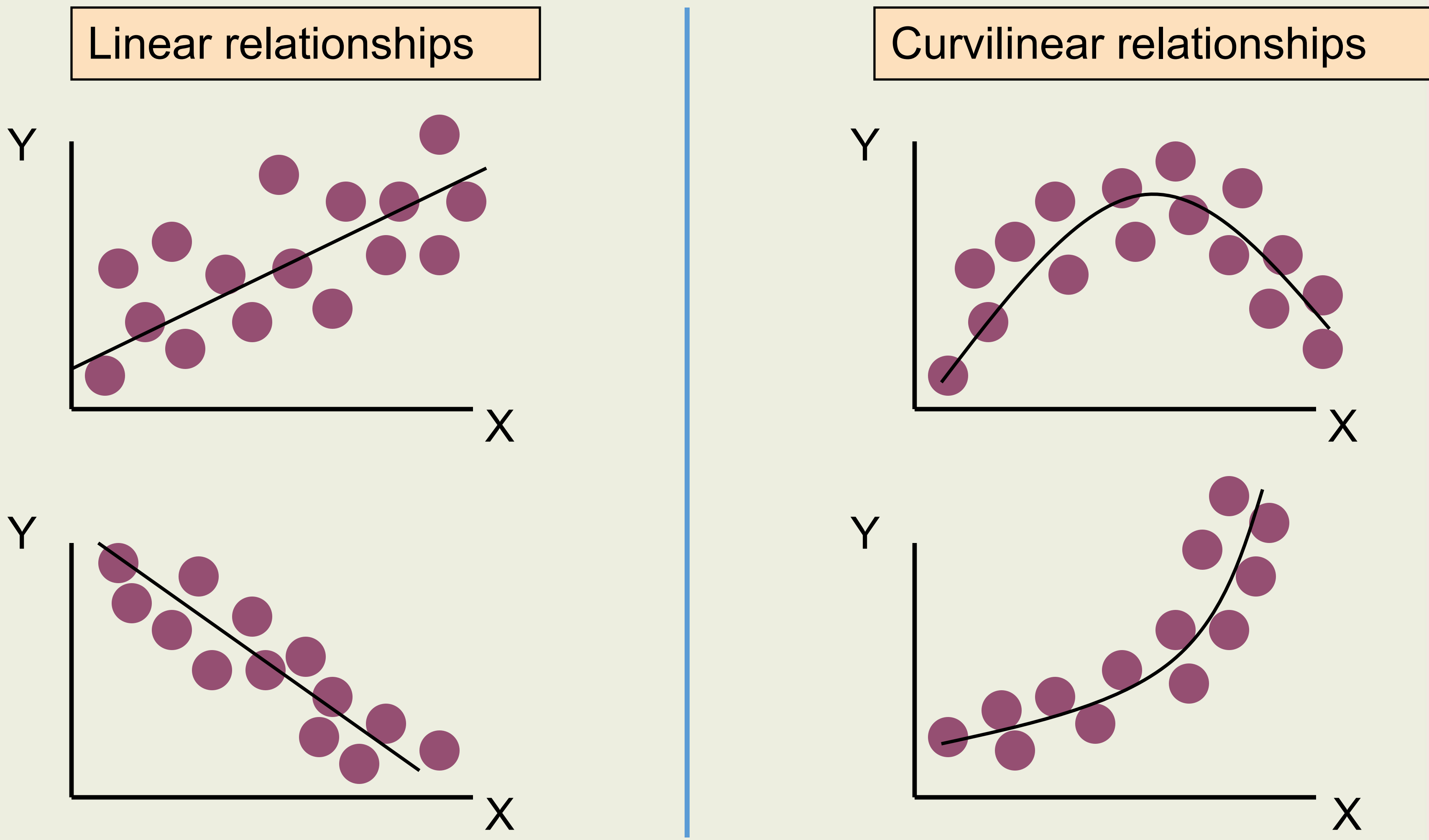


Scatter Plots of Data with Various Correlation Coefficients



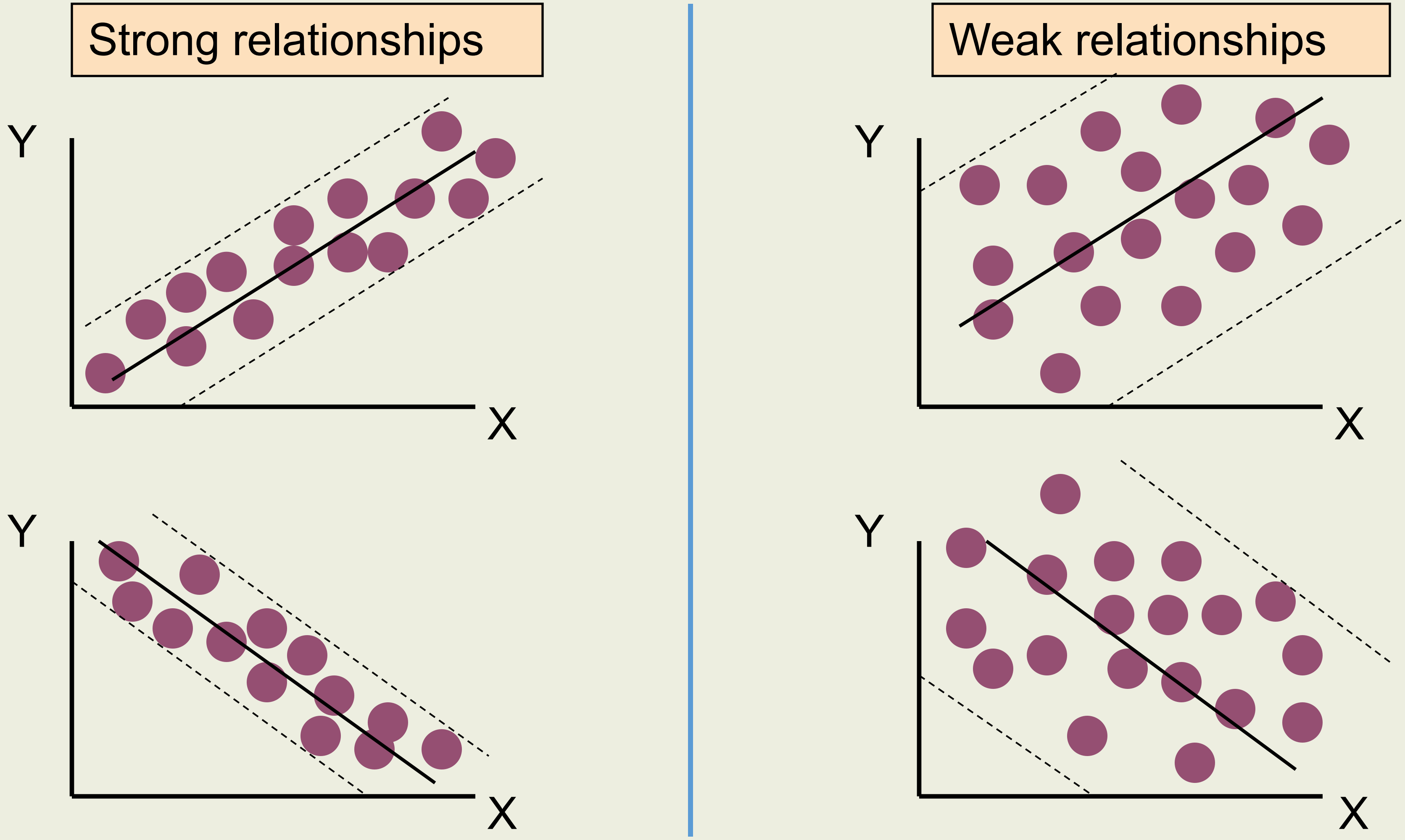
Slide from: Statistics for Managers Using Microsoft® Excel 4th Edition, 2004 Prentice-Hall

Linear Correlation



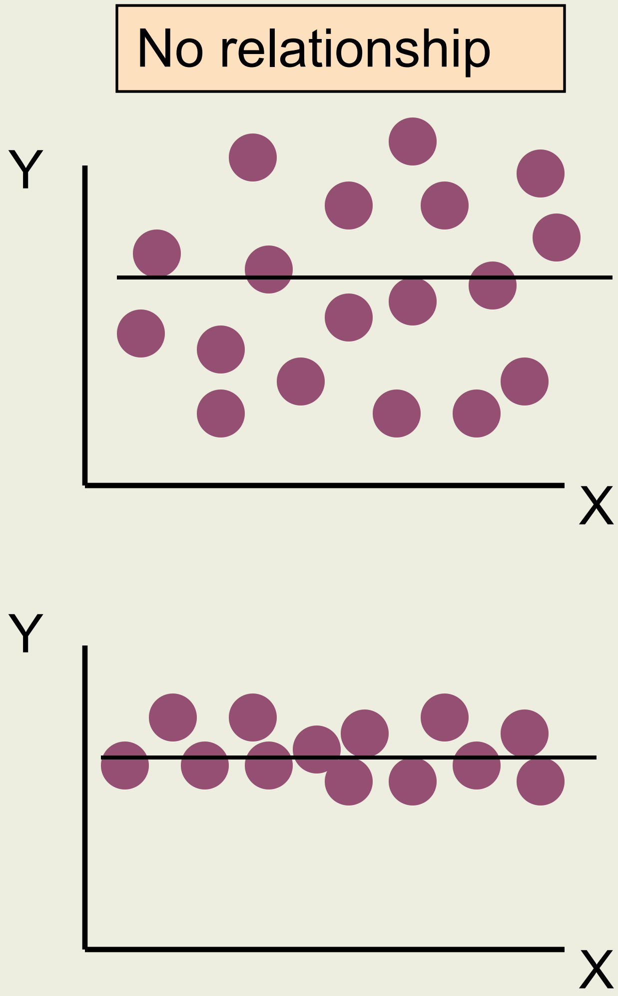
■Slide from: Statistics for Managers Using Microsoft® Excel 4th Edition, 2004 Prentice-Hall

Linear Correlation



Slide from: Statistics for Managers Using Microsoft® Excel 4th Edition, 2004 Prentice-Hall

Linear Correlation



Simpler calculation formula…

$$\hat{r} = \frac{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\cancel{n-1}}}{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\cancel{n-1}}}\sqrt{\dfrac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{\cancel{n-1}}}} =$$

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{SS_{xy}}{\sqrt{SS_x SS_y}}$$

Numerator of covariance

$$\hat{r} = \frac{SS_{xy}}{\sqrt{SS_x SS_y}}$$

Numerators of variance

Pearson r correlation assumptions

- Both variables should be normally distributed
- A straight-line (linear) relationship between two variables
- Data are normally distributed around the regression line

Roadmap

- Linear Correlation
- *Spearman's rho correlation*
- Kendall's tau correlation
- Linear regression

Spearman's Rank-Order Correlation

For Independence Questions

Welcome to the Spearman's Rho Test of Independence Learning Module

What is Spearman's Rho?

(i.e., does not assume data distribution)

- Spearman's "Rho' is a non-parametric analogue to the Pearson Product Moment Correlation.

What is Spearman's Rho?

- Spearman's "Rho' is a non-parametric analogue to the Pearson Product Moment Correlation.
- Spearman's Rho is designed to estimate the coherence or lack of coherence of two variables (as in the Pearson Product Moment Correlation).

What is Spearman's Rho?

- Spearman's "Rho' is a non-parametric analogue to the Pearson Product Moment Correlation.
- Spearman's Rho is designed to estimate the coherence or lack of coherence of two variables (as in the Pearson Product Moment Correlation).
- It is calculated based on the rank-ordered (ordinal) data rather than the means and standard deviation used in the Pearson Product Moment Correlation.

What is Spearman’s Rho?

- Here is an illustration of the difference between a Pearson Correlation and a Spearman’s Rho

What is Spearman’s Rho?

- Here is an illustration of the difference between a Pearson Correlation and a Spearman’s Rho
- Are race times of athletes who participated in both biking and running competitions independent of one another? *(This is a Pearson Correlation question because we are dealing with continuous variables)*

What is Spearman's Rho?

- Here is an illustration of the difference between a Pearson Correlation and a Spearman's Rho
- Are race times of athletes who participated in both biking and running competitions independent of one another? *(This is a Pearson Correlation question because we are dealing with continuous variables)*

Individuals	Biking Event race times	Running Event race times
Bob	4.5 hours	4.0 hours
Conrad	7.0 hours	2.5 hours
Dallen	5.2 hours	2.8 hours
Ernie	6.0 hours	2.9 hours
Fen	6.3 hours	3.3 hours
Gaston	5.1 hours	2.3 hours

What is Spearman's Rho?

- Here is an illustration of the difference between a Pearson Correlation and a Spearman's Rho
- Are race times of athletes who participated in both biking and running competitions independent of one another? *(This is a Pearson Correlation question because we are dealing with continuous variables)*
- *This is a Spearman's Rho question if we are dealing with rank ordered or ordinal data:*

What is Spearman's Rho?

- *This is a Spearman's Rho question if we are dealing with rank ordered or ordinal data:*

Individuals	Biking Event race times	Running Event race times
Bob	1st	6th
Conrad	6th	2nd
Dallen	3rd	3rd
Ernie	4th	4th
Fen	5th	5th
Gaston	2nd	1st

What is Spearman's Rho?

- In summary, if at least one of two variables to be correlated are based on an underlying ordinal measurement, the Spearman's Rho is an appropriate estimate.

What is Spearman's Rho?

- In summary, if at least one of two variables to be correlated are based on an underlying ordinal measurement, the Spearman's Rho is an appropriate estimate.

- For example -

What is Spearman's Rho?

	Interval or continuous Data	Ordinal or rank-ordered Data
Individuals	Biking Event race times in minutes	Running Event placement
Bob	55	6th
Conrad	25	2nd
Dallen	29	3rd
Ernie	33	4th
Fen	39	5th
Gaston	23	1st

What is Spearman's Rho?

- For example –

Individuals	Interval or continuous Data Biking Event race times in minutes	Ordinal or rank-ordered Data Running Event placement
Bob	55	6th
Conrad	25	2nd
Dallen	29	3rd
Ernie	33	4th
Fen	39	5th
Gaston	23	1st

Because this data is ordinal or rank ordered we will use Spearman's Rho

What is Spearman's Rho?

- For example –

	Interval or continuous Data	Ordinal or rank-ordered Data
Individuals	**Biking Event race times in minutes**	**Running Event placement**
Bob	55	6th
Conrad	25	2nd
Dallen	29	3rd
Ernie	33	4th
Fen	39	5th
Gaston	23	1st

- or

What is Spearman's Rho?

- For example –

Individuals	Biking Event race times in minutes (Interval or continuous Data)	Running Event placement (Ordinal or rank-ordered Data)
Bob	55	6th
Conrad	25	2nd
Dallen	29	3rd
Ernie	33	4th
Fen	39	5th
Gaston	23	1st

- or

Individuals	Biking Event placement (Ordinal or rank-ordered Data)	Running Event race times (Interval or continuous Data)
Bob	1st	4.0 hours
Conrad	6th	2.5 hours
Dallen	3rd	2.8 hours
Ernie	4th	2.9 hours
Fen	5th	3.3 hours
Gaston	2nd	2.3 hours

What is Spearman's Rho?

- For example –

Individuals	Biking Event race times in minutes	Running Event placement
	Interval or continuous Data	Ordinal or rank-ordered Data
Bob	55	6th
Conrad	25	2nd
Dallen	29	3rd
Ernie	33	4th
Fen	39	5th
Gaston	23	1st

- or

ls	Biking Event placement	Running Event race times
	Ordinal or rank-ordered Data	Interval or continuous Data
	1st	4.0 hours
	6th	2.5 hours
	3rd	2.8 hours
	4th	2.9 hours
	5th	3.3 hours
Gaston	2nd	2.3 hours

Because this data is ordinal or rank ordered we will use Spearman's Rho

What is Spearman's Rho?

- If both variables are on an interval scale, but one or both are significantly skewed, then Spearman's Rho is an appropriate estimate that compensates for distortion of the mean.

What is Spearman's Rho?

- If both variables are on an interval scale, but one or both are significantly skewed, then Spearman's Rho is an appropriate estimate that compensates for distortion of the mean.
- For example:

What is Spearman's Rho?

- If both variables are on an interval scale, but one or both are significantly skewed, then Spearman's Rho is an appropriate estimate that compensates for distortion of the mean.
- For example:

	Interval –heavily skewed data	Interval normally distributed Data
Individuals	**Biking Event race times**	**Running Event race times**
Bob	4.5 hours	4.0 hours
Conrad	4.6 hours	2.5 hours
Dallen	4.7 hours	2.8 hours
Ernie	5.0 hours	2.9 hours
Fen	20.0 hours	3.3 hours
Gaston	28.0 hours	2.3 hours

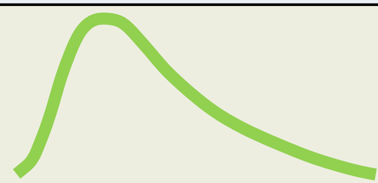
What is Spearman's Rho?

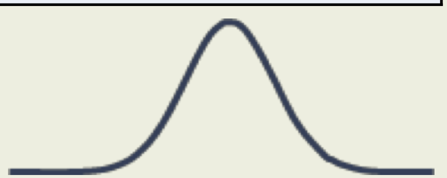
- If both variables are on an interval scale, but one or both are significantly skewed, then Spearman's Rho is an appropriate estimate that compensates for distortion of the mean.
- For example:

Interval –heavily skewed data

Interval normally distributed Data

Individuals	Biking Event race times	Running Event race times
Bob	4.5 hours	4.0 hours
Conrad	4.6 hours	2.5 hours
Dallen	4.7 hours	2.8 hours
Ernie	5.0 hours	2.9 hours
Fen	20.0 hours	3.3 hours
Gaston	28.0 hours	2.3 hours





How to calculate Rho?

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

Where:

P= Spearman rank correlation

di= the difference between the ranks of corresponding values Xi and Yi

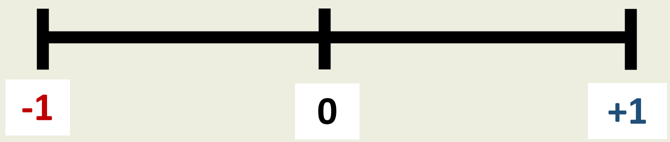
n= number of value in each data set

What is Spearman's Rho?

- Spearman's Rho renders a result that is similar to the Pearson Correlation

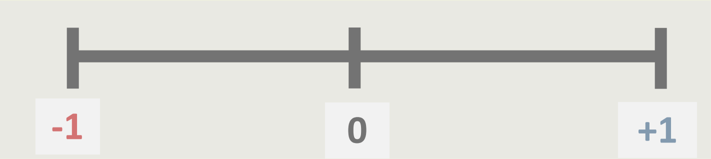
What is Spearman's Rho?

- Spearman's Rho renders a result that is similar to the Pearson Correlation



What is Spearman's Rho?

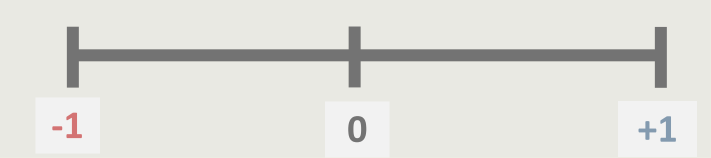
- Spearman's Rho renders a result that is similar to the Pearson Correlation



- Therefore it shares the same properties as these other methods:

What is Spearman's Rho?

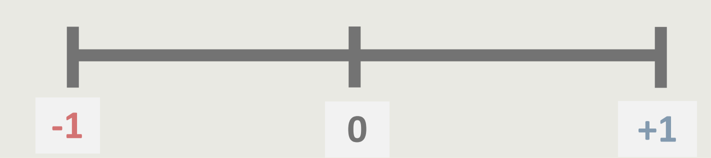
- Spearman's Rho renders a result that is similar to the Pearson Correlation



- Therefore it shares the same properties as these other methods:
  - It ranges from -1 to +1.

What is Spearman's Rho?

- Spearman's Rho renders a result that is similar to the Pearson Correlation



- Therefore it shares the same properties as these other methods:
  - It ranges from -1 to +1.
  - It's direction is determined by the sign (- +)

What is Spearman's Rho?

- Spearman's Rho renders a result that is similar to the Pearson Correlation

- Therefore it shares the same properties as these other methods:
  - It ranges from -1 to +1.
  - It's direction is determined by the sign (- +)
  - The closer the value is to -1 or +1, the stronger the relationship

What is Spearman's Rho?

- Spearman's Rho renders a result that is similar to the Pearson Correlation

- Therefore it shares the same properties as these other methods:
  - It ranges from -1 to +1.
  - It's direction is determined by the sign (- +)
  - The closer the value is to -1 or +1, the stronger the relationship
  - The closer the value is to 0, the weaker the relationship.

What is Spearman's Rho?

- Spearman's Rho renders a result that is similar to the Pearson Correlation

- Therefore it shares the same properties as these other methods:
  - It ranges from -1 to +1.
  - It's direction is determined by the sign (- +)
  - The closer the value is to -1 or +1, the stronger the relationship
  - The closer the value is to 0, the weaker the relationship.

A result like this would be evidence of independence

What is Spearman's Rho?

- It differs from Kendall's Tau in one simple way. The Spearman's Rho **CANNOT** handle ties. The Kendall's Tau can:

What is Spearman's Rho?

- It differs from Kendall's Tau in one simple way. The Spearman's Rho cannot handle ties. The Kendall's Tau can:
- For example:

What is Spearman's Rho?

- It differs from Kendall's Tau in one simple way. The Spearman's Rho cannot handle ties. The Kendall's Tau can:
- For example:

Individuals	Rank order for Biking Event	Rank order for Running Event
Bob	1st	**1st**
Conrad	**2nd**	**1st**
Dallen	**2nd**	2nd
Ernie	3rd	3rd
Fen	4th	**4th**
Gaston	5th	**4th**

*use Kendall's Tau when there are rank ordered ties.

Spearman's Rho Assumptions

- non-parametric: it does not assume any assumptions about the distribution of the data
- Is the appropriate correlation analysis when the variables are measured on a scale that is at least ordinal.
- Scores on one variable must be monotonically related to the other variable.
- Cannot deal with ties

Roadmap

- Linear Correlation
- Spearman's rho correlation
- *Kendall's tau correlation*
- Linear regression

What is a Kendall Tau?

What is Kendall’s Tau?

Kendall’s Tau is a nonparametric analogue to the Pearson Product Moment Correlation.

What is Kendall's Tau?

Similar to Spearman's Rho, Kendall's Tau operates on rank-ordered (ordinal) data but is particularly useful when there are tied ranks.

Let's consider an investigation that would lend itself to being analyzed by Kendall's Tau:

What is Kendall's Tau?

An iron man competition consists of three consecutive events:

What is Kendall's Tau?

An iron man competition consists of three consecutive events: Biking 110 miles,

What is Kendall's Tau?

An iron man competition consists of three consecutive events: Biking 110 miles, Swimming 2.5 miles

What is Kendall's Tau?

An iron man competition consists of three consecutive events: Biking 110 miles, Swimming 2.5 miles and Running 26.2 miles

What is Kendall's Tau?

An iron man competition consists of three consecutive events: Biking 110 miles, Swimming 2.5 miles and Running 26.2 miles. Researchers are interested in the relationship between the rank ordered results from the biking and the running events.

What is Kendall's Tau?

An iron man competition consists of three consecutive events: Biking 110 miles, Swimming 2.5 miles and Running 26.2 miles. Researchers are interested in the relationship between the rank ordered results from the biking and the running events.

What is Kendall's Tau?

An iron man competition consists of three consecutive events: Biking 110 miles, Swimming 2.5 miles and Running 26.2 miles. Researchers are interested in the relationship between the rank ordered results from the biking and the running events. Here is the data for 6 individuals who competed:

What is Kendall's Tau?

Individuals	Rank order for Biking Event	Rank order for Running Event
Bob		
Conrad		
Dallen		
Ernie		
Fen		
Gaston		

What is Kendall's Tau?

Individuals	Rank order for Biking Event	Rank order for Running Event
Bob	1st	
Conrad	2nd	
Dallen	2nd	
Ernie	3rd	
Fen	4th	
Gaston	5th	

What is Kendall's Tau?

Individuals	Rank order for Biking Event	Rank order for Running Event
Bob	1^{st}	1^{st}
Conrad	2^{nd}	1^{st}
Dallen	2^{nd}	2^{nd}
Ernie	3^{rd}	3^{rd}
Fen	4^{th}	4^{th}
Gaston	5^{th}	4^{th}

What is Kendall's Tau?

Because both variables are expressed as rank ordered data, we will use either a Kendall's Tau or a Spearman's Rho.

What is Kendall's Tau?

Because both variables are expressed as rank ordered data, we will use either a Kendall's Tau or a Spearman's Rho.

Note – even if only one variable were ordinal and the other were scaled or nominal, you would still use Kendall's Tau or a Spearman's Rho by virtue of having ***one ordinal variable***.

What is Kendall's Tau?

Because there are ties in the data, we will use Kendall's Tau ***instead*** of the Spearman's Rho.

How to calculate Tau?

Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be a set of observations of the joint random variables X and Y respectively, such that all the values of (x_i) and (y_i) are unique. Any pair of observations (x_i, y_i) and (x_j, y_j) are said to be *concordant* if the ranks for both elements agree: that is, if both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$. They are said to be *discordant*, if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant nor discordant.

The Kendall τ coefficient is defined as:

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\frac{1}{2}n(n-1)}.$$ [3]

What is Kendall's Tau?

Because there are ties in the data, we will use Kendall's Tau ***instead*** of the Spearman's Rho.

Individuals	Rank order for Biking Event	Rank order for Running Event
Bob	1st	**1st**
Conrad	**2nd**	**1st**
Dallen	**2nd**	2nd
Ernie	3rd	3rd
Fen	4th	**4th**
Gaston	5th	**4th**

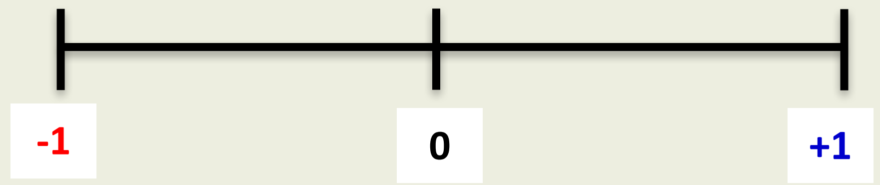
What is Kendall's Tau?

Kendall's Tau renders a result that is similar to Spearman's Rho and the Pearson Correlation

- Therefore it shares the same properties as these other methods:
  - It ranges from -1 to +1.
  - It's direction is determined by the sign (+ -)
  - The closer the value is to -1 or +1, the stronger the relationship
  - The closer the value is to 0, the weaker the relationship.

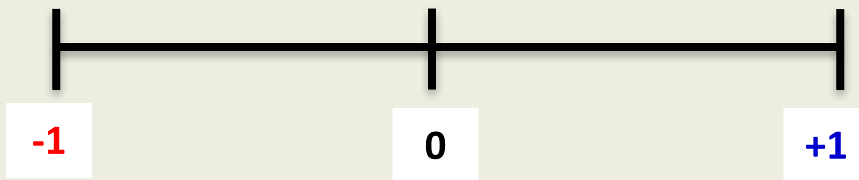
What is Kendall's Tau?

Kendall's Tau renders a result that is similar to Spearman's Rho and the Pearson Correlation



What is Kendall's Tau?

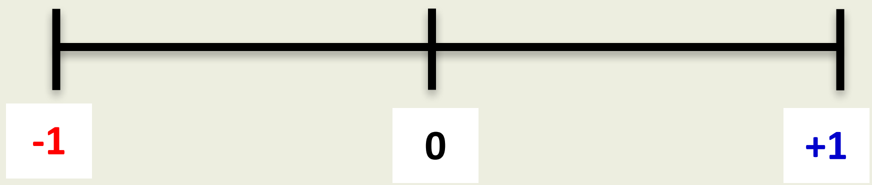
Kendall's Tau renders a result that is similar to Spearman's Rho and the Pearson Correlation



- Therefore it shares the same properties as these other methods:
  - It ranges from -1 to +1.
  - It's direction is determined by the sign (- +).
  - The closer the value is to -1 or +1, the stronger the relationship
  - The closer the value is to 0, the weaker the relationship.

What is Kendall’s Tau?

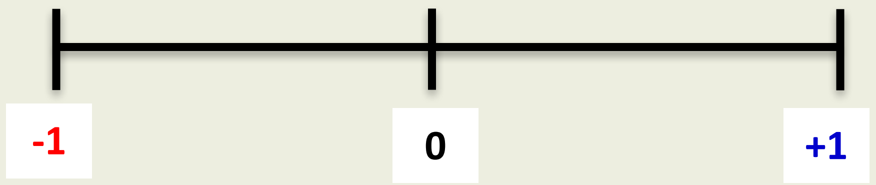
Kendall’s Tau renders a result that is similar to Spearman’s Rho and the Pearson Correlation



- Therefore it shares the same properties as these other methods:
  - **It ranges from -1 to +1.**
  - It’s direction is determined by the sign (- +)
  - The closer the value is to -1 or +1, the stronger the relationship
  - The closer the value is to 0, the weaker the relationship.

What is Kendall's Tau?

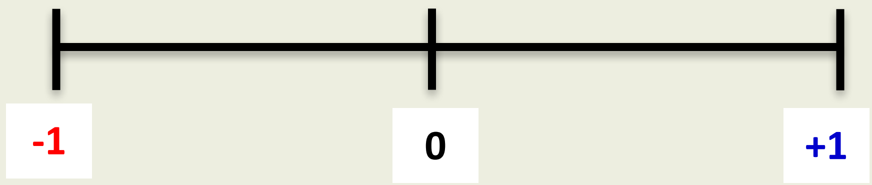
Kendall's Tau renders a result that is similar to Spearman's Rho and the Pearson Correlation



- Therefore it shares the same properties as these other methods:
  - It ranges from -1 to +1.
  - **It's direction is determined by the sign (- +)**
  - The closer the value is to -1 or +1, the stronger the relationship
  - The closer the value is to 0, the weaker the relationship.

What is Kendall's Tau?

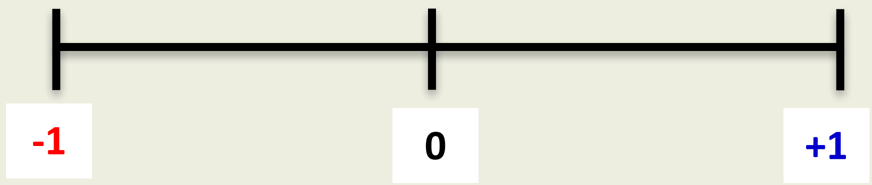
Kendall's Tau renders a result that is similar to Spearman's Rho and the Pearson Correlation



- Therefore it shares the same properties as these other methods:
  - It ranges from -1 to +1.
  - It's direction is determined by the sign (- +)
  - **The closer the value is to -1 or +1, the stronger the relationship**
  - The closer the value is to 0, the weaker the relationship.

What is Kendall’s Tau?

Kendall’s Tau renders a result that is similar to Spearman’s Rho and the Pearson Correlation



- Therefore it shares the same properties as these other methods:
  - It ranges from -1 to +1.
  - It’s direction is determined by the sign (- +)
  - The closer the value is to -1 or +1, the stronger the relationship
  - **The closer the value is to 0, the weaker the relationship.**

Kendall's Tau Assumptions

- non-parametric: does not assume any assumptions about the distribution of the data
- Is the appropriate correlation analysis when the variables are measured on a scale that is at least ordinal.
- Can deal with ties

Binned Kenall Correlation

Use this "binned" Kendall correlation under two scenarios:

- Skewed data distribution
  - To this end, we look at the average value for each bin and compute the correlation on the binned data.
- Amount of the data so large that rank correlation is computationally expensive
  - The binned correlation retains the qualitative properties that we want to highlight with lower compute cost.

Roadmap

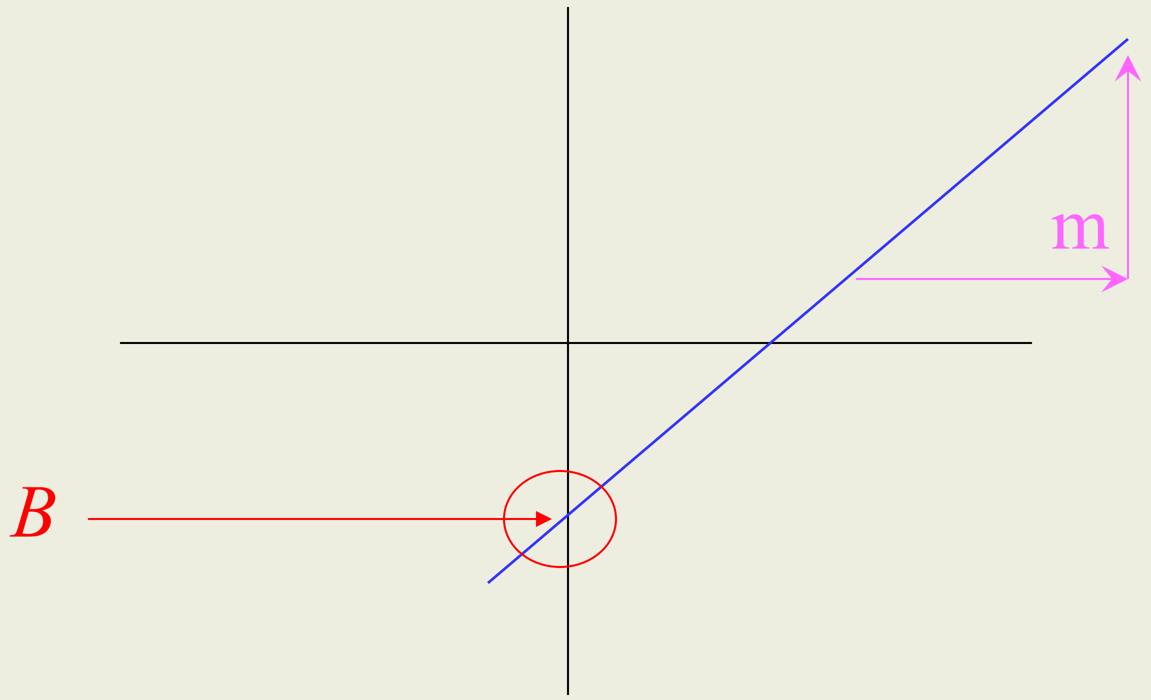
- Linear Correlation
- Spearman's rho correlation
- Kendall's tau correlation
- *Linear regression*

Linear regression

In correlation, the two variables are treated as equals. In regression, one variable is considered independent (=predictor) variable (X) and the other the dependent (=outcome) variable Y.

What is "Linear"?

- Remember this:
- *Y=mX+B?*



What’s Slope?

A slope of 2 means that every 1-unit change in X yields a 2-unit change in Y.

Prediction

If you know something about X, this knowledge helps you predict something about Y. (Sound familiar?…sound like conditional probabilities?)

Regression equation…

Expected value of y at a given level of *x*=

$$\boxed{E(y_i / x_i) = \alpha + \beta x_i}$$

Predicted value for an individual…

$$\hat{y}_i = \underbrace{\alpha + \beta * x_i}_{\text{Fixed – exactly on the line}} + \boxed{\text{random error}_i}$$

Assumption: Follows a normal distribution

Estimating the intercept and slope: least squares estimation

** Least Squares Estimation
A little calculus….
What are we trying to estimate? **β, the slope,** from

What's the constraint? We are trying to minimize the squared distance (hence the "least squares") between the observations themselves and the predicted values , or (<u>also called the "residuals", or left-over unexplained variability</u>)

Difference$_i$ = $y_i - (\beta x + \alpha)$ Difference$_i^2$ = $(y_i - (\beta x + \alpha))^2$

Find the β that gives the minimum sum of the squared differences. How do you maximize a function? Take the derivative; set it equal to zero; and solve. Typical max/min problem from calculus….

$$\frac{d}{d\beta}\sum_{i=1}^{n}(y_i - (\beta x_i + \alpha))^2 = 2(\sum_{i=1}^{n}(y_i - \beta x_i - \alpha)(-x_i))$$

$$2(\sum_{i=1}^{n}(-y_i x_i + \beta x_i^{\,2} + \alpha x_i)) = 0...$$

From here takes a little math trickery to solve for β...

Resulting formulas…

Slope (beta coefficient) = $$\hat{\beta} = \frac{Cov(x, y)}{Var(x)}$$

Intercept= $$\text{Calculate}: \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

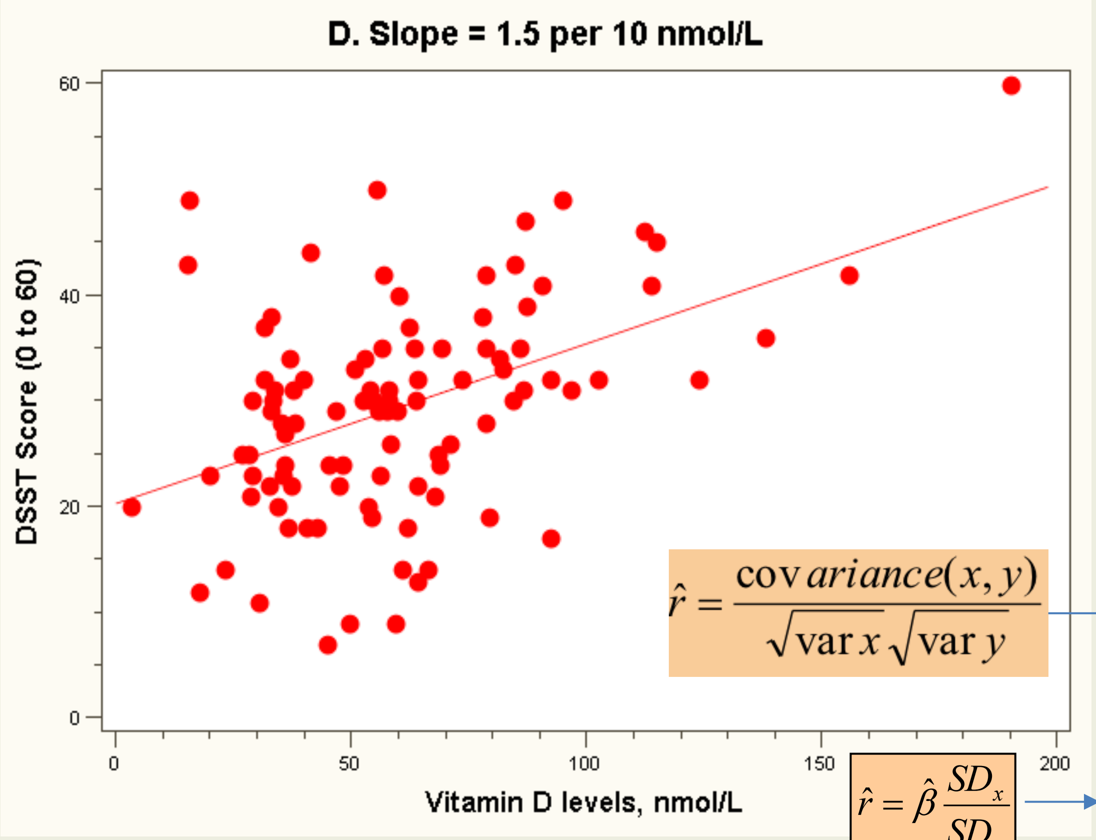
Regression line always goes through the point: $(\bar{x}, \bar{y})$

Relationship with correlation

$$\hat{r} = \hat{\beta}\frac{SD_x}{SD_y}$$

In correlation, the two variables are treated as equals. In regression, one variable is considered independent (=predictor) variable (X) and the other the dependent (=outcome) variable Y.

Example:



$$\hat{r} = \frac{\mathrm{cov}ariance(x,y)}{\sqrt{\mathrm{var}\,x}\sqrt{\mathrm{var}\,y}}$$

$$\hat{r} = \hat{\beta}\frac{SD_x}{SD_y}$$

SDx = 33 nmol/L

SDy= 10 points

Cov(X,Y) = 163 points*nmol/L

Beta = 163/33^2 = 0.15 points per nmol/L

= 1.5 points per 10 nmol/L

r = 163/(10*33) = 0.49

Or

r = 0.15 * (33/10) = 0.49

Pearson r correlation assumptions

- Both variables should be normally distributed
- A straight-line (linear) relationship between two variables
- Data are normally distributed around the regression line

Summary

Assumptions	Pearson r /linear regression	Spearman’s Rho	Kendall’s Tau
distributions of two variables	both are normally distributed	no assumption	no assumption
variable property	both are numbers	at least ordinal	at least ordinal
relationship between two variables	linear	scores on one variable must be monotonically related to the other variable	does not assume monotonic relationship
misc.	data are normally distributed around the regression line	cannot deal with tie	can deal with ties