WEEK 13

SIMPLE CORRELATION ANALYSIS

CORRELATION ANALYSIS

Investigates the association between two or more numerical variables

e.g. Chest depth and live weight

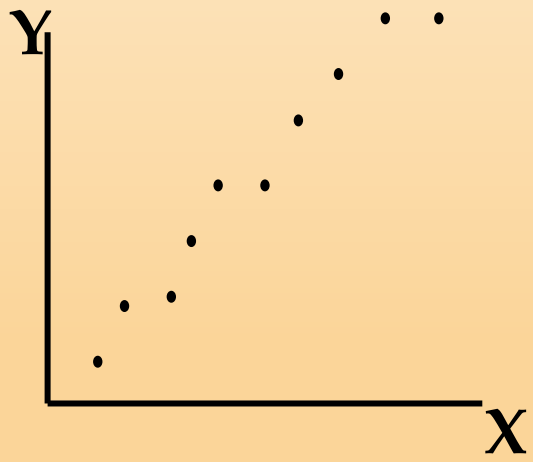
If there is association;

- Direction
- Strength

DIRECTION OF ASSOCIATION?

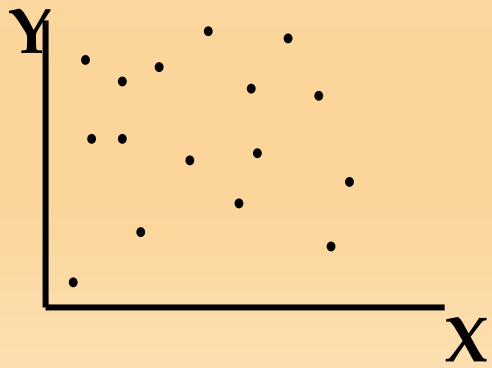
Positive

as one variable increases in value, the other variable increases



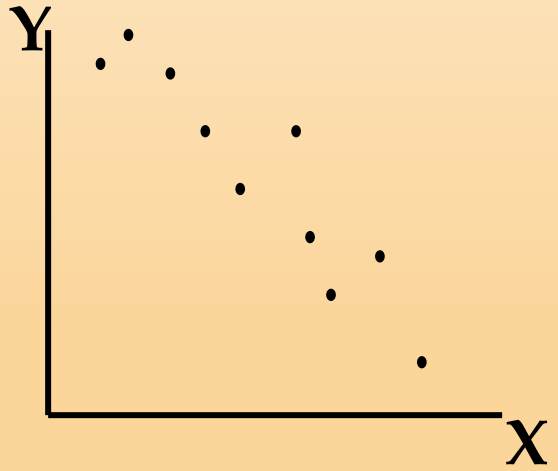
No linear association

there is a random scatter of points with no indication of a linear relation between the variables



Negative

as one variable increases in value, the other variable decreases



CORRELATION ANALYSIS

STRENGTH OF RELATIONSHIP?

Correlation coefficient (r)

- Correlation coefficient can take any value from -1 to +1. !!!!!
- The sign of the correlation coefficient (r) gives information about the direction of the relationship (positive or negative)
- The closer the value of the correlation coefficient is to either of its extreme values (-1 or +1), the stronger the relationship between the variables.
  - Interchanging x and y does not affect the value of r
  - A significant relationship between x and y does not provide evidence of a causal relationship

r = -1: perfect negative correlation
(as x increases, y decreases, or vice versa)
r = +1: perfect positive correlation
(as x increases, y increases, or vice versa)
r = 0: no association between x and y

Correlation coefficient (r) can be interpreted as following:

0.0 - 0.20	Very weak
0.21 - 0.40	Weak
0.41 - 0.60	Medium
0.61 - 0.84	Strong
0.85 - 1.0	Very strong

CORRELATION ANALYSIS

Pearson Correlation Coefficient

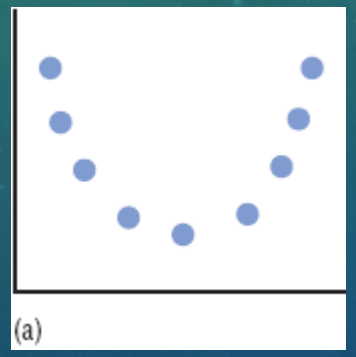
If,
- All of the variables are continouos
- Both variables are normally distributed
- Sample size is large enough
- There is a linear relationship between x and y

Spearman Rank Correlation Coefficient

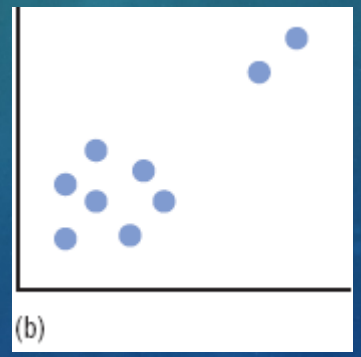
If
- one of the variables is ordinal
- both of the variables are not normally distributed
- sample size is small
- there is no linear relationship between x and y

Misuse of the correlation coefficient

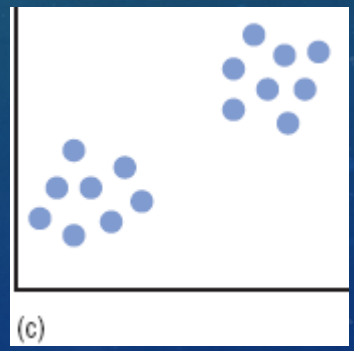
- Underlying relationship between x and y should be linear
- Observations should be independent
- There should be no outliers. Extreme values may distort the value of r.



(a) non-linear association



(b) Extreme values



(c) Subgrouped data set.

ASSUMPTIONS OF CORRELATION ANALYSIS

- Both of the variables, x and y, are numerical.

- The hypothesis test that the true population correlation coefficient is zero only requires *at east one* of the two variables to be Normally distributed in the population (strictly, one variable is Normally distributed with constant variance for any given value of the other variable).

If the data are measured on an ordinal scale or if we are concerned about the distributional assumptions in other circumstances, we calculate **Spearman's rank correlation coefficient**

STEPS OF CORRELATION ANALYSIS

Step 1: Establish your hypothesis

H_0: Correlation coefficient is equal to zero (no association between x and y)

H_1: Correlation coefficient is not equal to zero (there is a association between x and y)

Step 2: Collect the data and display them in a scatter diagram to see the relationship

Step 3: Calculate the correlation coefficient (r) and the related “p-value”

Step 4: Interpret your findings

- Interpret the direction of association by looking at the sign of r
- Interpret the strength of association by looking at value of r
- Interpret the significance of your correlation coefficient by looking at your p value

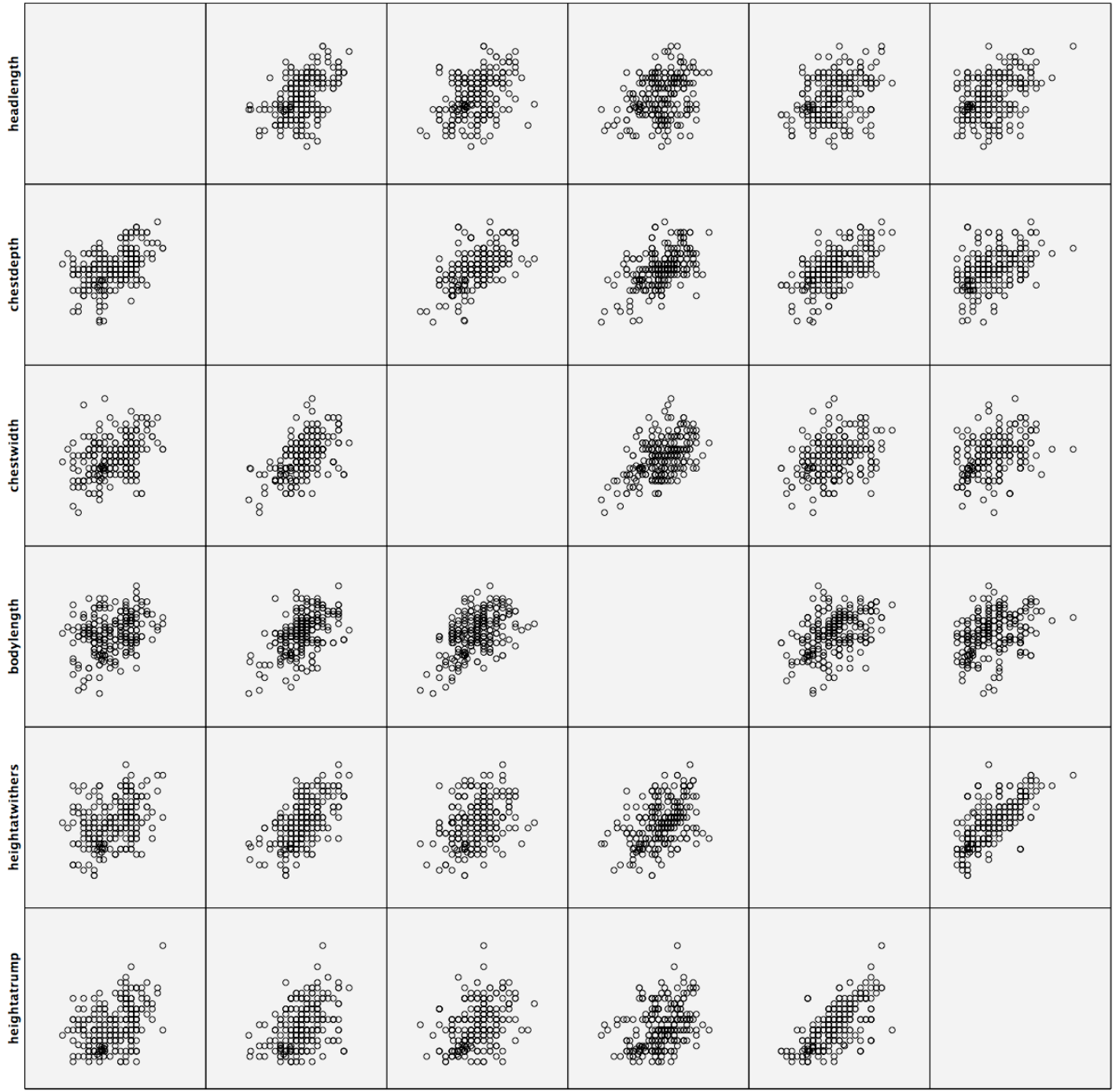
EXAMPLE:

	bodyweight	headlength	chestdepth	chestwidth	bodylength	heightatwithers	heightatr...
1	54,60	24,50	32,50	19,50	68,00	66,00	67,00
2	56,80	21,50	32,00	19,50	67,50	65,00	65,00
3	50,00	21,00	33,00	17,50	69,00	65,00	64,00
4	54,70	23,00	32,00	21,00	70,50	65,50	63,50
5	60,20	21,00	34,00	21,50	69,00	64,50	63,50
6	44,30	20,50	29,00	19,00	65,50	63,00	63,50
7	48,60	22,00	31,50	21,00	62,00	62,00	63,00
8	57,00	22,50	34,00	20,00	70,00	65,00	63,00
9	54,80	22,00	31,00	21,00	71,00	64,00	63,00
10	57,40	21,00	32,00	18,00	70,50	65,50	63,00
11	62,10	21,50	34,00	21,00	68,00	66,00	63,00
12	53,00	24,00	35,00	20,00	66,00	66,00	63,00

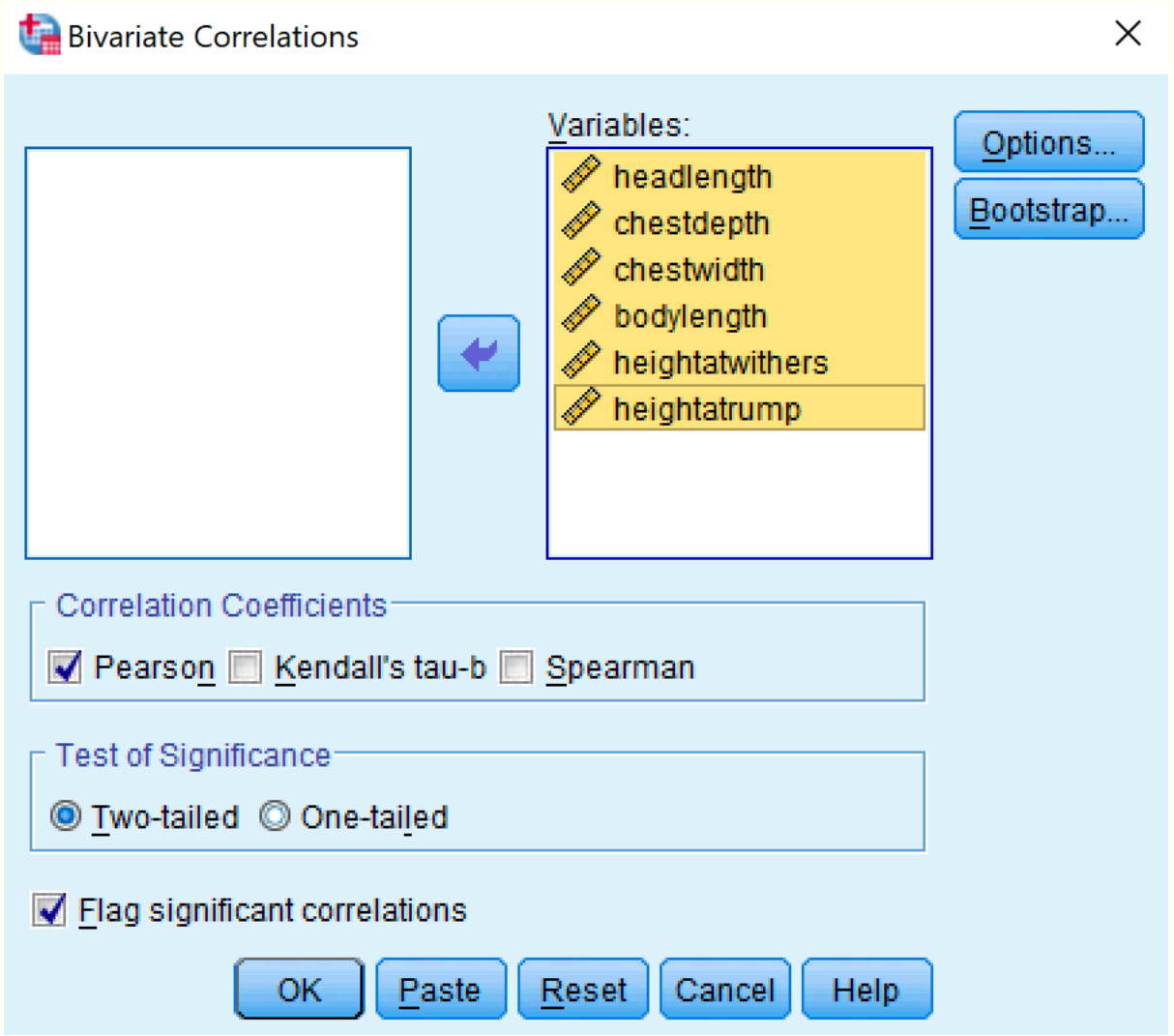
A researcher wants to examine the relationship between body measurements of awassi sheeps. For this reason, he collects various body measurements (*eg. Headlength, chestdepth, chest width, body length, height at withers and height at rump*) of 250 awasi sheeps.

What is the Hypothesis?

Graphs > Legacy Dialogs > Scatter/Dot



> Analyze > Correlate
> Bivariate Correlation



Correlations

		headlength	chestdepth	chestwidth	bodylength	heightatwithers	heightatrump
headlength	Pearson Correlation	1	,477**	,367**	,300**	,359**	,399**
	Sig. (2-tailed)		,000	,000	,000	,000	,000
	N	250	250	250	250	250	250
chestdepth	Pearson Correlation	,477**	1	,539**	,511**	,601**	,395**
	Sig. (2-tailed)	,000		,000	,000	,000	,000
	N	250	250	250	250	250	250
chestwidth	Pearson Correlation	,367**	,539**	1	,494**	,312**	,397**
	Sig. (2-tailed)	,000	,000		,000	,000	,000
	N	250	250	250	250	250	250
bodylength	Pearson Correlation	,300**	,511**	,494**	1	,422**	,317**
	Sig. (2-tailed)	,000	,000	,000		,000	,000
	N	250	250	250	250	250	250
heightatwithers	Pearson Correlation	,359**	,601**	,312**	,422**	1	,665**
	Sig. (2-tailed)	,000	,000	,000	,000		,000
	N	250	250	250	250	250	250
heightatrump	Pearson Correlation	,399**	,395**	,397**	,317**	,665**	1
	Sig. (2-tailed)	,000	,000	,000	,000	,000	
	N	250	250	250	250	250	250

**. Correlation is significant at the 0.01 level (2-tailed).

INTERPRETATION ?

There is a strong positive correlation between withers height and chest depth ($r=0,601$, $p<0.001$)