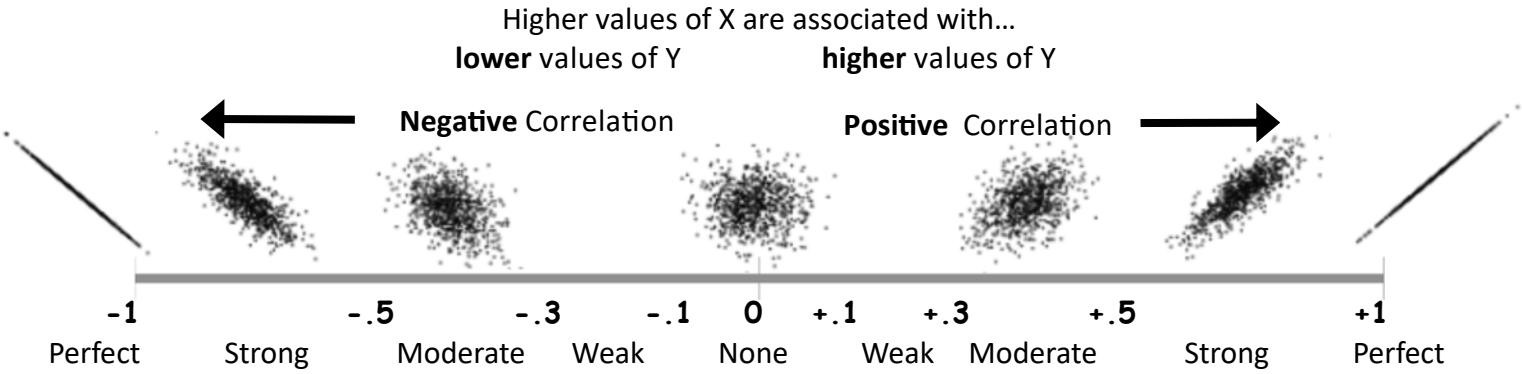
Identifying Bivariate Relationships

2 Numeric Variables: Correlation

The word *correlation* implies a Bivariate Pearson [Linear] Correlation (r). If you do not mean this, say *relationship*.



The exact values at which the relationship is considered weak, medium or strong depends on your field and the topic.

1 Numeric & 1 Categorical Variable: Means Table

Relationships are always described as differences between 2 or more groups on the mean of the numeric variable. .

The standard deviation is useful to evaluate differences between means. A difference of about a half of a standard deviation is often meaningful, but use your judgement.

	$\bar{x}$	*sd*	*n*
Group A	49.3	17.4	322
Group B	53.7	16.8	275
	mean of the sample	*standard deviation*	*number of obs*

2 Categorical Variables: Contingency Table

A contingency (2-way) table may show a difference in 1 variable or a relationship across 2 (a difference of differences).

n	1	2	
A	**196**	**49**	245
B	**112**	**48**	160

hard to compare!

⇨

n	1	2	
A	196/**245**	49/**245**	245/**245**
B	112/**160**	48/**160**	160/**160**

divide each value in a row by the total

⇨

%	1	2	
A	**80**	**20**	100%
B	**70**	**30**	100%

A & B *are comparable*

Percentages equalize the sample size in each group allowing comparisons. Include %'s for the predictor variable (whichever "comes 1st").

Compare the percentages parallel to the 100%'s. If they are approximately equal, there is *not* a relationship.

If *both* A & B are more likely to be in group 1 or 2, that is just a 1 variable ***difference***.

%	1	2	
A	**50**	**50**	100%
B	**50**	**50**	100%

no relationship, no difference

%	1	2	
A	**80**	**20**	100%
B	**80**	**20**	100%

no relationship, a difference (more 1s)

A ***relationship*** is a difference of differences, or a dependence.

%	1	2	
A	**80**	**20**	100%
B	**60**	**40**	100%

a relationship: Bs *more likely to be* 2s

%	1	2	
A	**50**	**50**	100%
B	**60**	**40**	100%

a relationship: Bs *less likely to be* 2s