Module 5: Correlation

The Applied Research Center

Module 5 Overview

- Definition of Correlation
- Relationship Questions
- Scatterplots
- Strength and Direction of Correlations
- Running a Pearson Product Moment Correlation
- Factors Affecting the Correlation
- Other Types of Correlations

Correlation

- The term correlation is used to describe the relationship between two variables
- Pearson's *r* is used to quantify the relationship between two continuous variables

Relationship Questions

- Relationship questions ask the following:
- As scores on one variable go up (or down), what happens to scores on the second variable?
- We are trying to identify a **trend**

Characteristics of Relationship Questions

- Note that the previous questions could all be phrased in terms of the following question:
- As scores on *X* go up (or down), what happens to scores on *Y* ?
- Such a statement was not possible with comparative questions (e.g., scores on gender can't go up or down!)

Scatterplots

- Data for a correlation can be visually displayed using a scatterplot
- The scales of the two variables are plotted on the X and Y axes
- For each observation, a dot is placed at the point at which the X and Y scores intersect

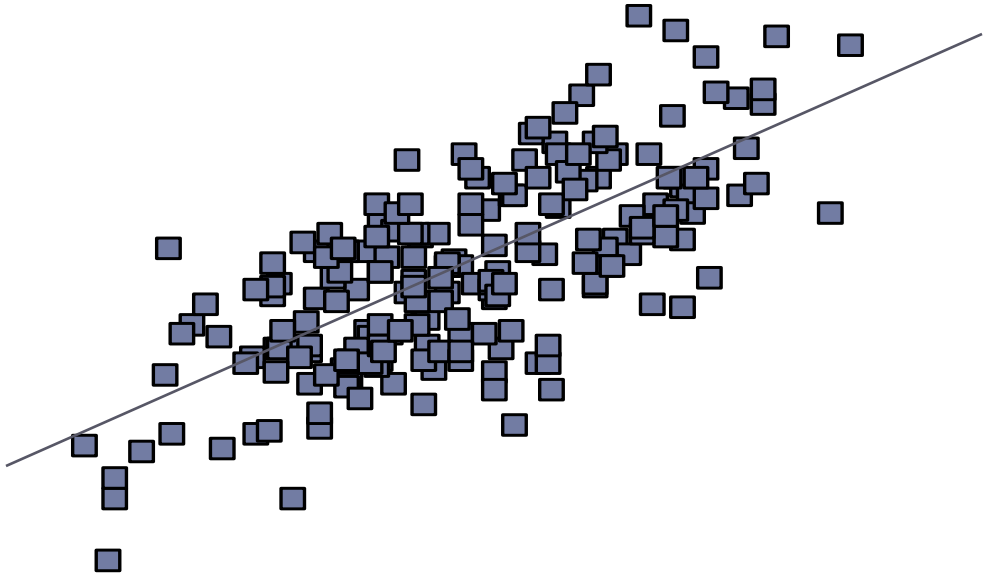
Characteristics of a Correlation

- Scatterplots can help us describe both the direction and the strength of the correlation
- The direction of a correlation can either be positive or negative
- The strength of a correlation can range from weak (or none = 0) to strong (perfect = 111)

Positive Correlations

- Scores on the two variables tend to move in the same direction
- Those who score high on one variable tend to score high on the other (and vice versa)
- Examples:
  - SAT scores and college GPA
  - Age and reading ability
  - Years of education and salary

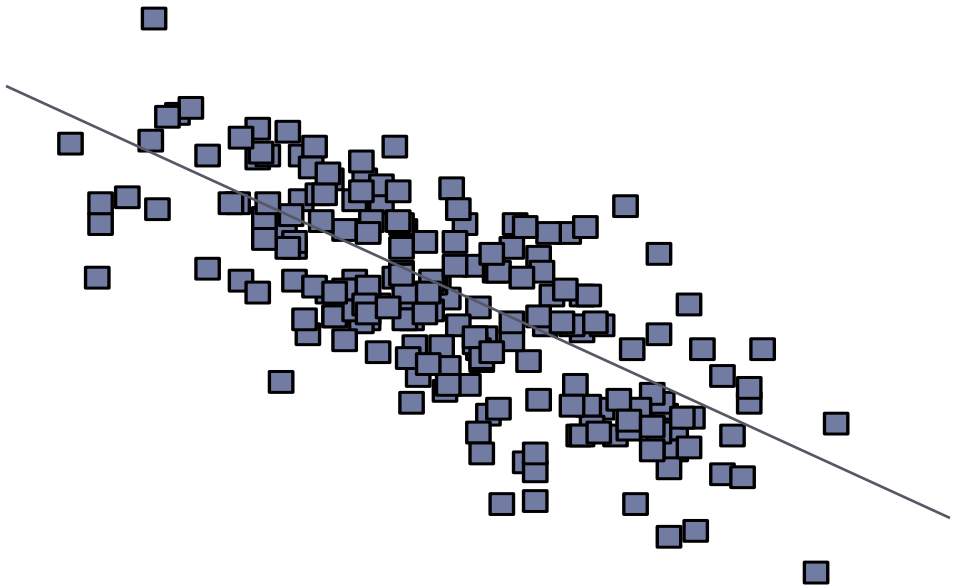
Scatterplot of a Positive Correlation



Negative Correlations

- The two variables are inversely related
- Those who score high on one variable tend to score low on the other (and vice versa)
- Examples:
  - Percentage free/reduced lunch and FCAT scores
  - Anxiety and test performance
  - Stress and job satisfaction

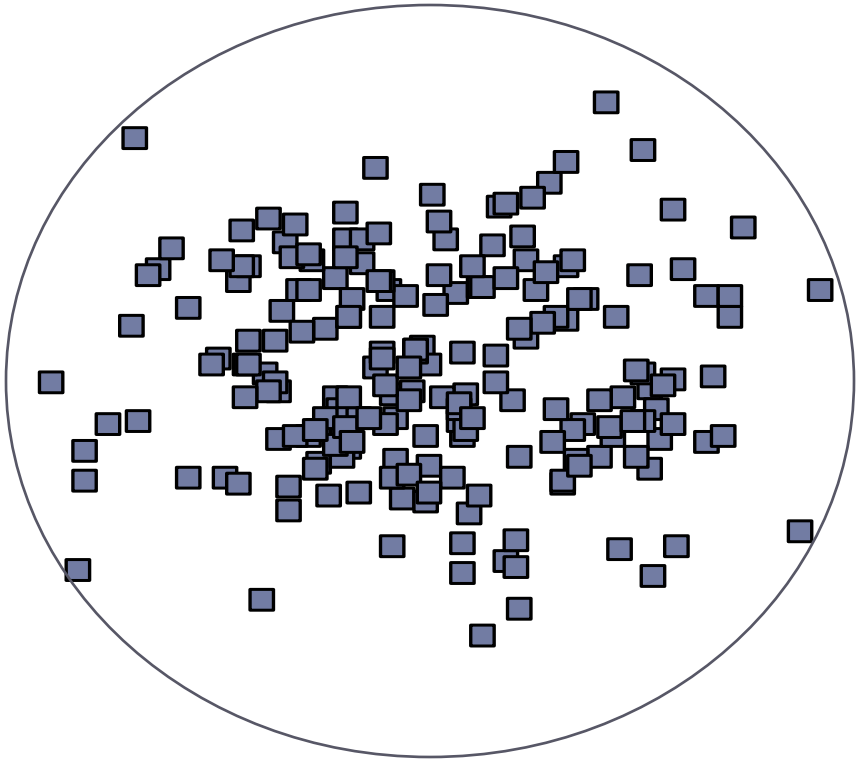
Scatterplot of a Negative Correlation



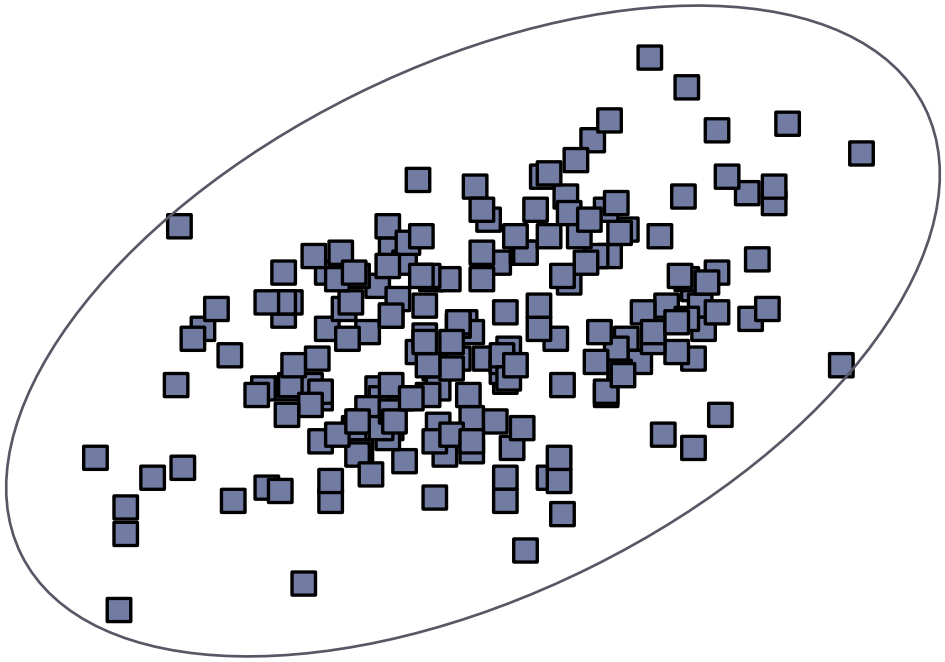
Strength of the Relationship

- The strength of the relationship can be visually estimated by degree to which the data fall on a straight line (i.e., the degree of **linear** trend)
- The correlation gets stronger as the plot approaches a straight line, and reaches a maximum when all data points fall directly on a line

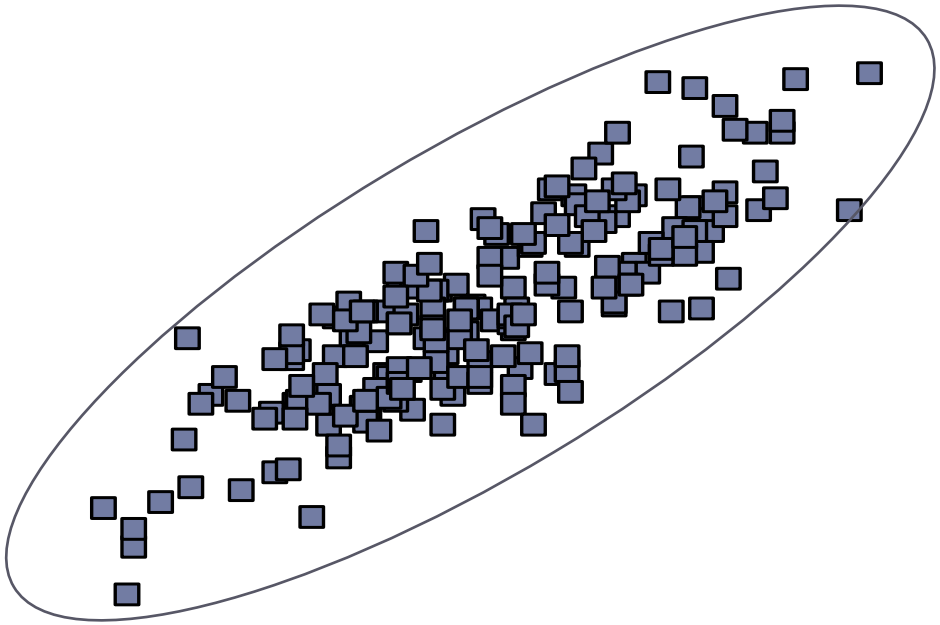
No Correlation



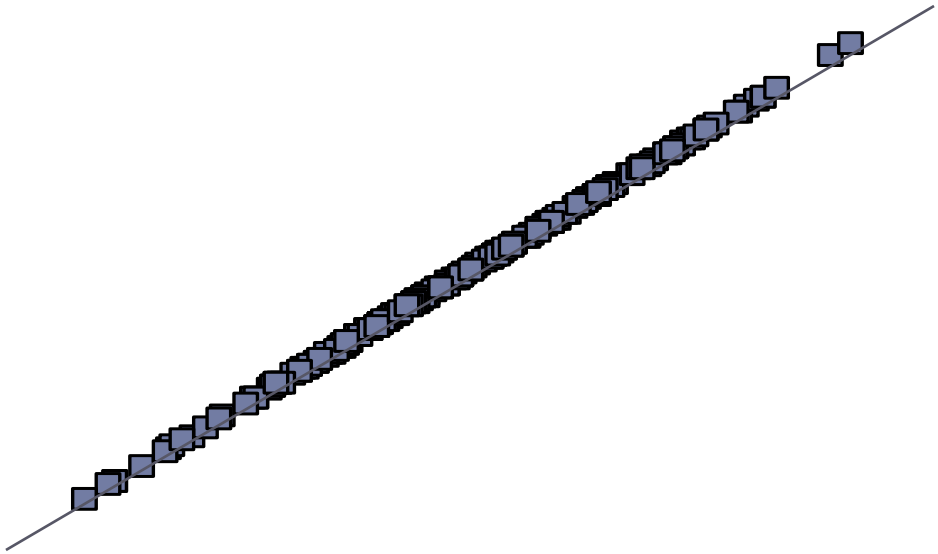
Weak Correlation



Strong Correlation



Perfect Correlation

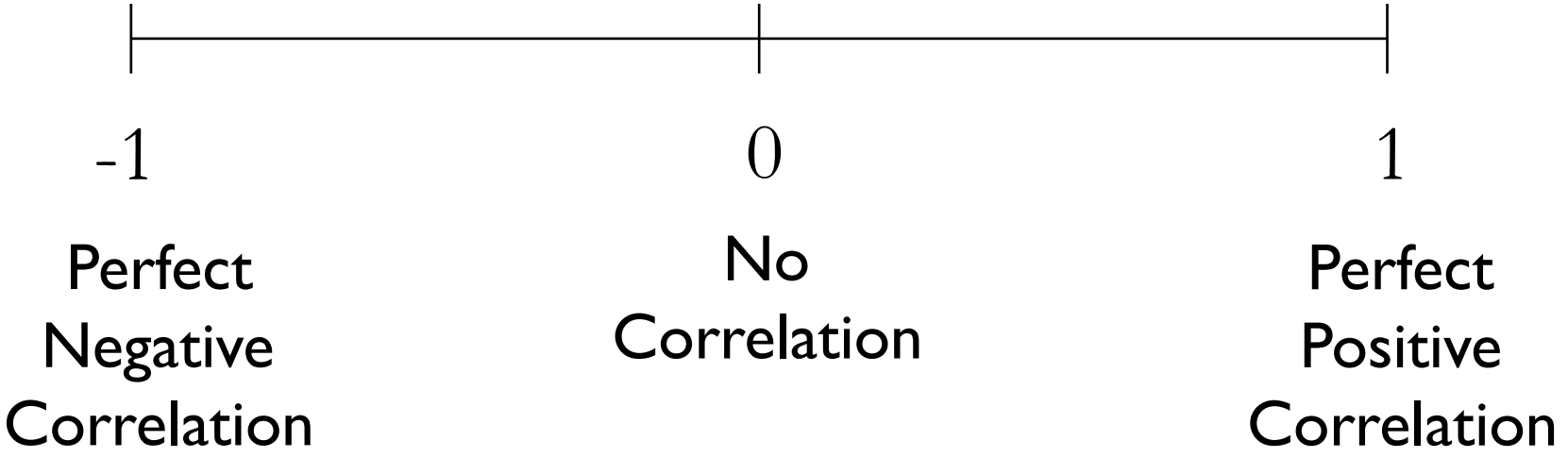


Pearson's *r*

- When describing the correlation between two continuous variables, Pearson's *r* is used
- This index quantifies the degree (and direction) of the linear trend in the data
- The sign of *r* (+ or -) gives the direction of the correlation
- The magnitude of *r* gives the strength of the relationship

Pearson's *r* Scale

- Pearson's *r* ranges in value from -1 to 1



Example 1

- Using the online course survey, it is of interest to determine if there is a relationship between age and satisfaction
- As age increases, what happens to satisfaction?
- What type of trend exists?

Example 1 (cont'd)

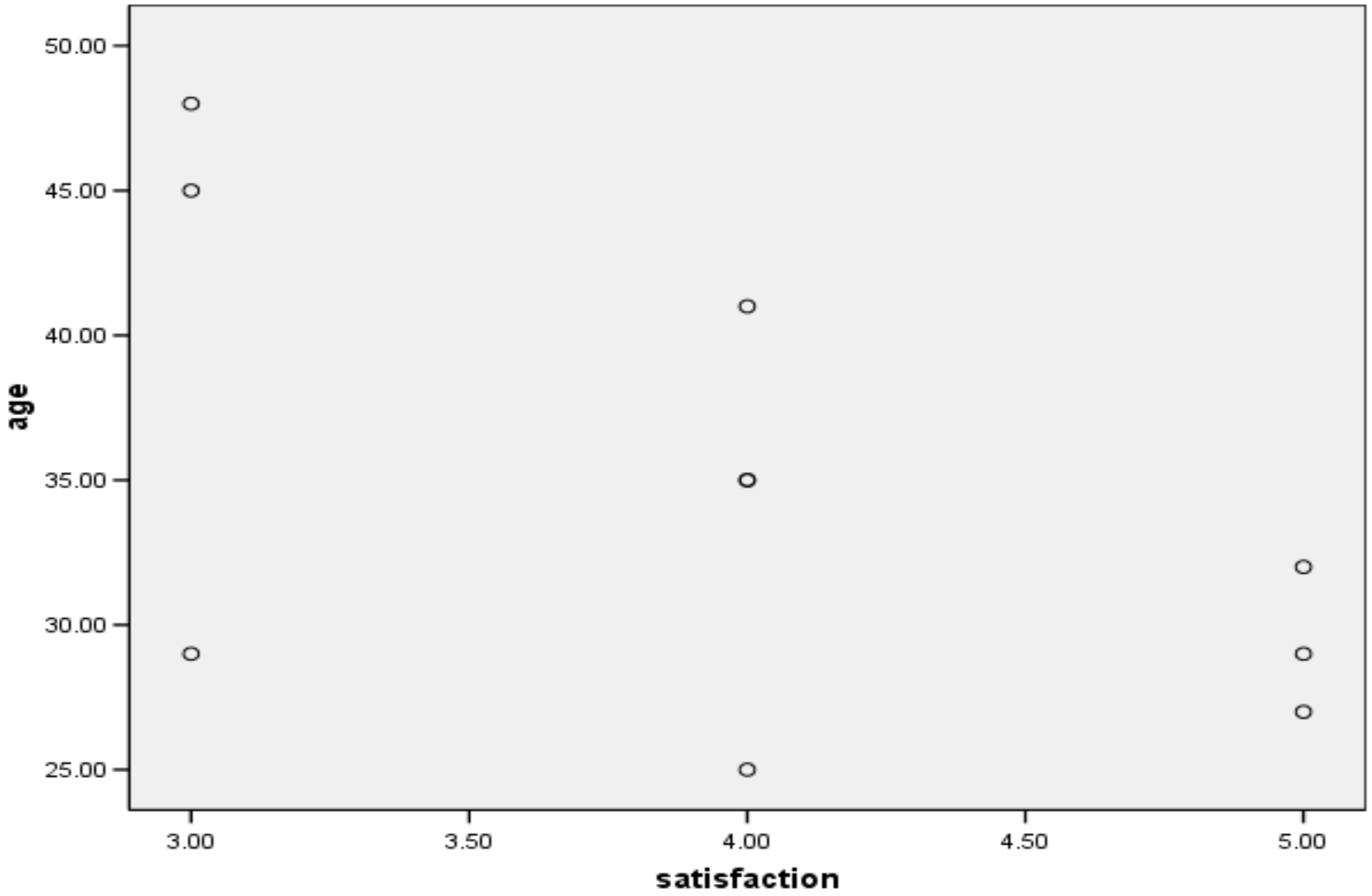
- Step 1: Write your null and alternate hypotheses:
- What is the null hypothesis?
  - Ho: X = Y
  - Written out?
- What is the alternate hypothesis?
  - Ha: X ≠ Y, Ha: X > Y, or Ha: X < Y ?
  - Written out?

- Is this a one-tailed or two-tailed test? Why?

Example 1 (cont'd)

- Step 2: Create a simple scatterplot with age on the Y axis and satisfaction on the X axis.
  - Graph → Legacy Dialogue → Scatter/Dot → Simple Scatter → Define
  - Age → Y-axis
  - Satisfaction → the X-axis.
  - Click OK

Example 1 (cont'd)



Example 1 (cont'd)

- Step 3: Run the Bivariate Correlation
  - Analyze → Correlate → Bivariate
  - Select the variables, select Pearson, and select two-tailed → OK

Example 1 (cont'd)

- From SPSS, the correlation between satisfaction and age was -.593 (r = -.593)
- The correlation is negative
- Thus, as age increases satisfaction with the course tends to decrease

Correlations

		age	satisfaction
age	Pearson Correlation	1	-.593
	Sig. (2-tailed)		.071
	N	10	10
satisfaction	Pearson Correlation	-.593	1
	Sig. (2-tailed)	.071	
	N	10	10

Incorrect Interpretations

- Pearson's r is not a percentage (i.e., there is not a 59% relationship)
- A correlation of .59 is not twice as strong as a correlation of .29
- A correlation of .59 does not mean that satisfaction scores can be predicted with 59% accuracy

Correlation and Causation

- The correlation coefficient simply describes the degree of relationship between two variables
- A correlation does not tell us why two variables are related, nor does it allow for causal statements
- As always, causality is **NOT** a statistical issue, and can only be inferred when using true experiments with random assignment.

Testing r for Statistical Significance

- When taking samples out of a population, correlation values will differ somewhat simply due to random chance (i.e., sampling error)
- The question is, could an r value of -.593 have occurred simply due to chance, or does this represent a "true" relationship between these variables in the population

Probability Value

Correlations

		age	satisfaction
age	Pearson Correlation	1	-.593
	Sig. (2-tailed)		.071
	N	10	10
satisfaction	Pearson Correlation	-.593	1
	Sig. (2-tailed)	.071	
	N	10	10

- The p-value for the correlation is .07 (p = .07)
- Do we reject or fail to reject the null?

Example 1 (cont'd)

- Step 4: Write up your results

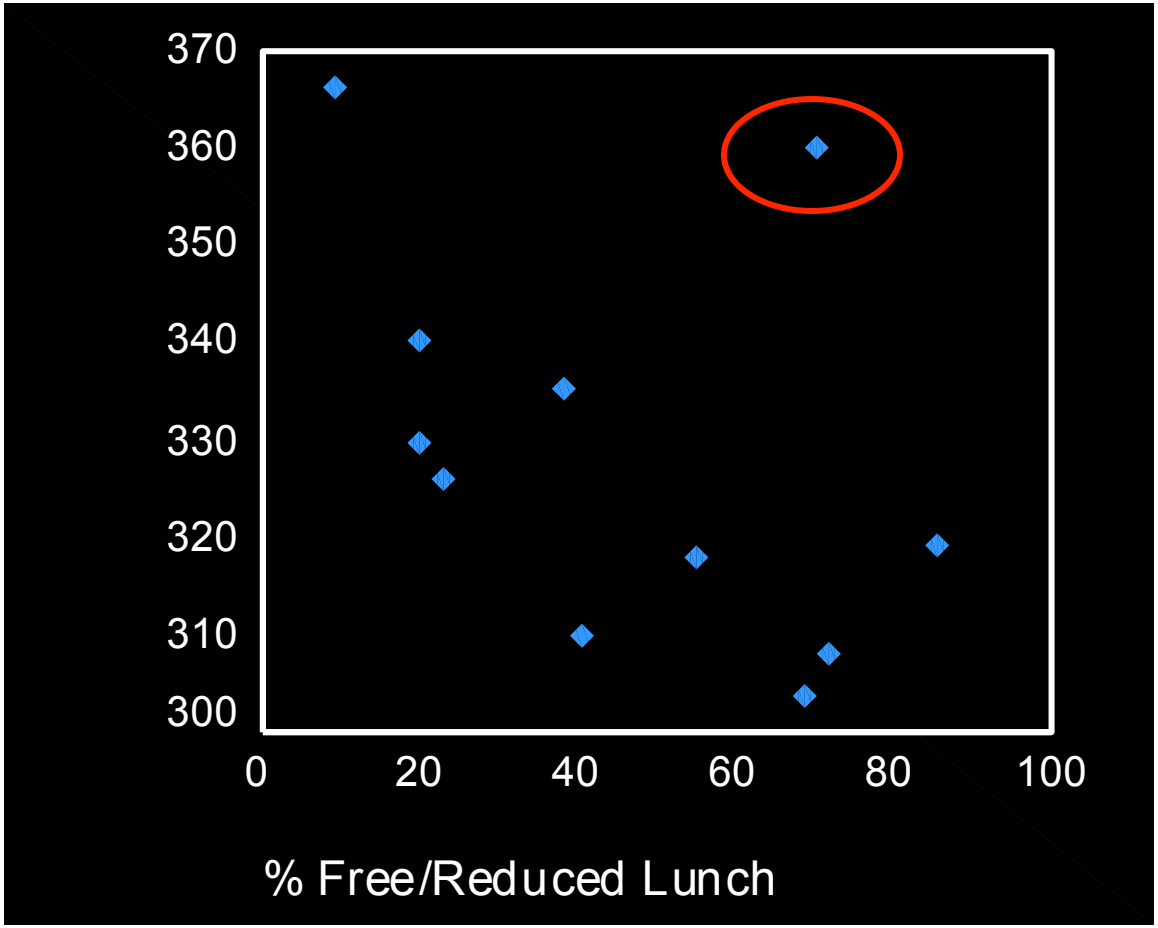
- The hypothesis was tested through the computation of a Pearson Product Moment Correlation between age and satisfaction in an online course. The correlation was not significant (r = -.593, p = .07).

Factors Influencing r

- Outliers
  - Observations that do not fit the overall trend of the relationship
    - Can increase or decrease the value of r
- Non-linear trends
  - Relationships described by a curved line (a polynomial, not linear trend)
    - Will decrease the value of r
- Restriction of range
  - Scores are not obtained along the entire range of the scale for X and/or Y
    - Will decrease the value of r

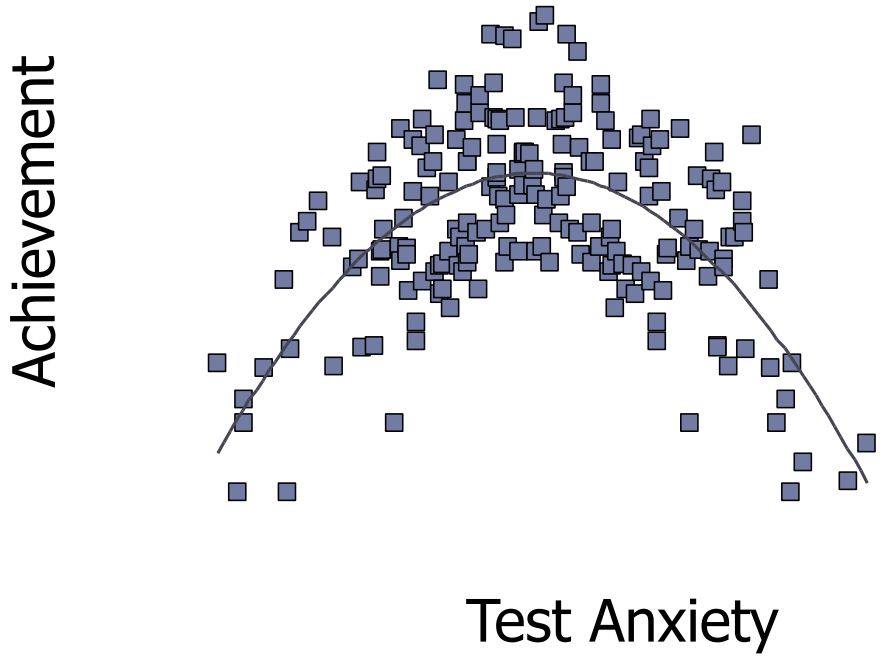
Outlier Example

- The presence of a single outlier that was inconsistent with the trend reduced the r value from -.689 to -.452



Non-Linear Trend Example

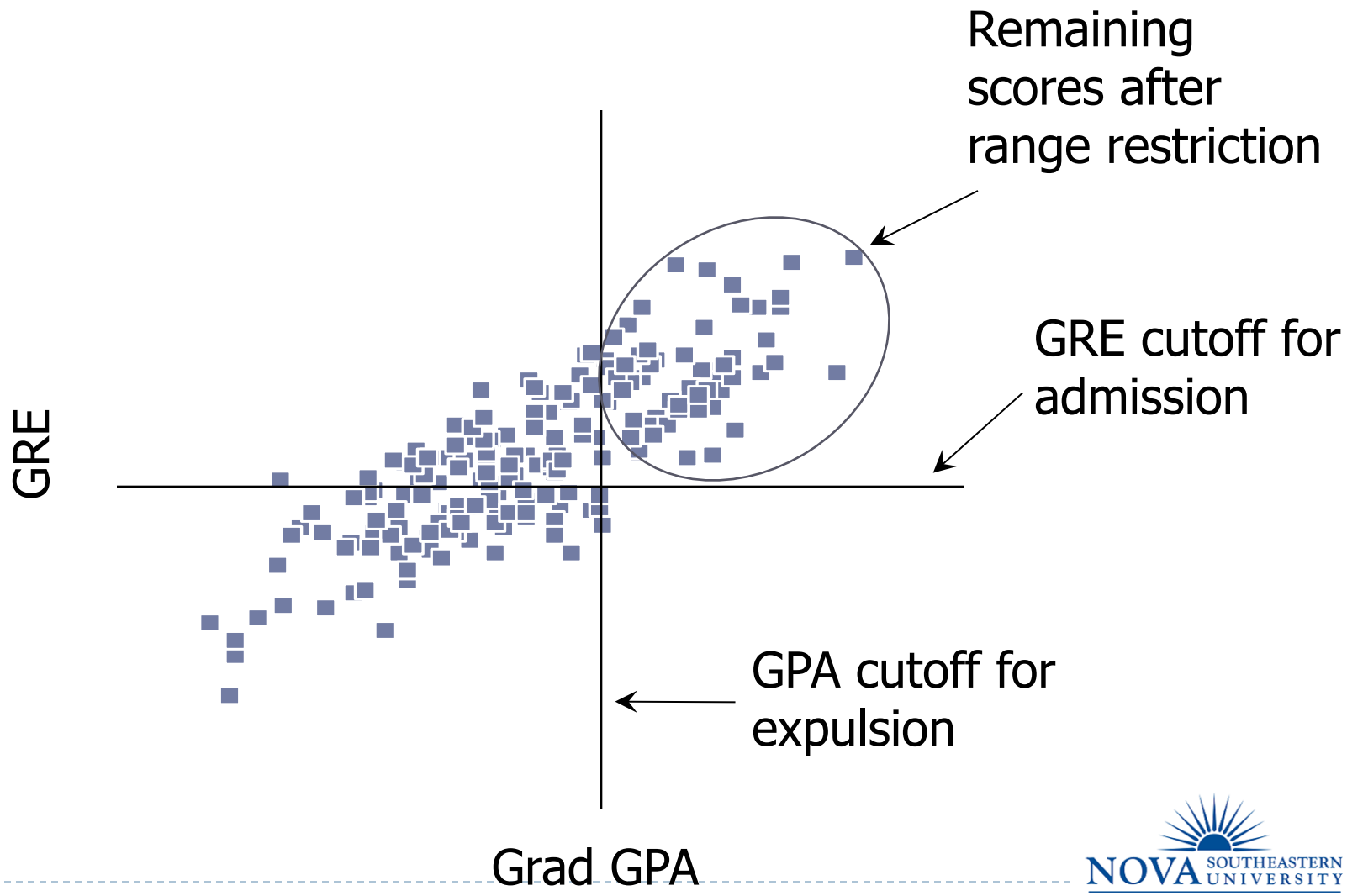
- Some anxiety is good (i.e., in order to reach an "optimal level of arousal")
- Beyond a certain point, an increase in anxiety results in detrimental performance
- r will be near 0 in this case



Restriction of Range Example

- It is known that there is only moderate correlation between the GRE and graduate school GPA, in part, due to range restriction
- In this example, both variables likely have restricted ranges
- Only the highest GRE scores and GPA values are used in the computation of r

Graphic Representation



Coefficient of Determination

- = r^2
- Proportion of variability in one variable explained by variability in the other variable.
- Multiply by 100 to discuss the percentage of explained variability b/n two variables.
- Implication is that r tends to exaggerate how strong the relationship is.

Other Correlation Coefficients

- Point-biserial correlation
  - One continuous and one categorical variable with only two groups
- Spearman's rho
  - At least one variable is ordinal (the other is ordinal or continuous)
- Phi
  - Two dichotomous categorical variables
- Cramer's C (or V)
  - Two categorical variables with any number of categories

Module 5 Summary

- Definition of Correlation
- Relationship Questions
- Scatterplots
- Strength and Direction of Correlations
- Running a Pearson Product Moment Correlation
- Factors Affecting the Correlation
- Other Types of Correlations

Review Activity

- Please complete the review activity at the end of the module.
- All modules build on one another. Therefore, in order to move onto the next module you must successfully complete the review activity before moving on to next module.
- You can complete the review activity and module as many times as you like.

Upcoming Modules

- Module 1: Introduction to Statistics
- Module 2: Introduction to SPSS
- Module 3: Descriptive Statistics
- Module 4: Inferential Statistics
- Module 5: Correlation
- Module 6: *t*-Tests
- Module 7: ANOVAs
- Module 8: Linear Regression
- Module 9: Nonparametric Procedures