

# ***ANOVA and Linear Regression***

***ScWk 242 – Week 13 Slides***

# ANOVA – Analysis of Variance

- Analysis of variance is used to test for differences among more than two populations. It can be viewed as an extension of the t-test we used for testing two population means.
- The specific analysis of variance test that we will study is often referred to as the oneway ANOVA. ANOVA is an acronym for ANalysis Of VAriance. The adjective oneway means that there is a single variable that defines group membership (called a factor). Comparisons of means using more than one variable is possible with other kinds of ANOVA analysis.

# Logic of ANOVA

- The logic of the analysis of variance test is the same as the logic for the test of two population means.
- In both tests, we are comparing the differences among group means to a measure of dispersion for the sampling distribution.
- In ANOVA, differences of group means is computed as the difference for each group mean from the mean for all subjects regardless of group. The measure of dispersion for the sampling distribution is a combination of the dispersion within each of the groups.
- Don't be fooled by the name. ANOVA does not compare variances.



## ANOVA Example 2: Diet vs. Weight Comparisons

<b>Treatment Group</b>	<b>N</b>	<b>Mean weight in pounds</b>
Low Fat	5	150
Normal Fat	5	180
High Fat	5	200
	15	

# Uses of ANOVA

- **The one-way analysis of variance for independent groups applies to an experimental situation where there might be more than two groups. The t-test was limited to two groups, but the Analysis of Variance can analyze as many groups as you want.**
- **Examine the relationship between variables when there is a nominal level independent variable has 3 or more categories and a normally distributed interval/ratio level dependent variable. Produces an F-ratio, which determines the statistical significance of the result.**
- **Reduces the probability of a Type I error (which would occur if we did multiple t-tests rather than one single ANOVA).**

# ANOVA - ASSUMPTIONS & LIMITATIONS

## *Assumptions*

- **NORMALITY ASSUMPTION.**  
The dependent variable can be modeled as a normal population.
- **HOMOGENEITY OF VARIANCE.**  
The dispersion of any populations in our model will be relatively equal.

## *Limitations*

- **The amount of variance for each sample among the dependent variables is relatively equivalent.**

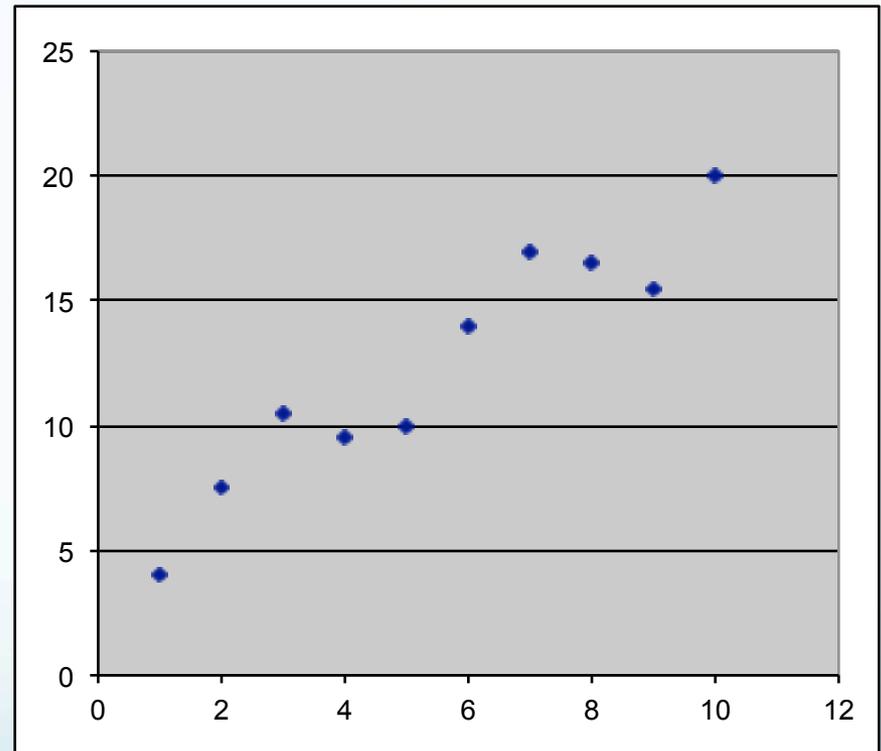
# Linear Regression - Definition

***What is 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  $Y$ .

***Prediction:*** If you know something about  $X$ , this knowledge helps you predict something about  $Y$ .

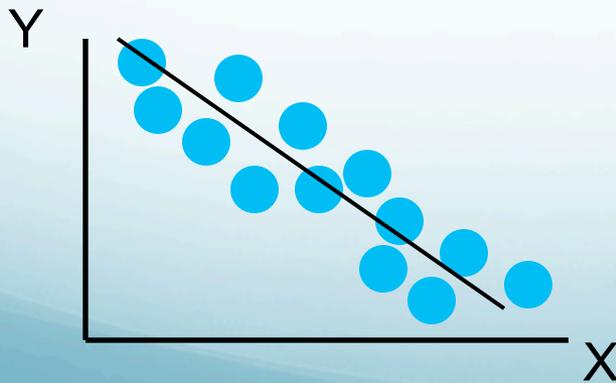
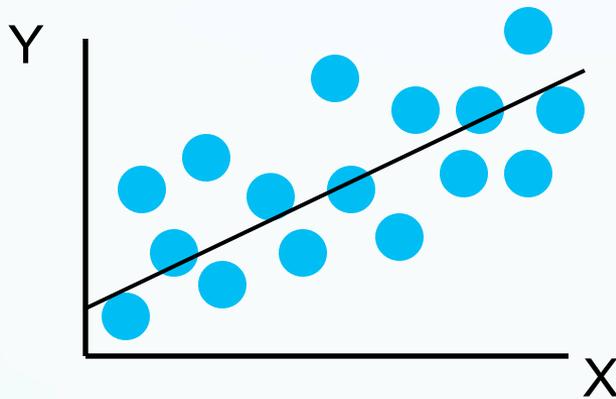
# Linear Regression - Example

- Does there seem to be a linear relationship in the data?
- Is the data perfectly linear?
- Could we fit a line to this data?

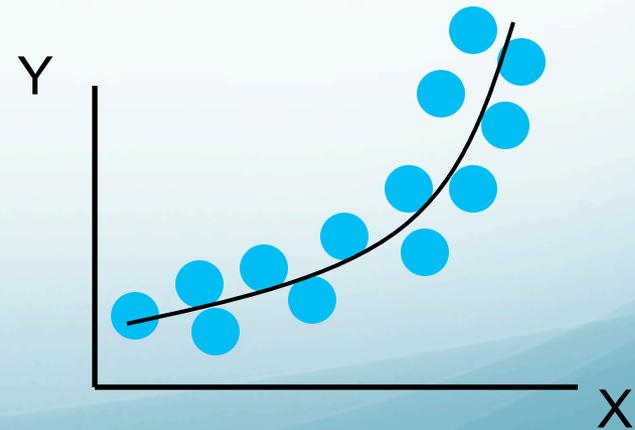
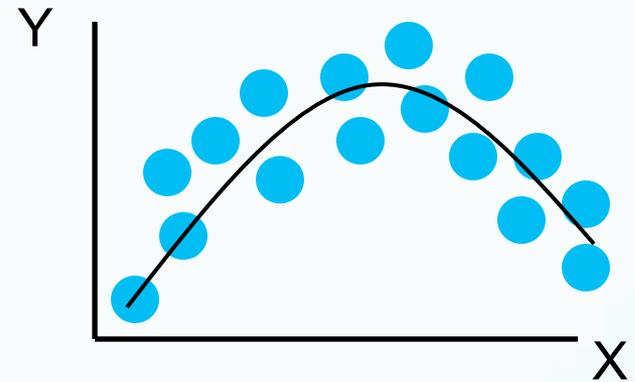


# Linear vs. Curvilinear Relationships

Linear relationships

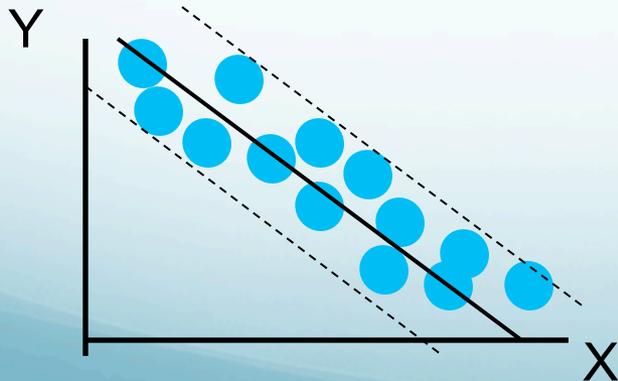
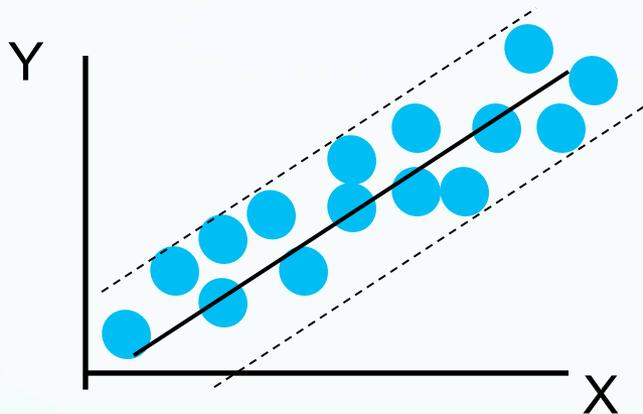


Curvilinear relationships

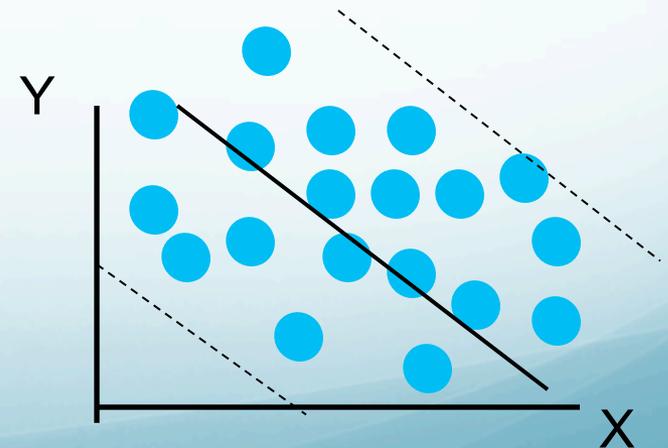
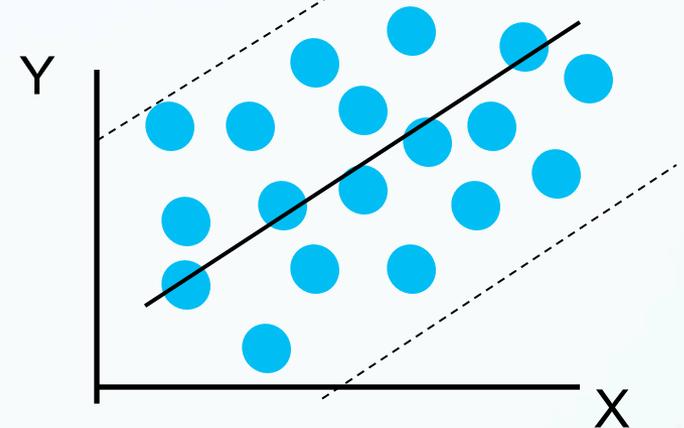


# Strong vs. Weak Linear Correlations

Strong relationships



Weak relationships



# Simple Linear Regression

- **Predicting a criterion value based upon a known predictor(s) value.**
- **Predictor variable (X): what is used as the basis for the prediction (test score, frequency of behavior, amount of something).**
- **Criterion variable (Y): what we want to know (self-esteem, graduate school GPA, violent tendencies).**

# Limitations - Simple Linear Regression

- **Interval or Ratio data only**
- **Can only use predictor values that lie within the existing data range (outliers do not work).**
- **Assume normally distributed values for both the predictor and the criterion variables.**

# Interpreting Results - Linear Regression

- **Know what you are predicting. It should make sense.**
- **Value of prediction is directly related to strength of correlation between the variables. As 'r' decreases, the accuracy of prediction decreases**
- **$Y = 3.5 + 6.8(X)$ , For every unit increase in X, there will be a 6.8 unit increase in Y. The client's education (X) and assertiveness level (Y) for each 1 year increase in a client's education level, her assertiveness level will increase by 6.8 points.**

# Multivariate Analysis

- So far we have tended to concentrate on two-way relationship (such as chi-square and t-tests). But we have started to look at about three-way relationships.
- Social relationships and phenomena are usually more complex than is allowed for in only a bivariate analysis.
- Multivariate analyses are thus commonly used as a reflection of this complexity.

# Multivariate Analysis - Summary

- Multivariate analyses can utilize a variety of techniques (depending on the form of the data, research questions to be addressed, etc., in order to determine whether the relationship between two variables persists or is altered when we ‘control for’ a third (or fourth, or fifth...) variable.
- Multivariate analysis can also enable us to establish which variable(s) has/have the greatest impact on a dependent variable – e.g. *Is ‘sex’ more important than ‘race’ in determining income?*
- It is often important for a multivariate analysis to check for *interactions* between the effects of independent variables, as discussed earlier under the heading of specification.