Plan:

1. Define Simple linear regression
2. Explain the underlying assumptions of linear regression

Inferential Analysis: Regression

Shannon E. Ellis, Ph.D
UC San Diego

Department of Cognitive Science
sellis@ucsd.edu

CORRELATION

ASSOCIATION BETWEEN VARIABLES

i.e. Pearson Correlation, Spearman Correlation, chi-square test

COMPARISON OF MEANS

DIFFERENCE IN MEANS BETWEEN VARIABLES

i.e. t-test, ANOVA

REGRESSION

DOES CHANGE IN ONE VARIABLE MEAN CHANGE IN ANOTHER?

I.e. simple regression, multiple regression

NON-PARAMETRIC TESTS

FOR WHEN ASSUMPTIONS IN THESE OTHER 3 CATEGORIES ARE NOT MET

i.e. Wilcoxon rank-sum test, Wilcoxon sign-rank test, sign test

CORRELATION

ASSOCIATION BETWEEN VARIABLES

i.e. Pearson Correlation, Spearman Correlation, chi-square test

COMPARISON OF MEANS

DIFFERENCE IN MEANS BETWEEN VARIABLES

i.e. t-test, ANOVA

REGRESSION

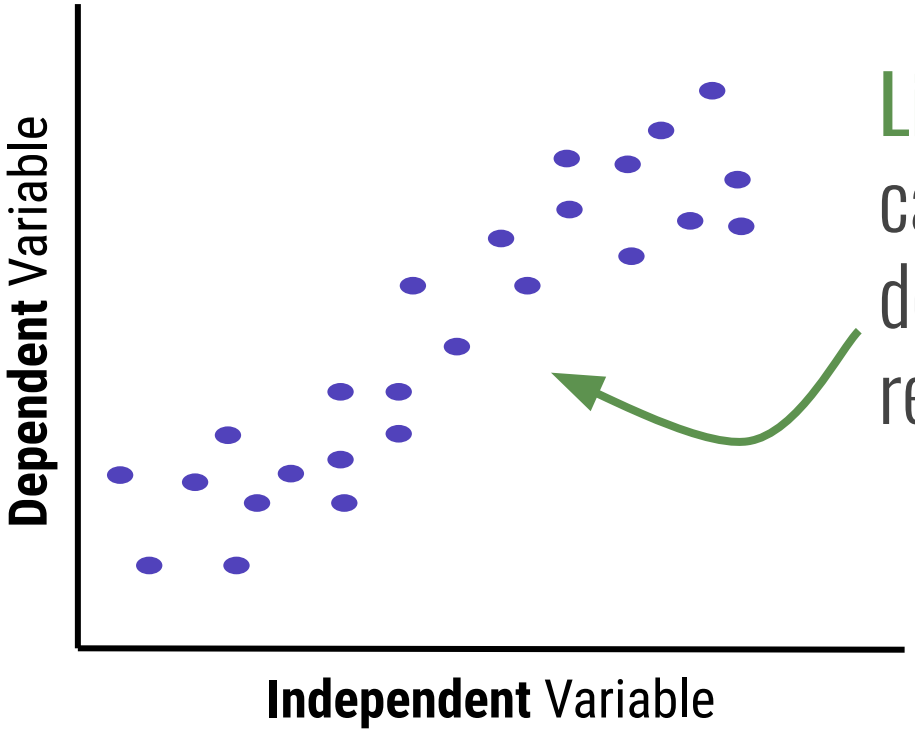
DOES CHANGE IN ONE VARIABLE MEAN CHANGE IN ANOTHER?

I.e. simple regression, multiple regression

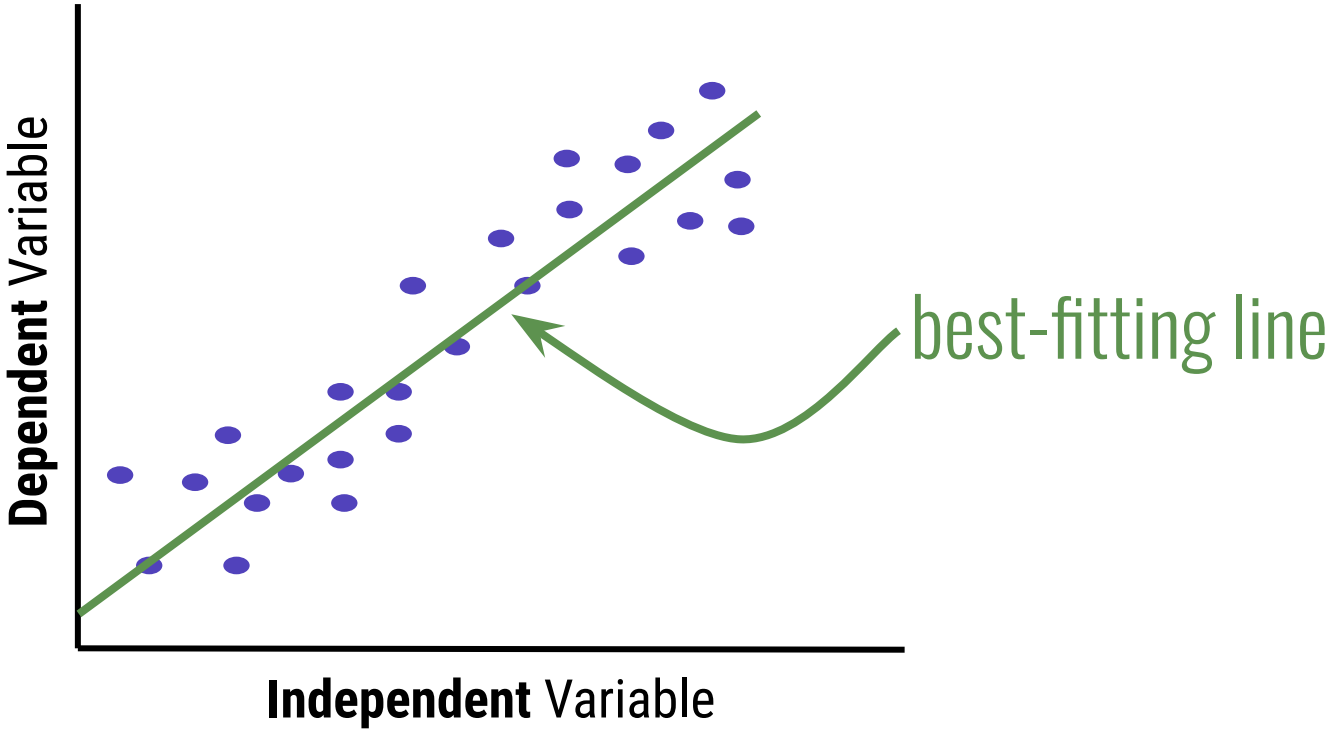
NON-PARAMETRIC TESTS

FOR WHEN ASSUMPTIONS IN THESE OTHER 3 CATEGORIES ARE NOT MET

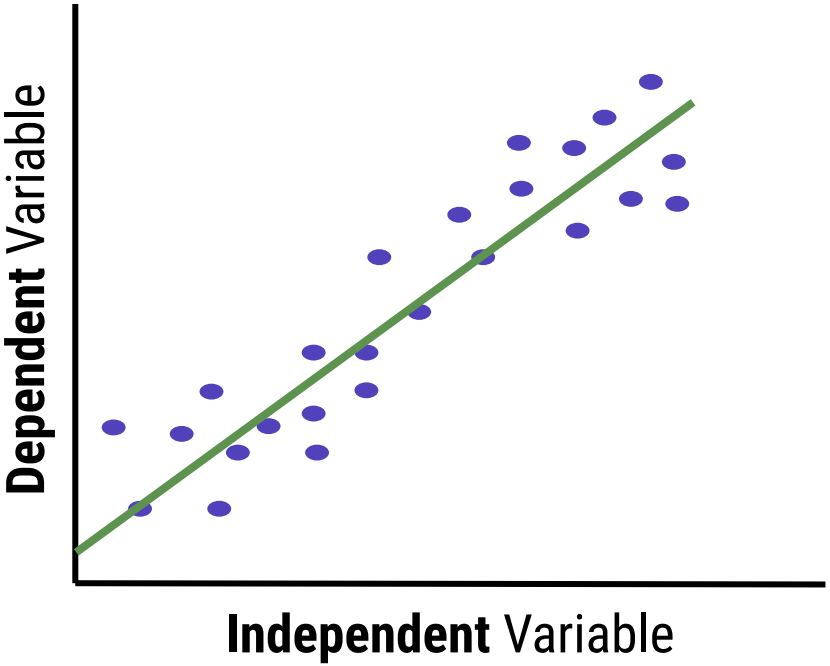
i.e. Wilcoxon rank-sum test, Wilcoxon sign-rank test, sign test



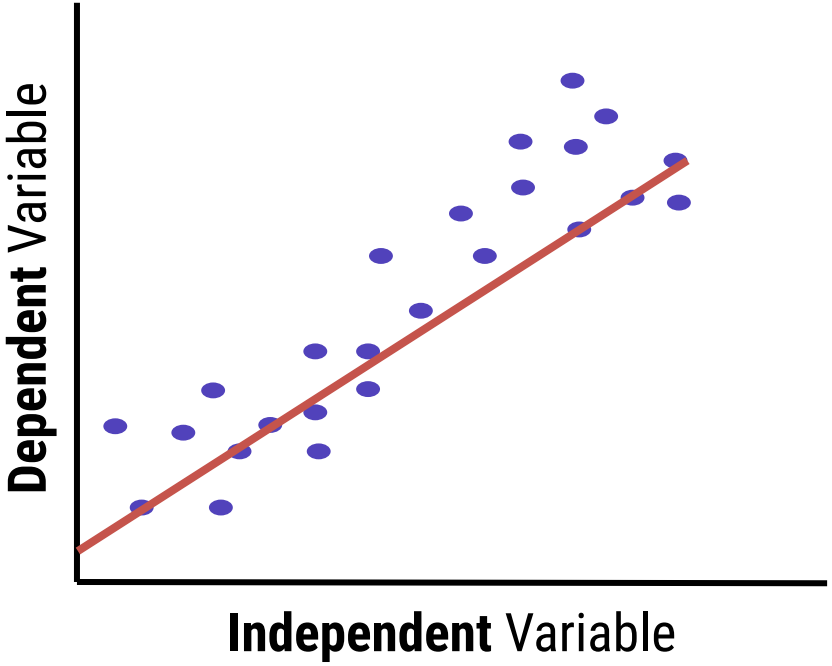
Linear regression can be used to describe this relationship

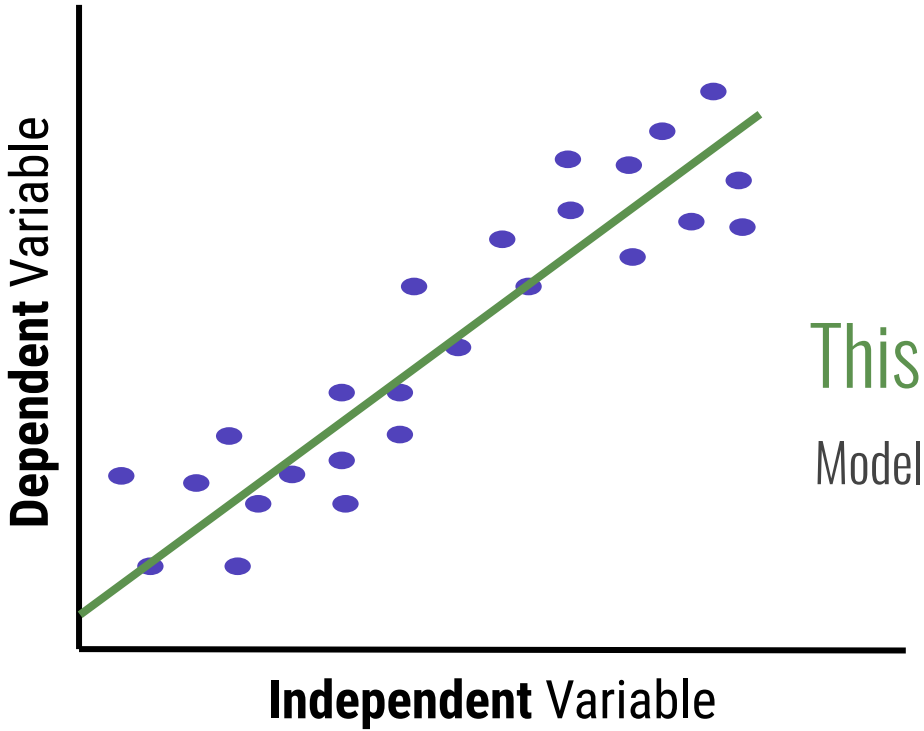

Dependent Variable
Independent Variable
best-fitting line

Best-fitting line



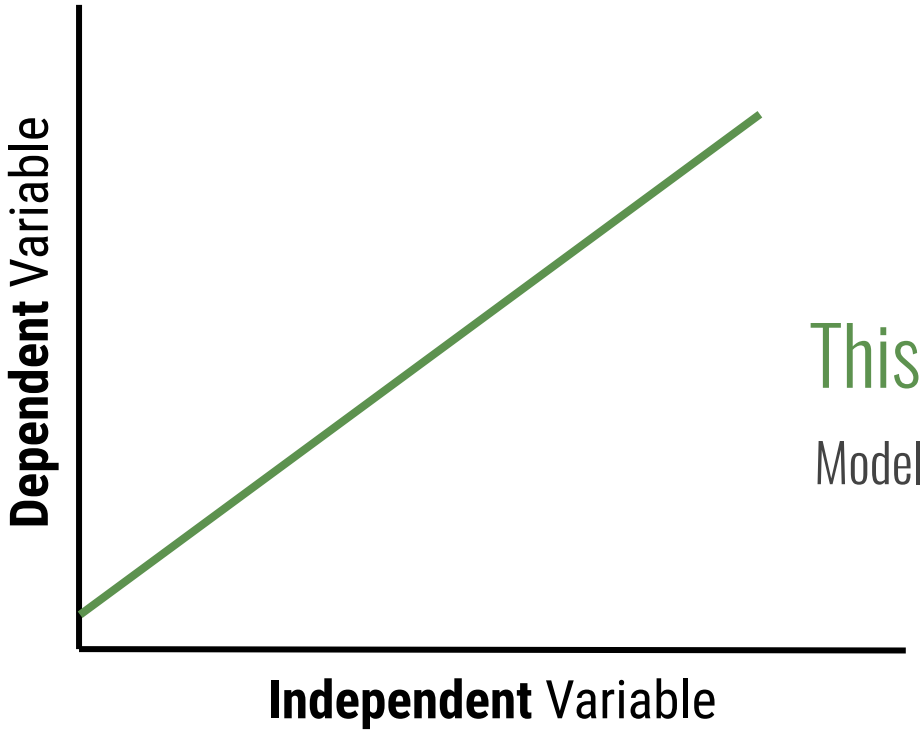
NOT a best-fitting line





This line is a **model** of the data

Models are mathematical equations generated to *represent* the real life situation

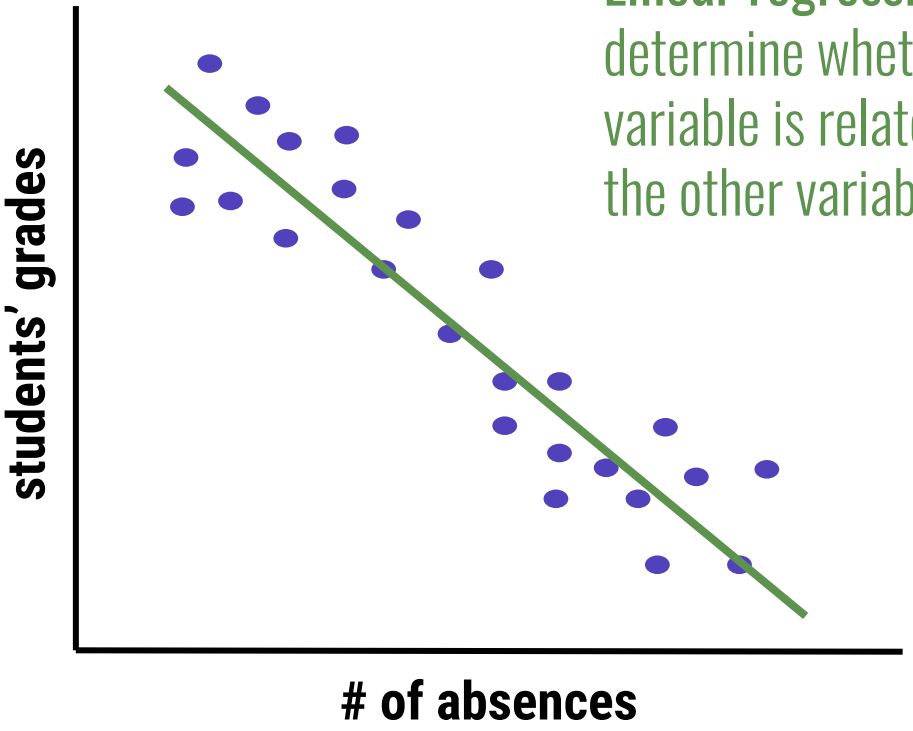


This line is a **model** of the data

Models are mathematical equations generated to *represent* the real life situation

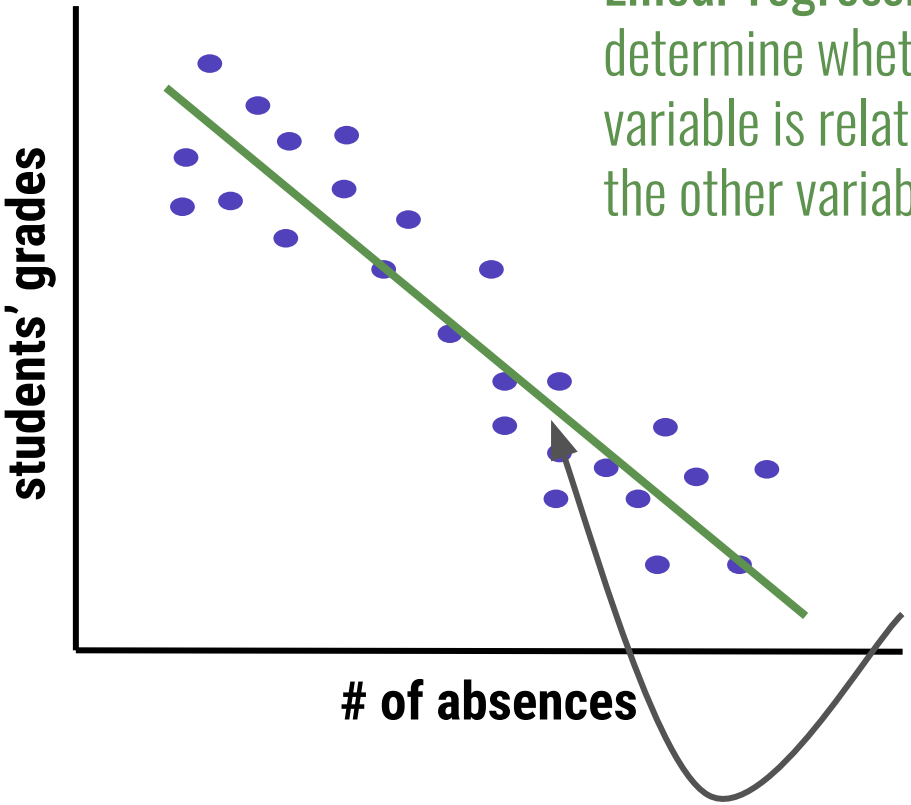
“*All models are wrong, but some are useful*”

-George Box (British Statistician, *JASA* 1976)



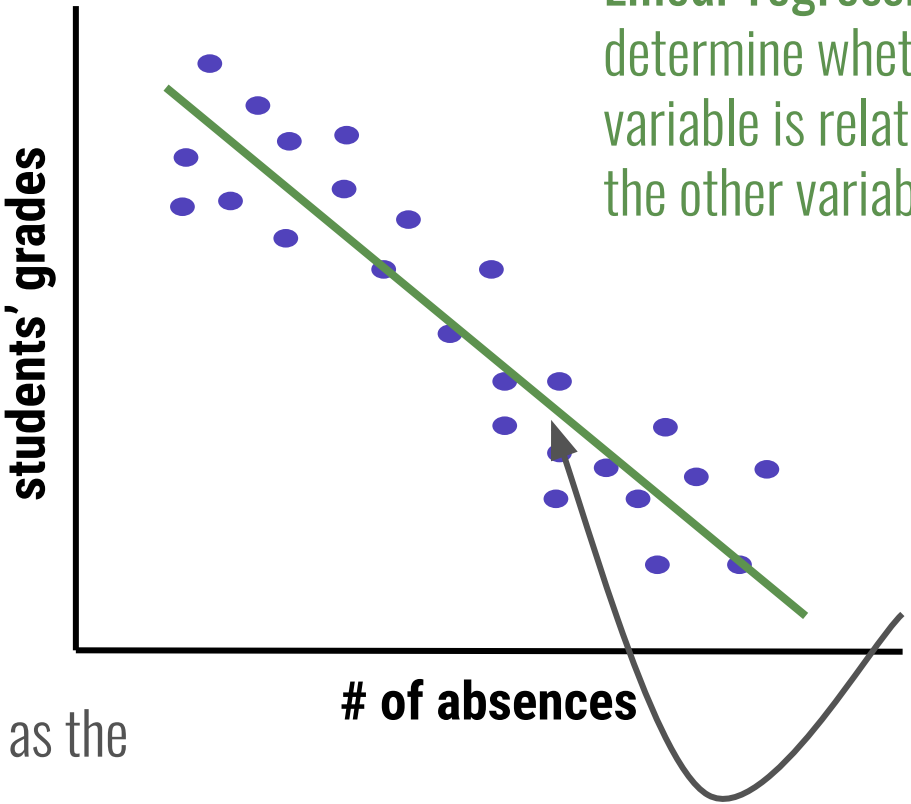
Linear regression can be used to determine whether a change in one variable is related to the change in the other variable

Linear regression can be used to determine whether a change in one variable is related to the change in the other variable



The magnitude of the relationship is measured by the slope of the line

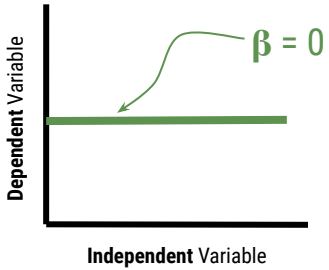
Linear regression can be used to determine whether a change in one variable is related to the change in the other variable



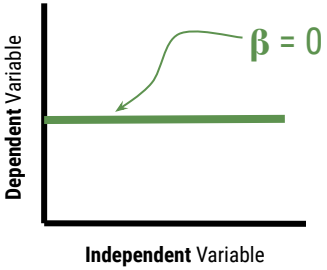
The magnitude of the relationship is measured by the slope of the line

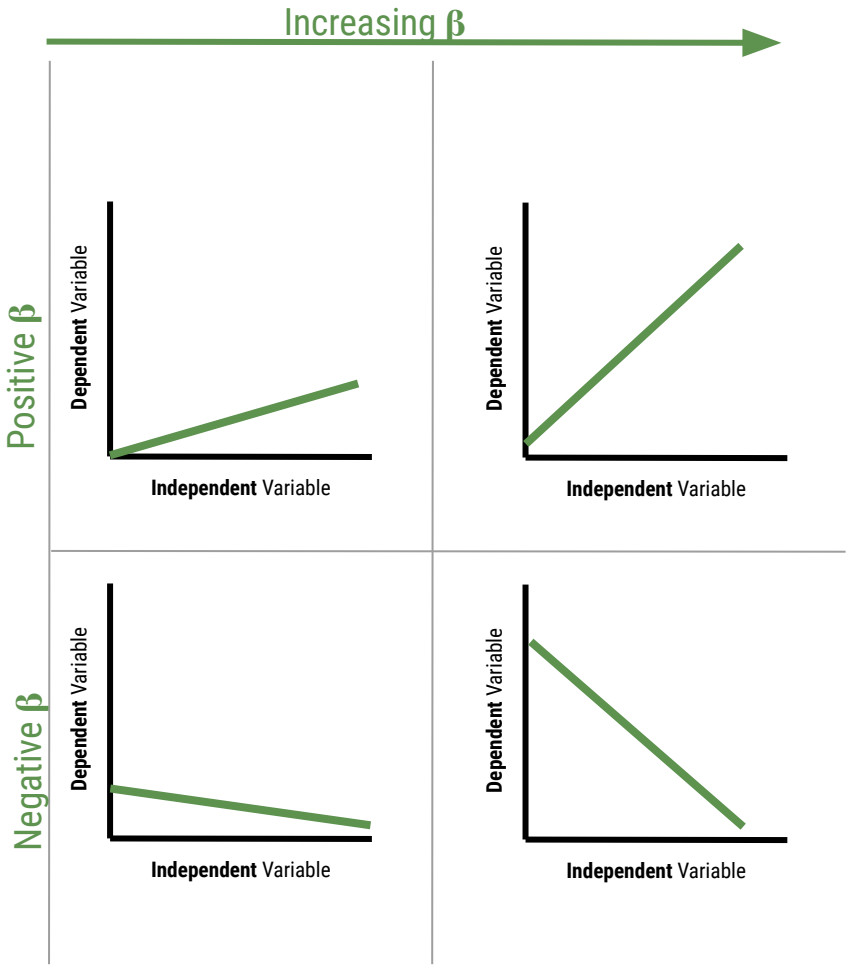
This is also referred to as the model's effect size (β)

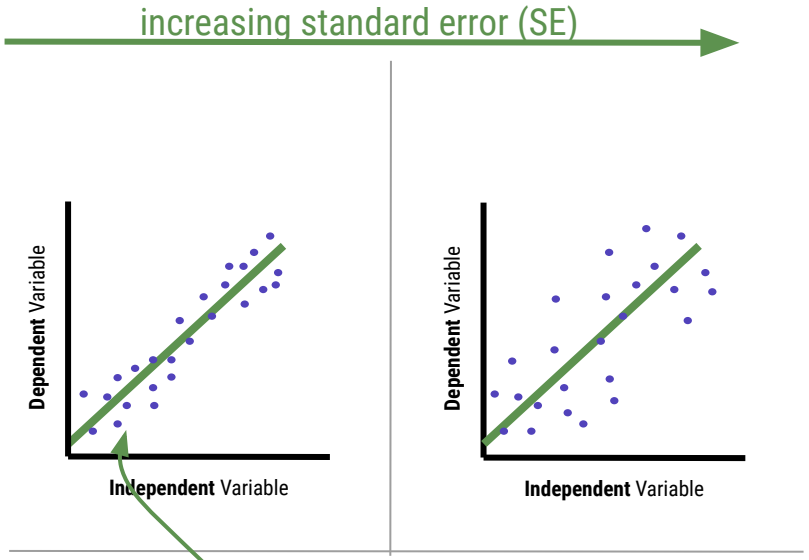
Effect size (β) can be estimated using the slope of the line



Effect size (β) can be estimated using the slope of the line



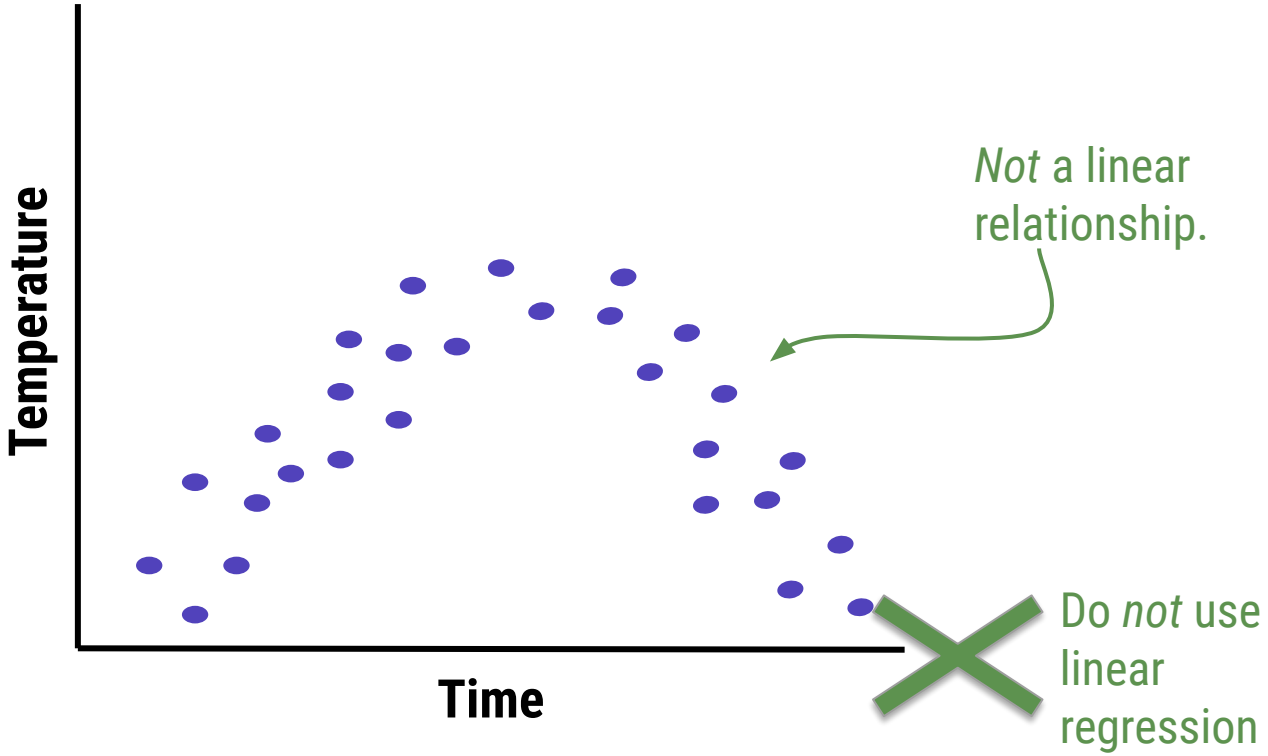




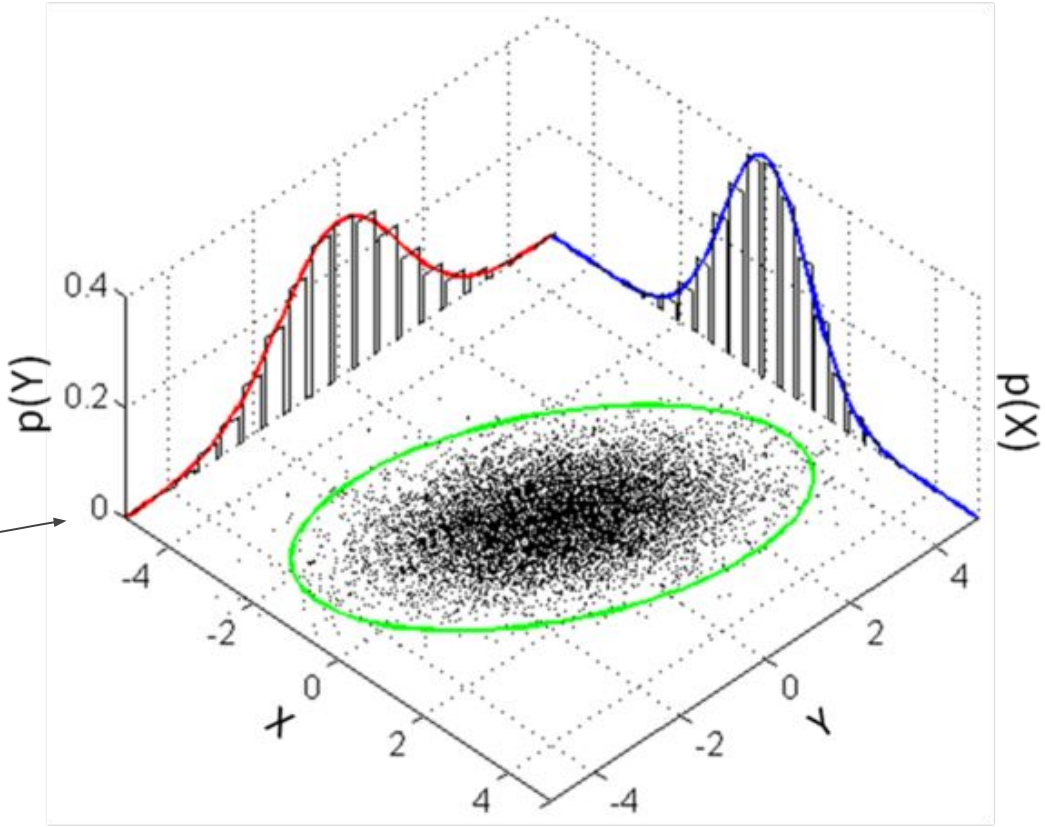
The *closer* the points are to the regression line, the *less uncertain* we are in our estimate

Assumptions of linear regression

1. Linear relationship
2. Multivariate normality
3. No multicollinearity
4. No auto-correlation
5. Homoscedasticity

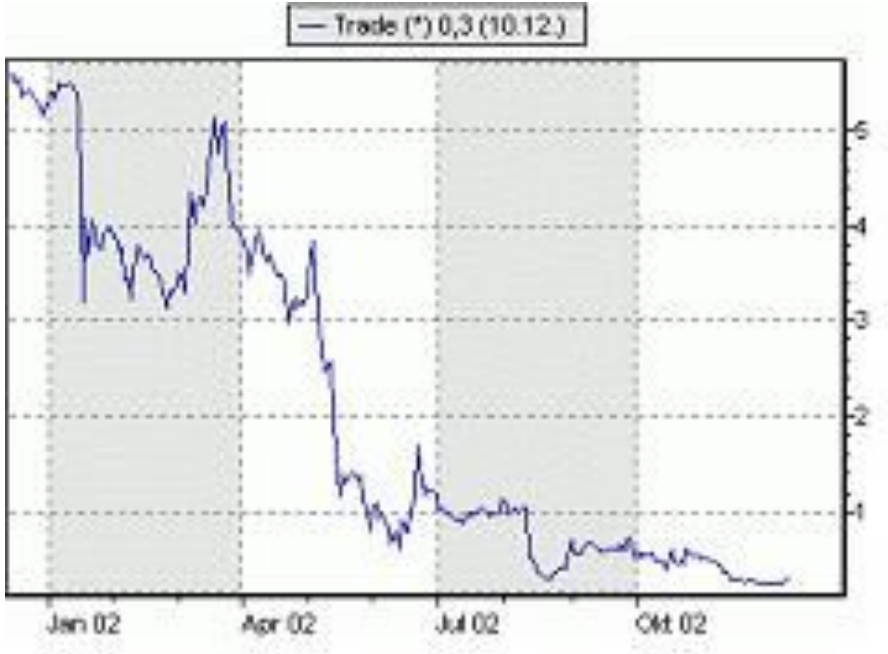

Temperature
Time
Not a linear relationship.
Do *not* use linear regression

A multivariate normal probability distribution (joint normal)

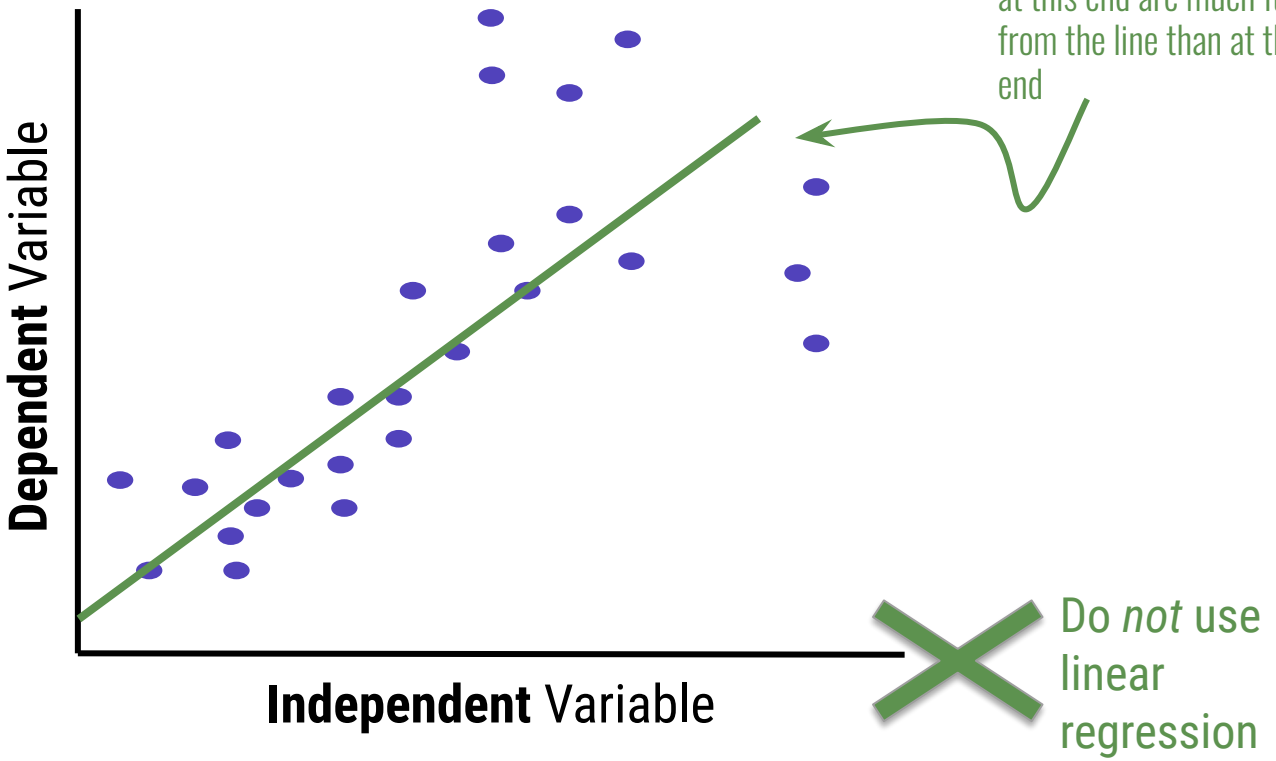


Linear regression assumes no multicollinearity. **Multicollinearity** occurs when the independent variables (in multiple linear regression) are too highly correlated with each other.

Autocorrelation occurs when the observations are *not* independent of one another (i.e. stock prices)



https://www.statisticssolutions.com/assumptions-of-linear-regression/


Not homoscedastic: points at this end are much further from the line than at the other end
Dependent Variable
Independent Variable
Do *not* use linear regression