Hypothesis Testing Example

Test the claim that the true mean # of TV sets in US homes is equal to 3. (Assume $\sigma = 0.8$)

- State the appropriate null and alternative hypotheses
  - $H_0: \mu = 3$, $H_1: \mu \neq 3$ (This is a two tailed test)
- Specify the desired level of significance
  - Suppose that $\alpha = .05$ is chosen for this test
- Choose a sample size
  - Suppose a sample of size $n = 100$ is selected

Hypothesis Testing Example

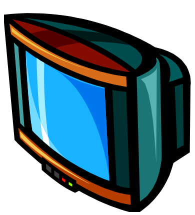
(continued)

- Determine the appropriate technique
  - σ is known so this is a z test
- Set up the critical values
  - For $\alpha = .05$ the critical z values are ± 1.96
- Collect the data and compute the test statistic
  - Suppose the sample results are

    $n = 100$, $\bar{x} = 2.84$ ($\sigma = 0.8$ is assumed known)

So the test statistic is:

$$z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{2.84 - 3}{\frac{0.8}{\sqrt{100}}} = \frac{-.16}{.08} = -2.0$$

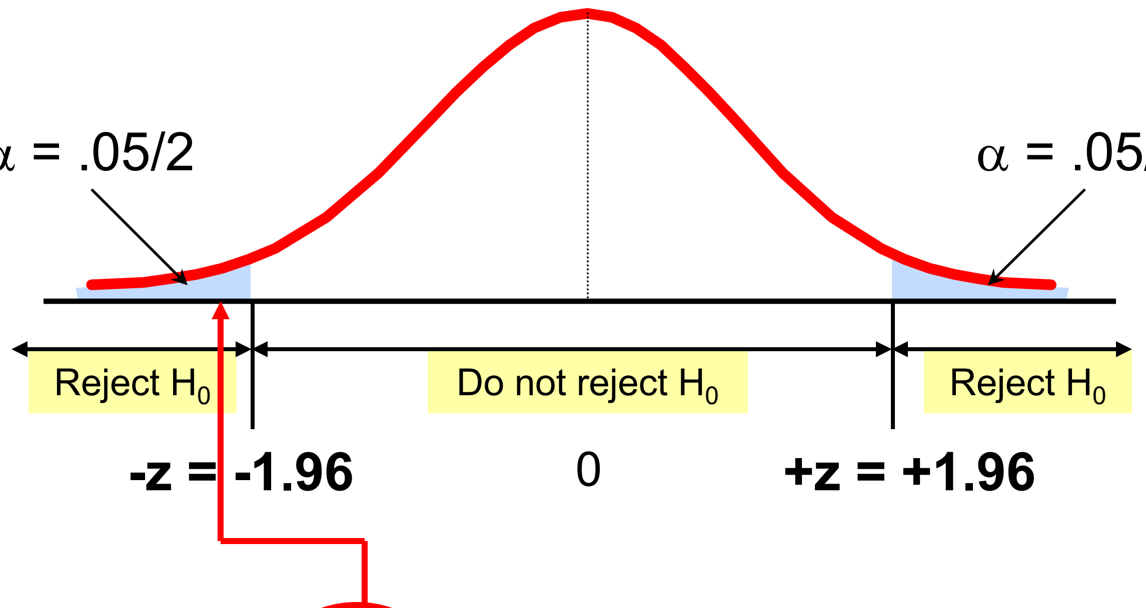


Hypothesis Testing Example

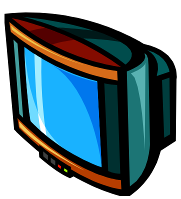
(continued)

- Is the test statistic in the rejection region?

Reject H_0 if $z < -1.96$ or $z > 1.96$; otherwise do not reject H_0



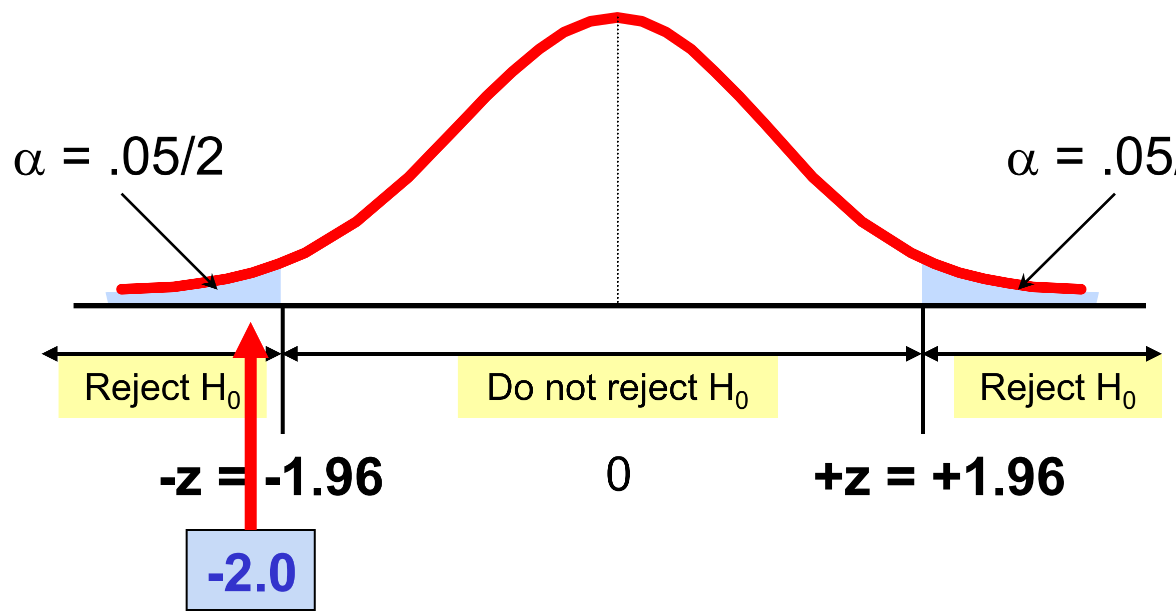
Here, $z = -2.0 < -1.96$, so the test statistic is in the rejection region



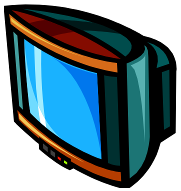
Hypothesis Testing Example

(continued)

- Reach a decision and interpret the result



Since z = -2.0 < -1.96, we reject the null hypothesis and conclude that there is sufficient evidence that the mean number of TVs in US homes is not equal to 3



Example: p-Value

- Example: How likely is it to see a sample mean of 2.84 (or something further from the mean, in either direction) if the true mean is μ = 3.0?

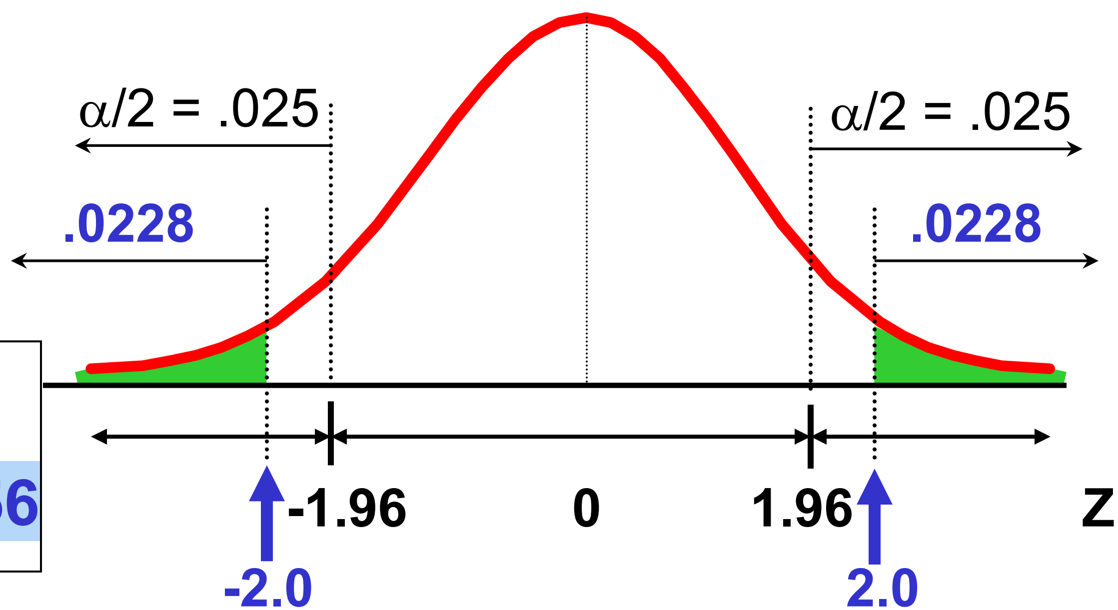
$\bar{x}$ = 2.84 is translated to a z score of z = -2.0

$$P(z < -2.0) = .0228$$

$$P(z > 2.0) = .0228$$

p-value

= .0228 + .0228 = .0456



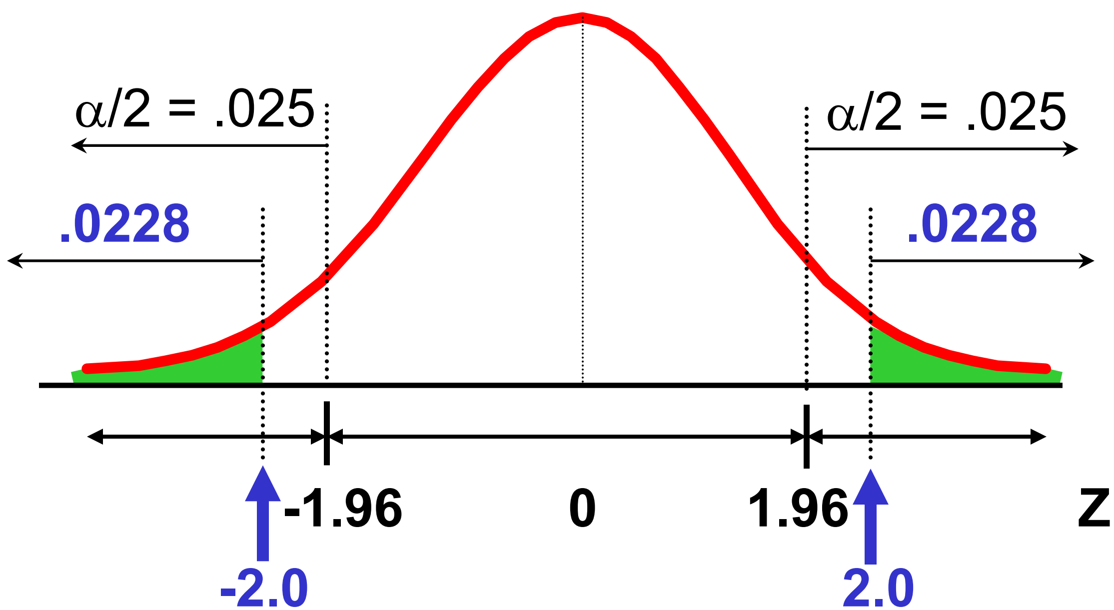
Example: p-Value

(continued)

- Compare the p-value with α
  - If p-value $<$ α , reject H_0
  - If p-value $\geq$ α , do not reject H_0

Here: p-value = .0456
α = .05

Since .0456 < .05, we reject the null hypothesis



An Alternative Way

Beer Consumption [redacted] Human Attractiveness to Malaria Mosquitoes

Beer (25):

27 20 21 26 27 31 24 21 20 19
23 24 28 19 24 29 18 20 17 31
20 25 28 21 27

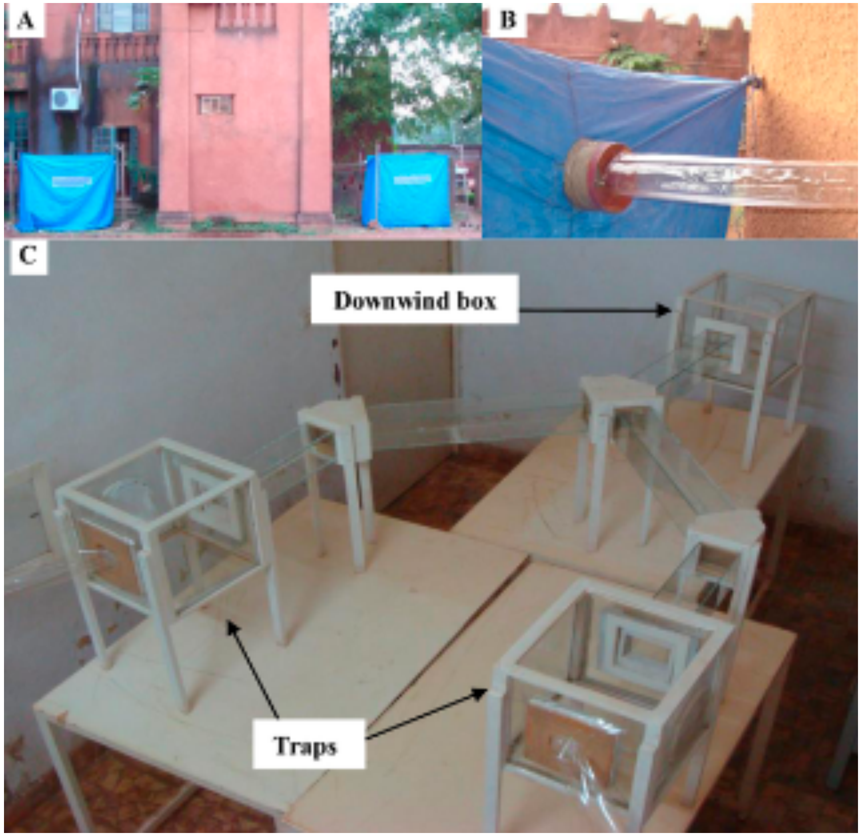
Mean: 23.6

Water (18):

21 22 15 12 21 16 19 15 22 24
19 23 13 22 20 24 18 20

Mean: 19.2

Is a difference of 4.4 significant?



Beer Consumption #Human Attractiveness to Malaria Mosquitoes
Thierry Lefèvre , Louis-Clément Gouagna, Kounbobr Roch Dabiré, Eric Elguero, Didier Fontenille, François Renaud, Carlo Costantini, Frédéric Thomas

Permutation Test

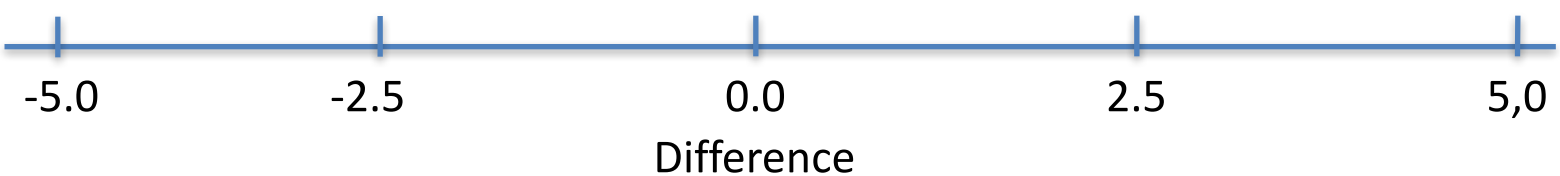
Beer (25)

27	23	20	31	29
20	24	25	24	18
21	28	28	21	20
26	19	21	20	17
27	24	27	19	31

Water (18)

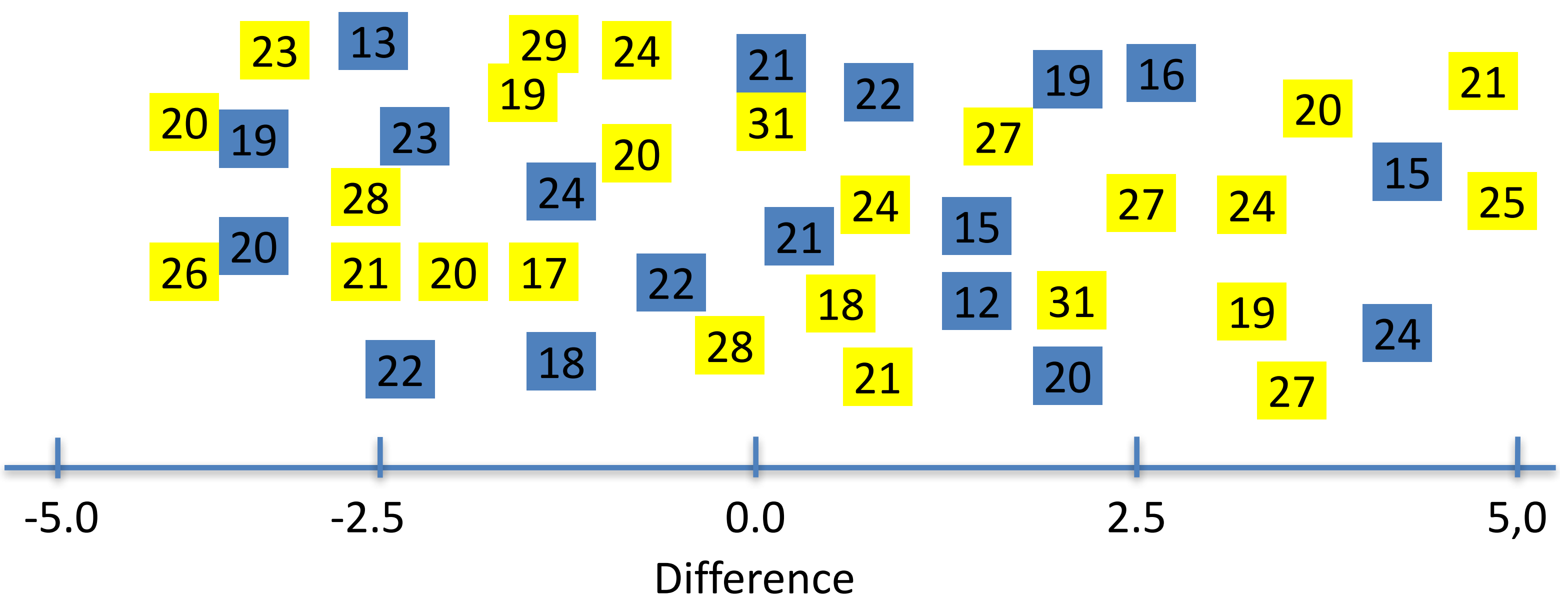
21	19	16	24
22	23	19	18
15	13	15	20
12	22	22	
21	20	24	

Difference: 4.4

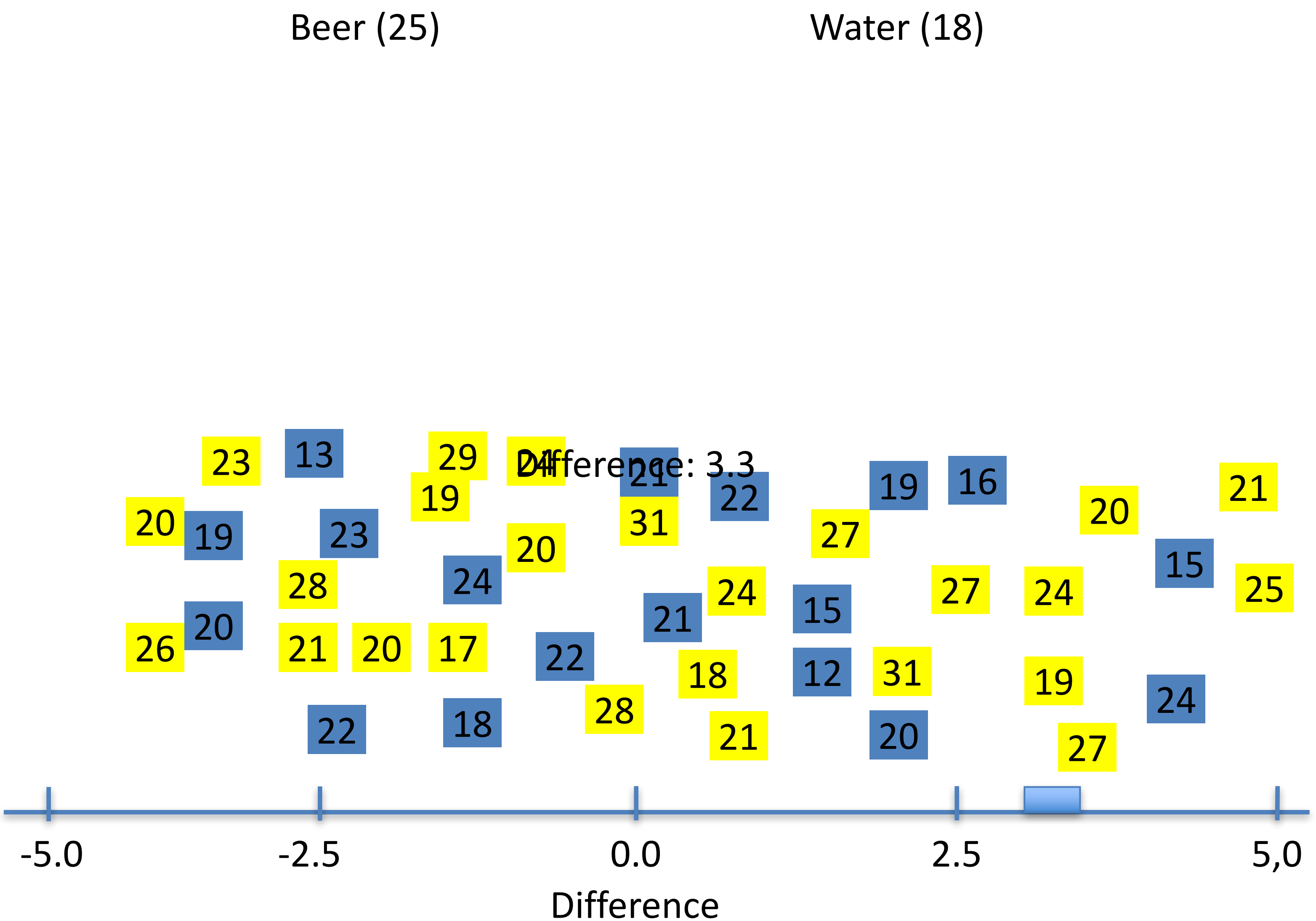


Permutation Test

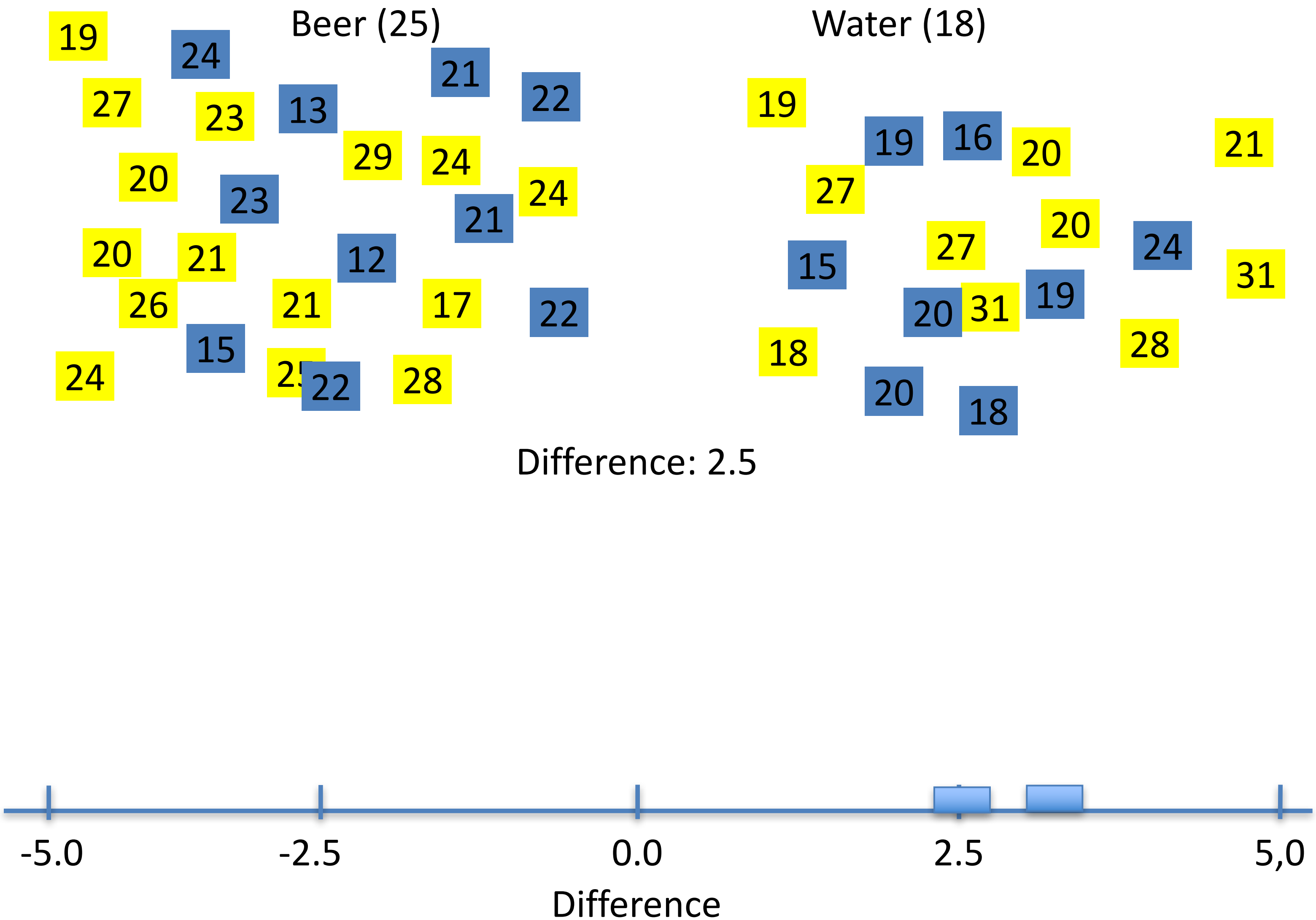
Beer (25) Water (18)



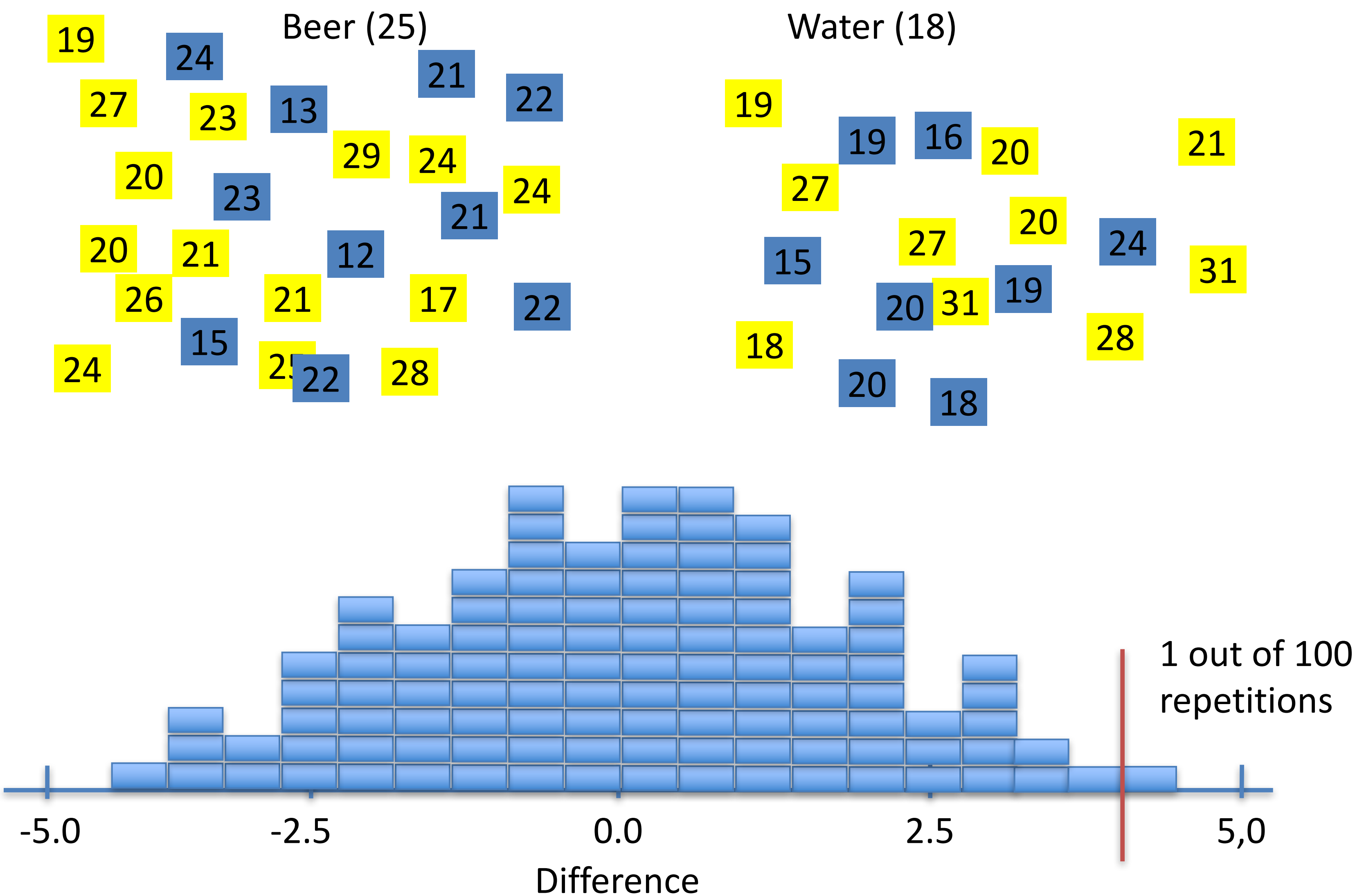
Permutation Test



Permutation Test



Permutation Test



t Test of Hypothesis for the Mean (σ Unknown)

- Convert sample result ($\bar{x}$) to a t test statistic

Hypothesis Tests for μ

σ Known | **σ Unknown**

Consider the test

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

(Assume the population is normal)

The decision rule is:

$$\text{Reject } H_0 \text{ if } \; t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} > t_{n-1,\alpha}$$

t Test of Hypothesis for the Mean (σ Unknown)

(continued)

- For a two-tailed test:

  Consider the test

  $$H_0 : \mu = \mu_0$$
  $$H_1 : \mu \neq \mu_0$$

  (Assume the population is normal, and the population variance is unknown)

  The decision rule is:

  Reject H_0 if $t = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} < -t_{n-1,\, \alpha/2}$ or if $t = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} > t_{n-1,\, \alpha/2}$

What is a t-distribution?

- A t-distribution is like a Z distribution, except has slightly fatter tails to reflect the uncertainty added by estimating σ.
- The bigger the sample size (i.e., the bigger the sample size used to estimate σ), then the closer t becomes to Z.
- Invented in 1908 by William Sealy Gosset, a chemist working for the Guinness who was not allowed to publish – so published as "student" t-test.

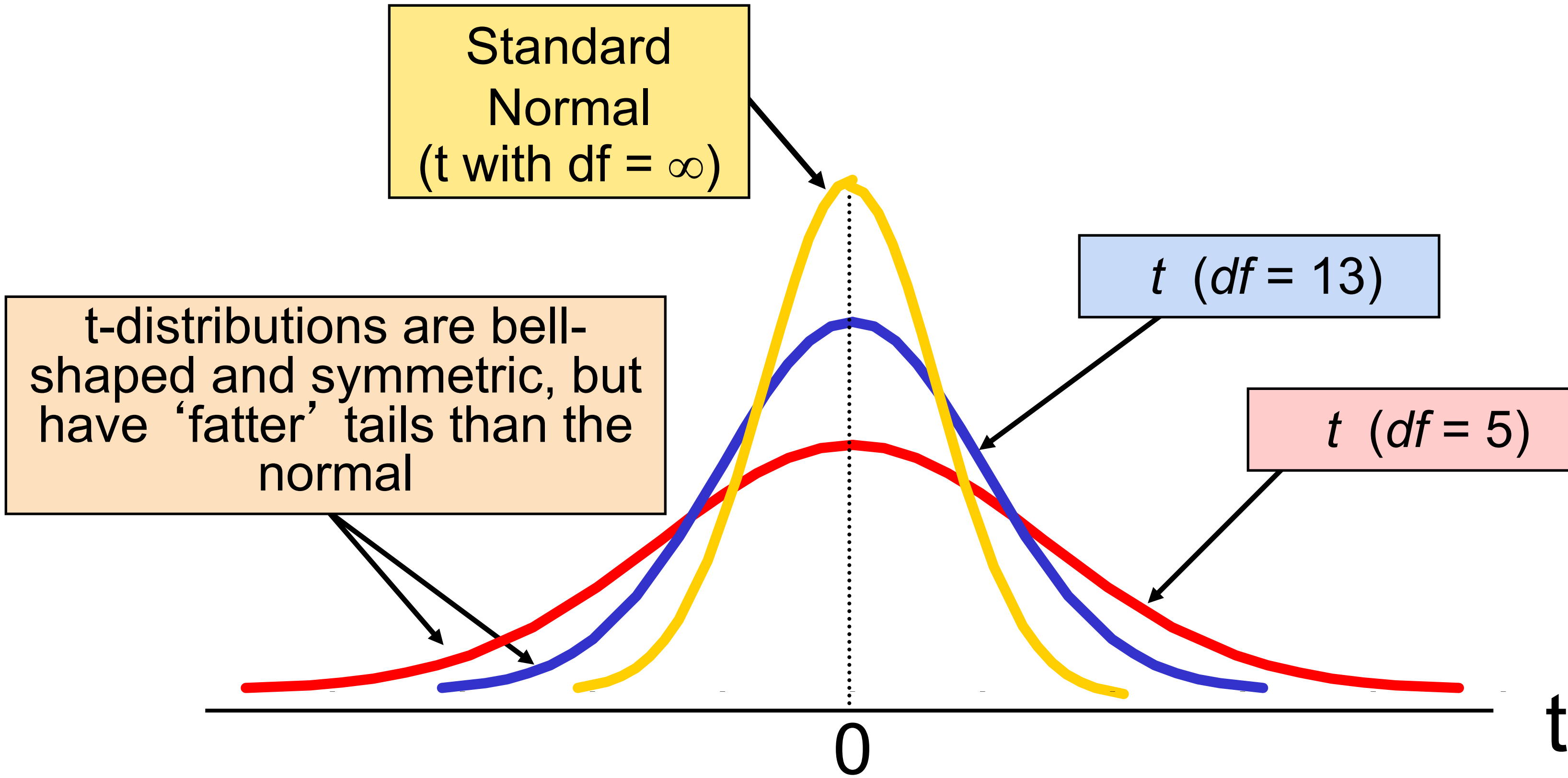
The t-distribution density function

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Where:
v is the degrees of freedom
Γ (gamma) is the Gamma function
is the constant Pi (3.14...)

Student's t Distribution

Note: t ⟶ Z as n increases



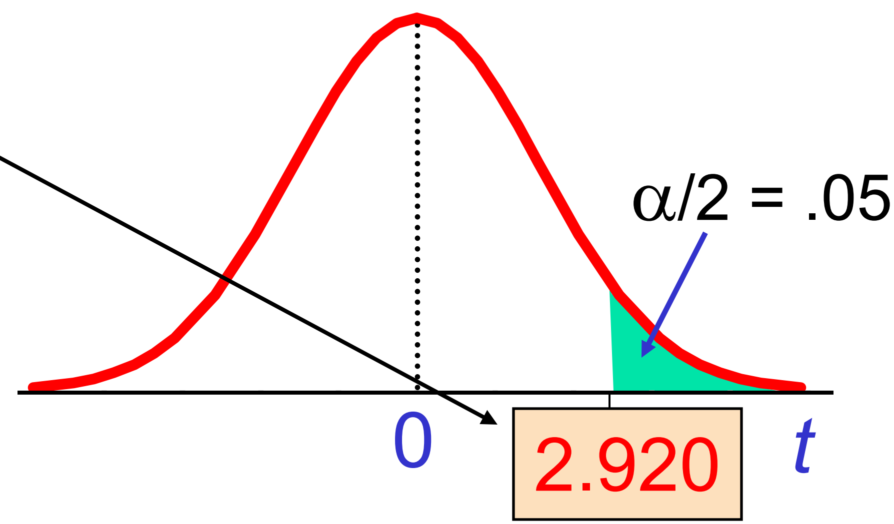
from "Statistics for Managers" Using Microsoft® Excel 4th Edition, Prentice-Hall 2004

Student’s t Table

	Upper Tail Area		
df	.25	.10	.05
1	1.000	3.078	6.314
2	0.817	1.886	2.920
3	0.765	1.638	2.353

The body of the table contains t values, not probabilities

Let: n = 3
df = n - 1 = 2
α = .10
$\alpha/2$ =.05



from “Statistics for Managers” Using Microsoft® Excel 4th Edition, Prentice-Hall 2004

t distribution values

With comparison to the Z value

Confidence Level	t (10 d.f.)	t (20 d.f.)	t (30 d.f.)	Z
.80	1.372	1.325	1.310	1.28
.90	1.812	1.725	1.697	1.64
.95	2.228	2.086	2.042	1.96
.99	3.169	2.845	2.750	2.58

Note: $t \longrightarrow Z$ as n increases

from "Statistics for Managers" Using Microsoft® Excel 4th Edition, Prentice-Hall 2004

P-Value Has Problems!

BASIC AND APPLIED SOCIAL PSYCHOLOGY, 37:1–2, 2015

ISSN: 0197-3533 print/1532-4834 online
DOI: 10.1080/01973533.2015.1012991

Routledge
Taylor & Francis Group

Editorial

David Trafimow and Michael Marks
New Mexico State University

The *Basic and Applied Social Psychology* (BASP) 2014 Editorial emphasized that the null hypothesis significance testing procedure (NHSTP) is invalid, and thus authors would be not required to perform it (Trafimow, 2014). However, to allow authors a grace period, the Editorial stopped short of actually banning the NHSTP. The purpose of the present Editorial is to announce that the grace period is over. From now on, BASP is banning the NHSTP.

With the banning of the NHSTP from BASP, what are the implications for authors? The following are anticipated questions and their corresponding answers.

Question 1. *Will manuscripts with p-values be desk rejected automatically?*

Answer to Question 1. No. If manuscripts pass the

a strong case for rejecting it, confidence intervals do not provide a strong case for concluding that the population parameter of interest is likely to be within the stated interval. Therefore, confidence intervals also are banned from BASP.

Bayesian procedures are more interesting. The usual problem with Bayesian procedures is that they depend on some sort of Laplacian assumption to generate numbers where none exist. The Laplacian assumption is that when in a state of ignorance, the researcher should assign an equal probability to each possibility. The problems are well documented (Chihara, 1994; Fisher, 1973; Glymour, 1980; Popper, 1983; Suppes, 1994; Trafimow, 2003, 2005, 2006). However, there have been Bayesian proposals that at least somewhat circumvent

The p value is the probability to obtain an effect equal to or more extreme than the one observed presuming the null hypothesis is true

It's not the probability that the null or the alternative hypothesis are correct or incorrect

Misconception 1

"In my experience teaching many academic physicians, when physicians are presented with a single-sentence summary of a study that produced a surprising result with *P* = 0.05, the overwhelming majority will confidently state that there is a 95% or greater chance that the null hypothesis is incorrect.

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999;130:995-1004.

“In my experience teaching many academic physicians, when physicians are presented with a single-sentence summary of a study that produced a surprising result with *P* = 0.05, the overwhelming majority will confidently state that there is a 95% or greater chance that the null hypothesis is incorrect.

This is an understandable but categorically wrong interpretation because the *P* value is calculated on the assumption that the null hypothesis is true. It cannot, therefore, be a direct measure of the probability that the null hypothesis is false. This logical error reinforces the mistaken notion that the data alone can tell us the probability that a hypothesis is true. “

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999;130:995-1004.

Misconception #1

“If P=.05, the null hypothesis has only a 5% chance of being true”

Let us suppose we flip a penny four times and observe four heads, two-sided P =.125. This does not mean that the probability of the coin being fair is only 12.5%.

Steven Goodman: “A Dirty Dozen: Twelve P-Value Misconceptions”

Misconception #2

A non significant difference (eg, P .05) means there is no difference between groups.

- A non significant difference only means the null effect is statistically consistent with the observation
- It does not make the null effect most likely
- In fact, the observed effect best explains the effect regardless the significance.

Steven Goodman: "A Dirty Dozen: Twelve P-Value Misconceptions"

Misconception #3

A statistically significant finding is (clinical) important

The P value carries no information about the magnitude of an effect, which is captured by the effect estimate and confidence interval.

Misconception #4

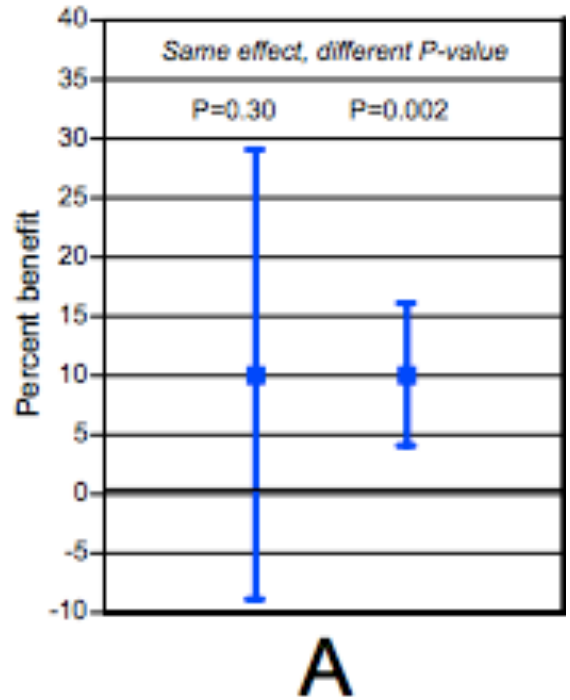
“Studies with P values on opposite sides of .05 are conflicting”

H_0:Drug T has no effect

H_1:Drug T has a positive effect

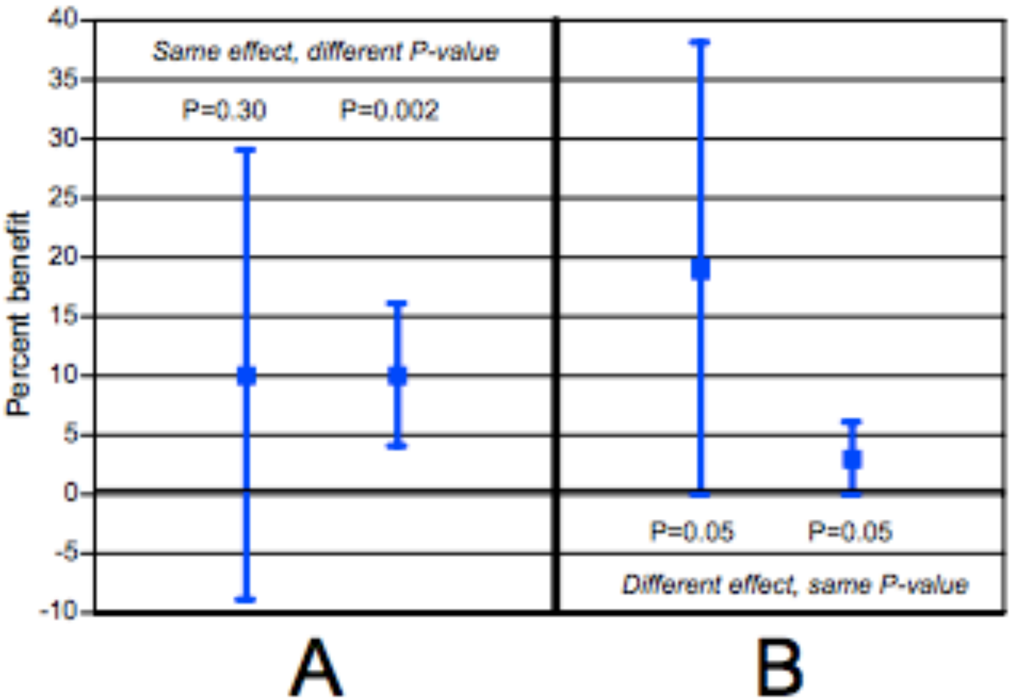
Study I: P=0.3

Study II: P=0.002



Misconception #5

Studies with the same P value provide the same evidence against the null hypothesis



Misconception #6

P = .05 means that we have observed data that would occur only 5% of the time under the null hypothesis

The probability of the observed data, plus more extreme data, under the null hypothesis.

REGRESSION

INTRODUCTION TO DATA SCIENCE

teaching datascience .org

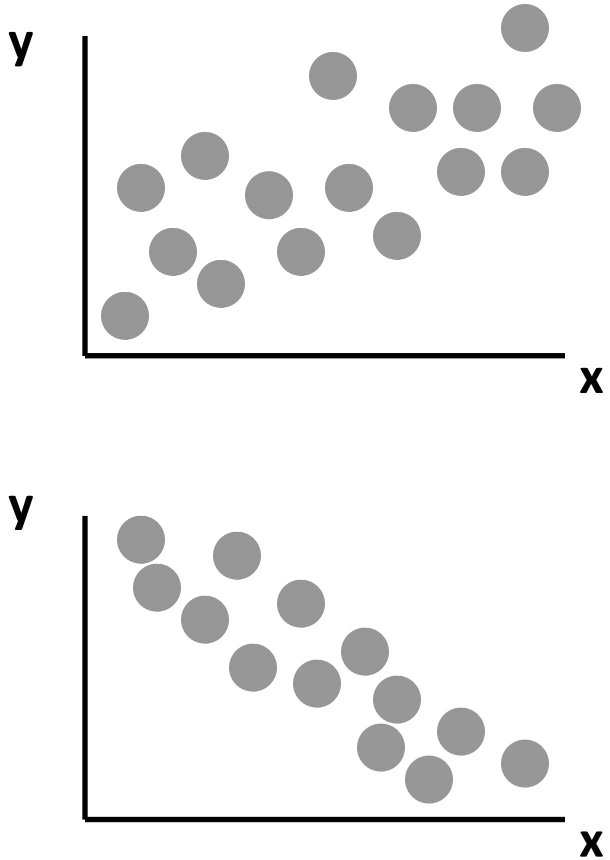
MATRIX

	Supervised Learning	Unsupervised Learning
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

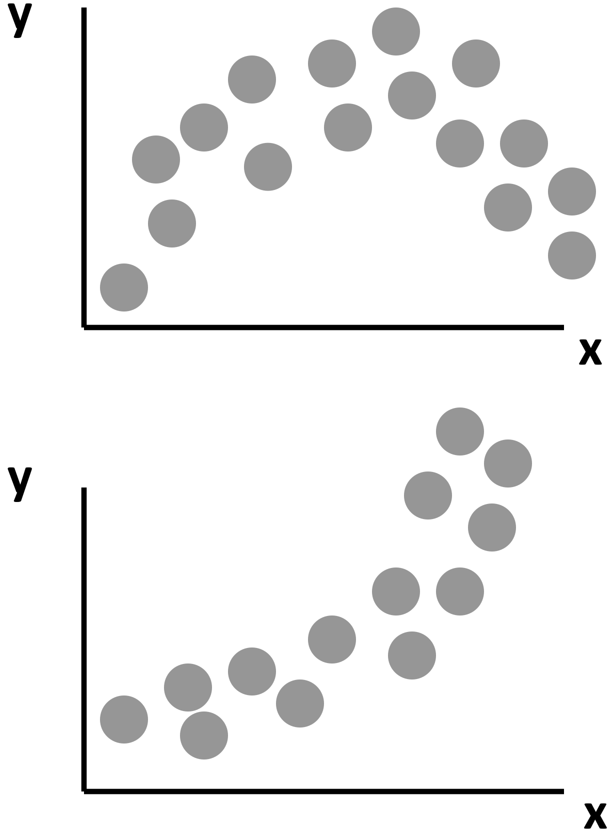
LINEAR REGRESSION

SCATTER PLOT EXAMPLES

Linear relationships



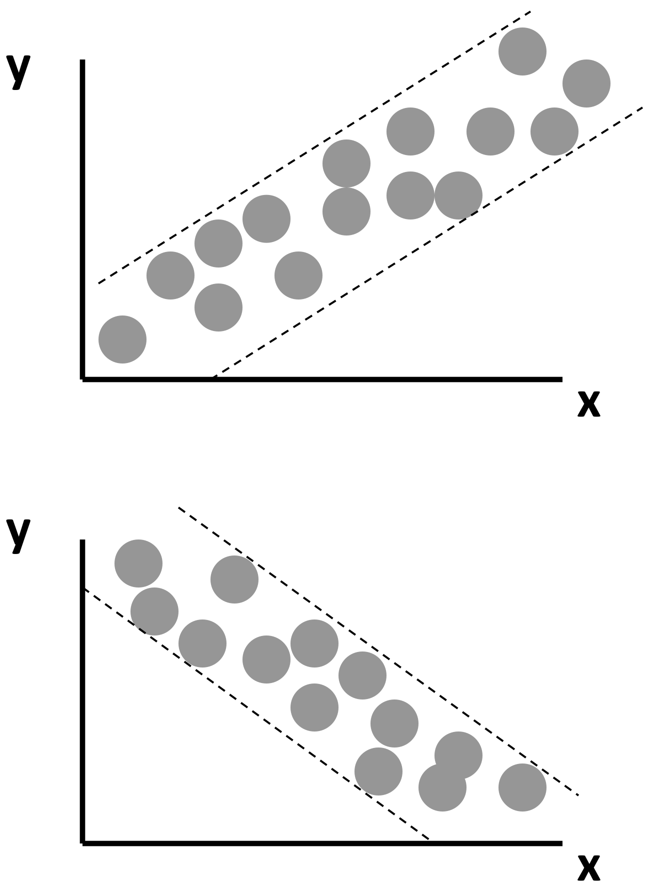
Curvilinear relationships



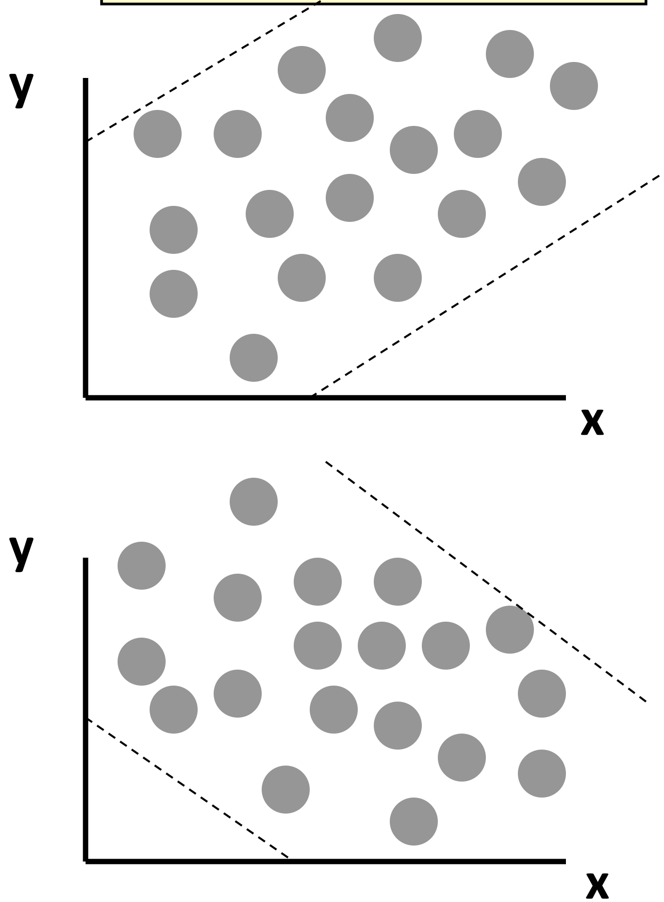
SCATTER PLOT EXAMPLES

(continued)

Strong relationships



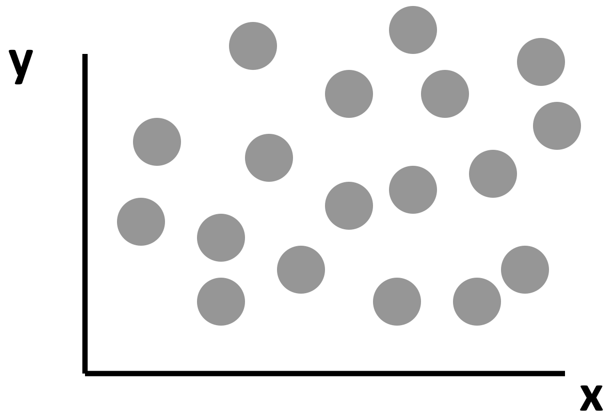
Weak relationships

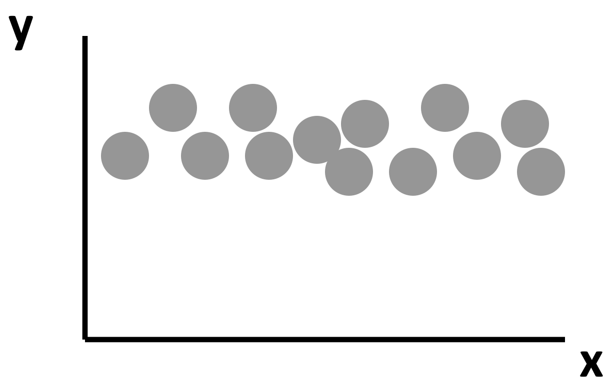


SCATTER PLOT EXAMPLES

(continued)

No relationship





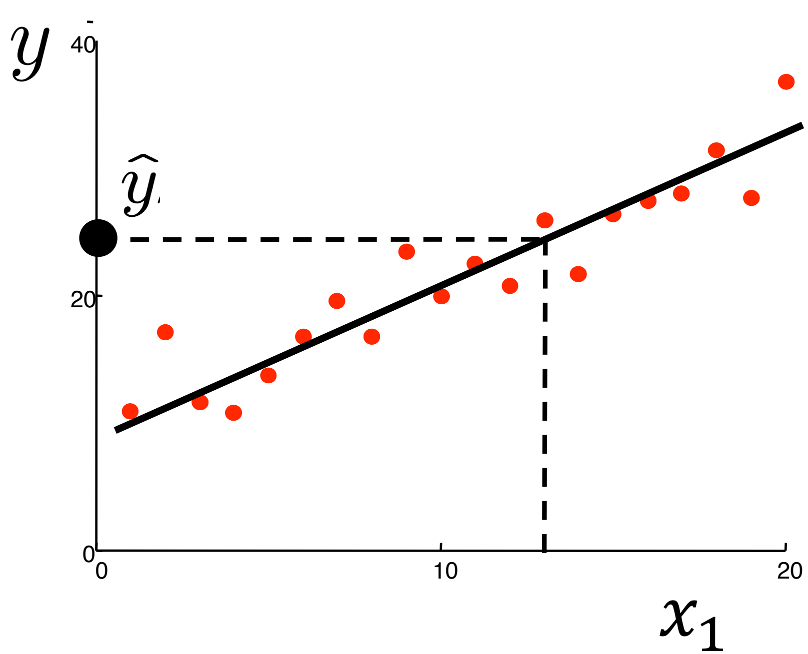
LINEAR REGRESSION

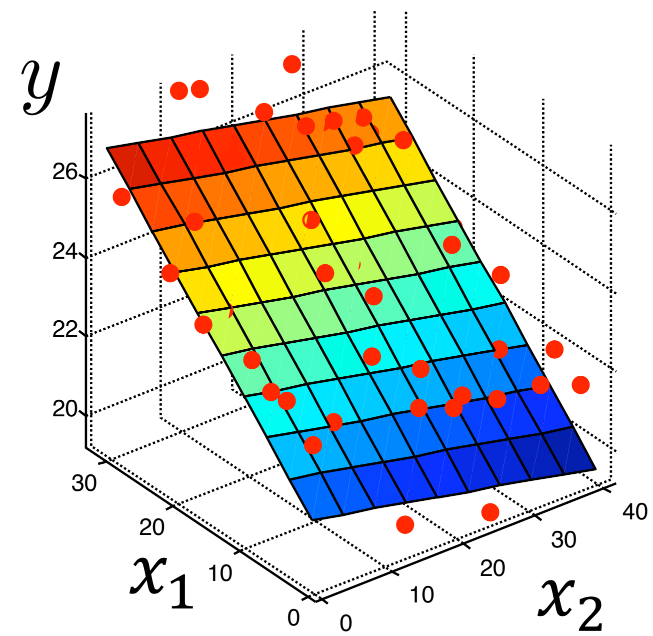
We wish to estimate $\hat{y}$ by a linear function of form:

$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2$$

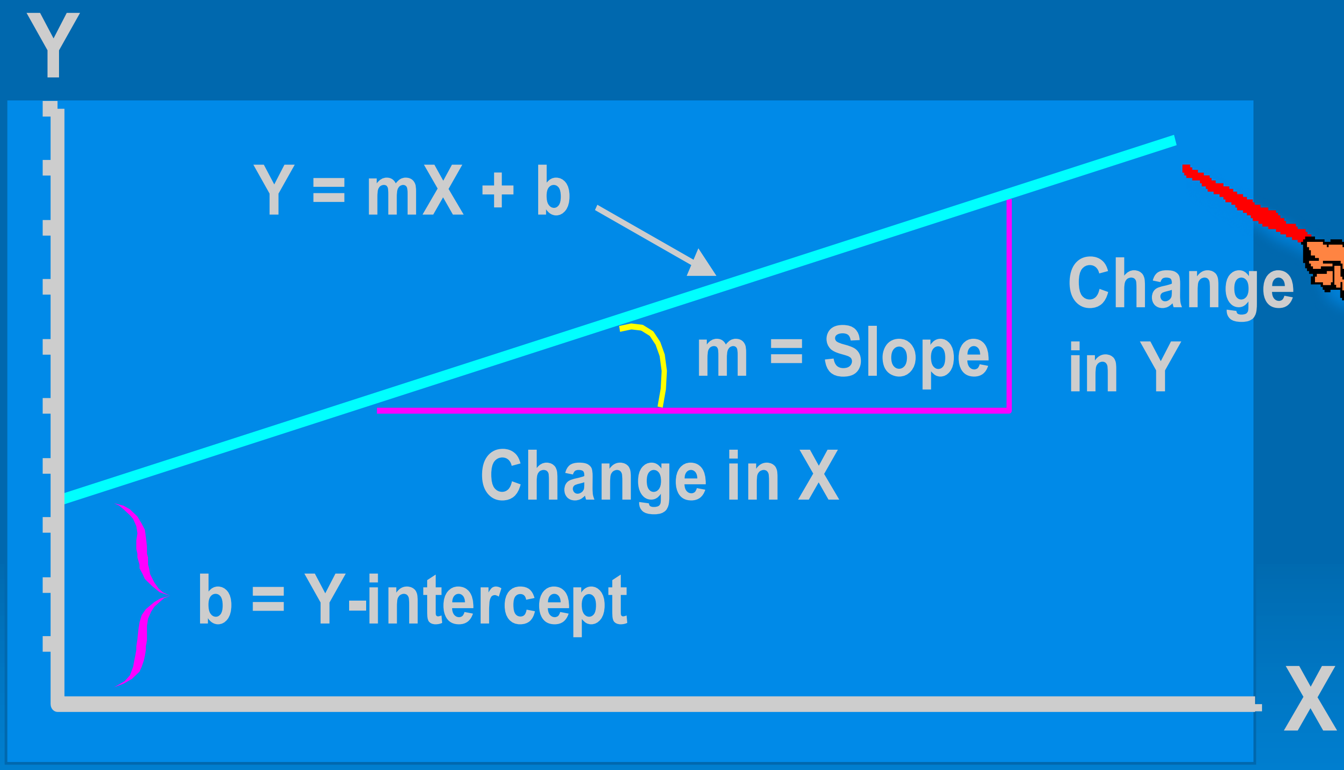
$$= w^T x \qquad w = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} \qquad \text{x}= \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix}$$

Standard convention of letting the first component of x be 1 (intercept term)





Linear Equations



Linear Regression Model

- 1. Relationship Between Variables Is a Linear Function

Population Y-Intercept

Population Slope

Random Error

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

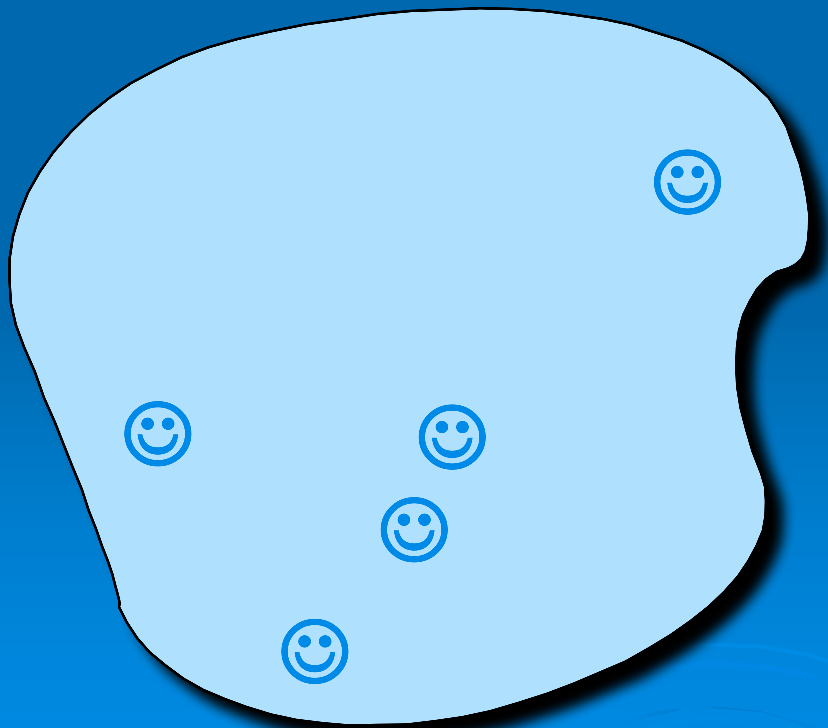
Dependent (Response) Variable (e.g., CD+ c.)

Independent (Explanatory) Variable (e.g., Years s. serocon.)

Population & Sample Regression Models

Population & Sample Regression Models

Population



Population & Sample Regression Models

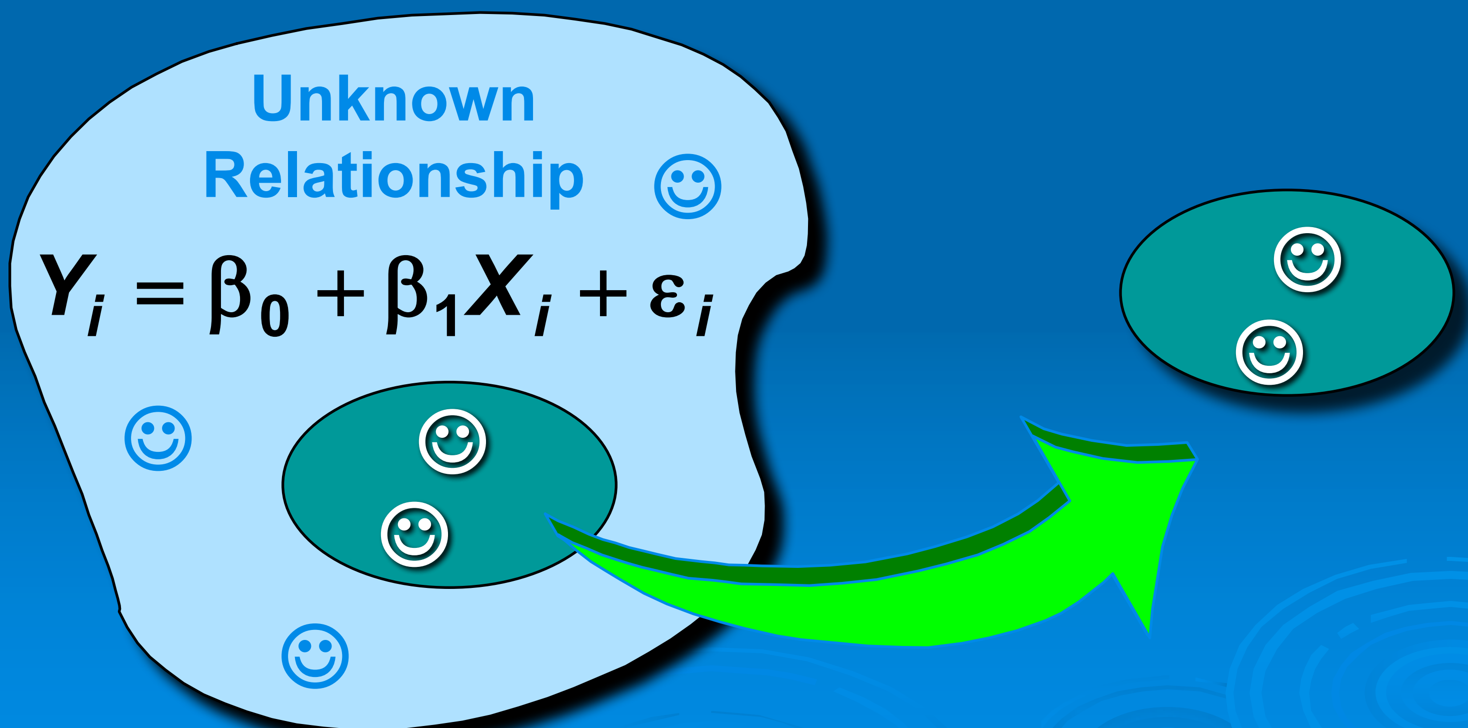
Population

Unknown Relationship

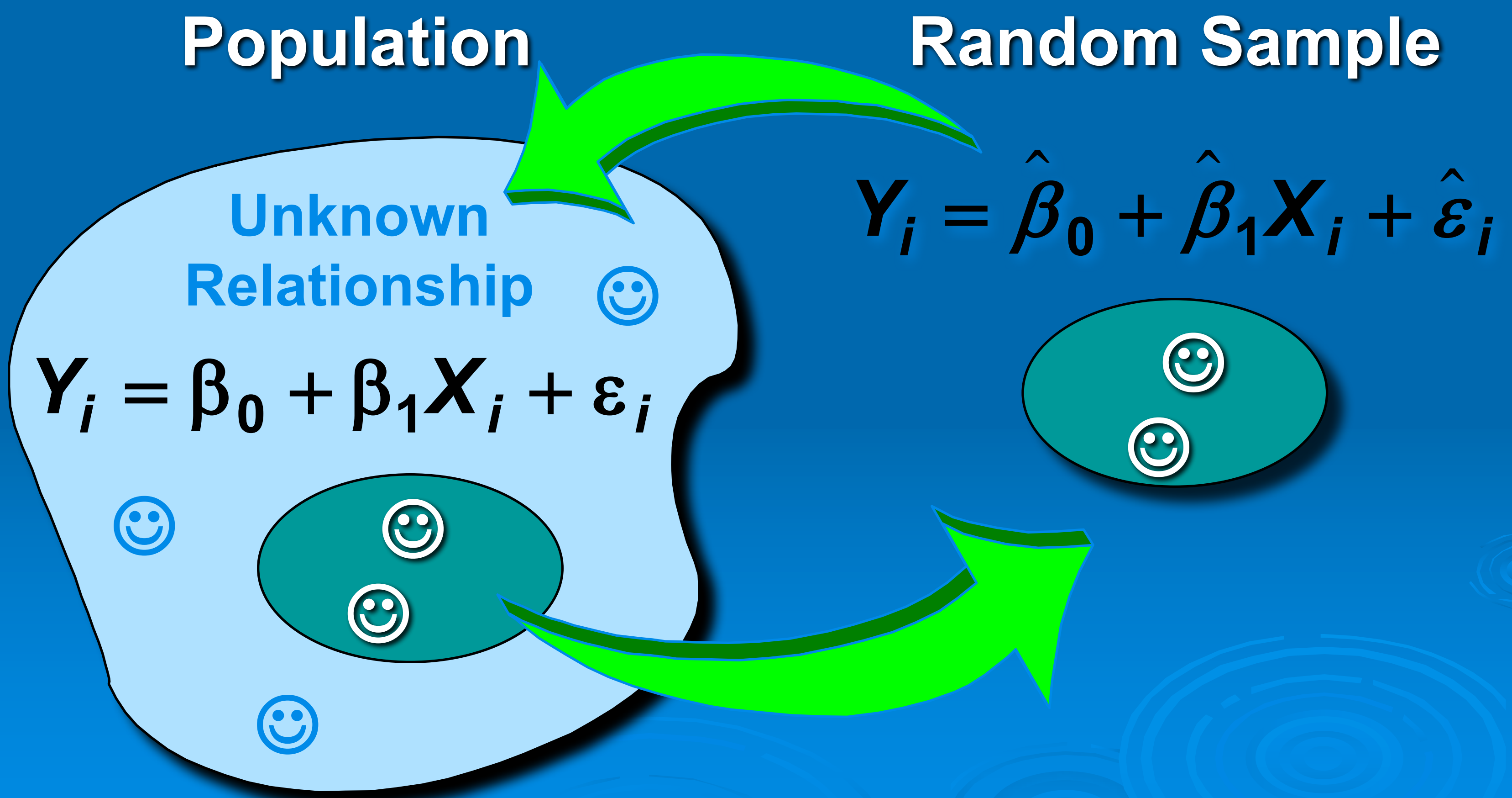
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Population & Sample Regression Models

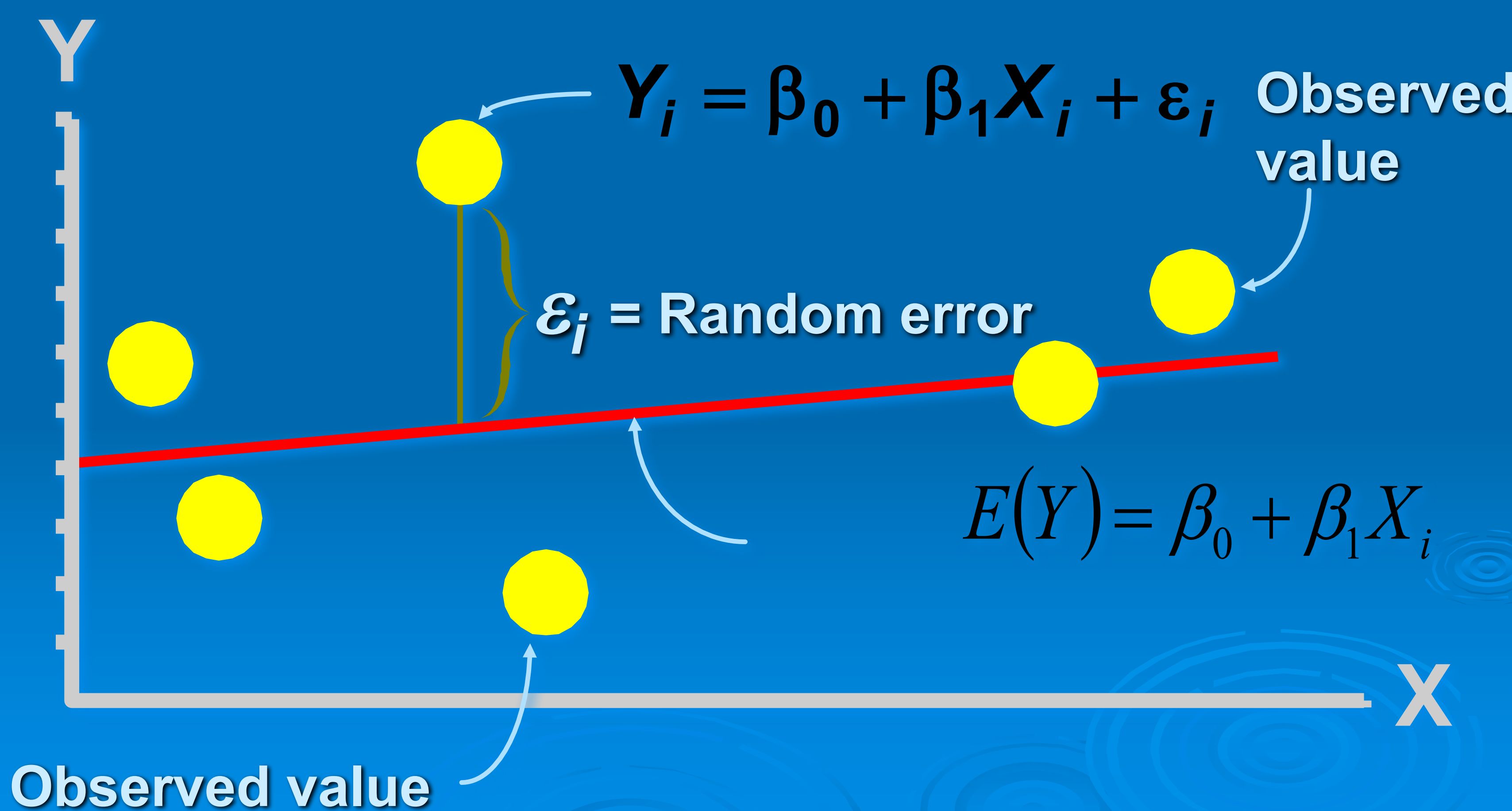
Population **Random Sample**



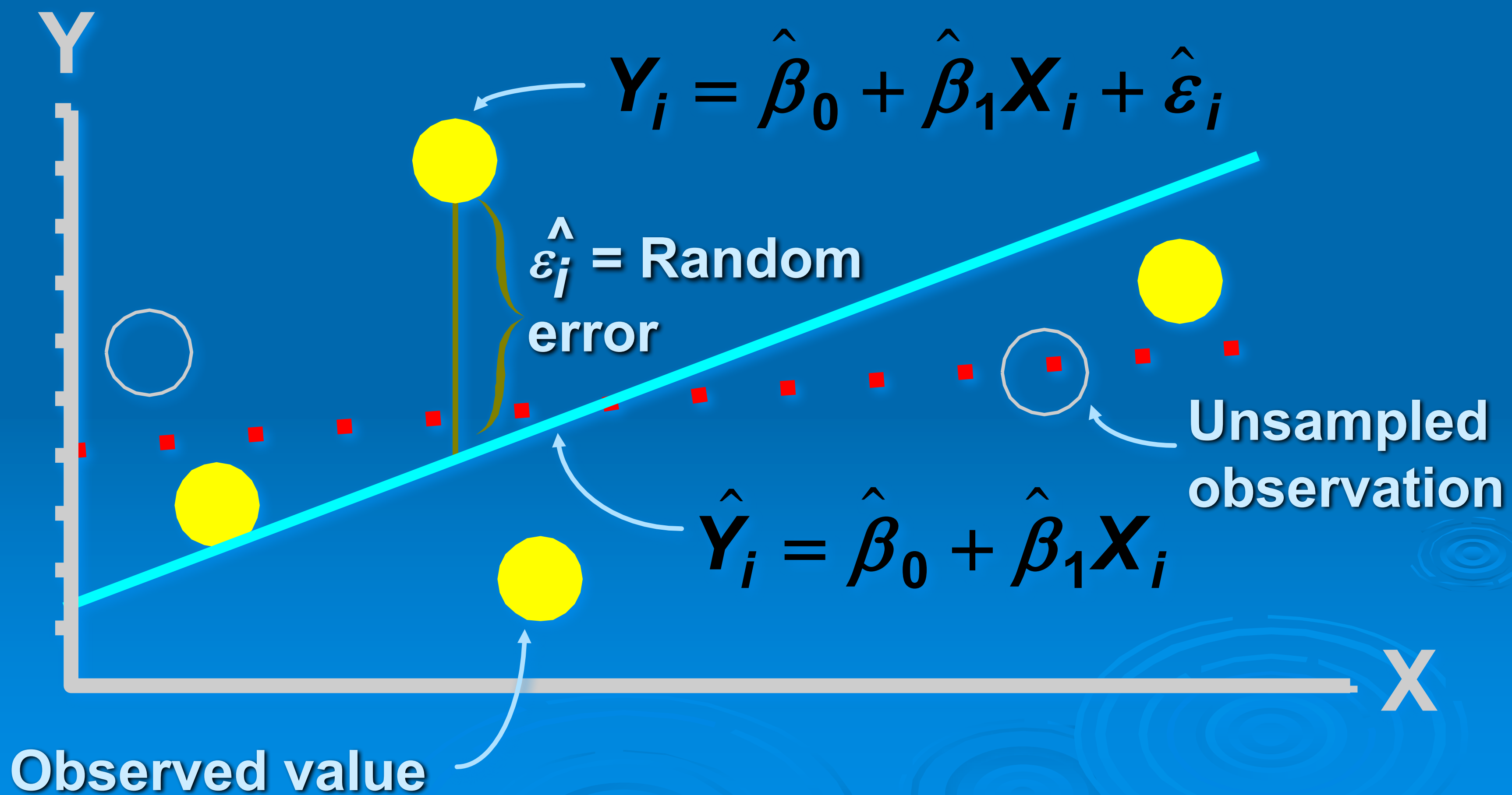
Population & Sample Regression Models



Population Linear Regression Model



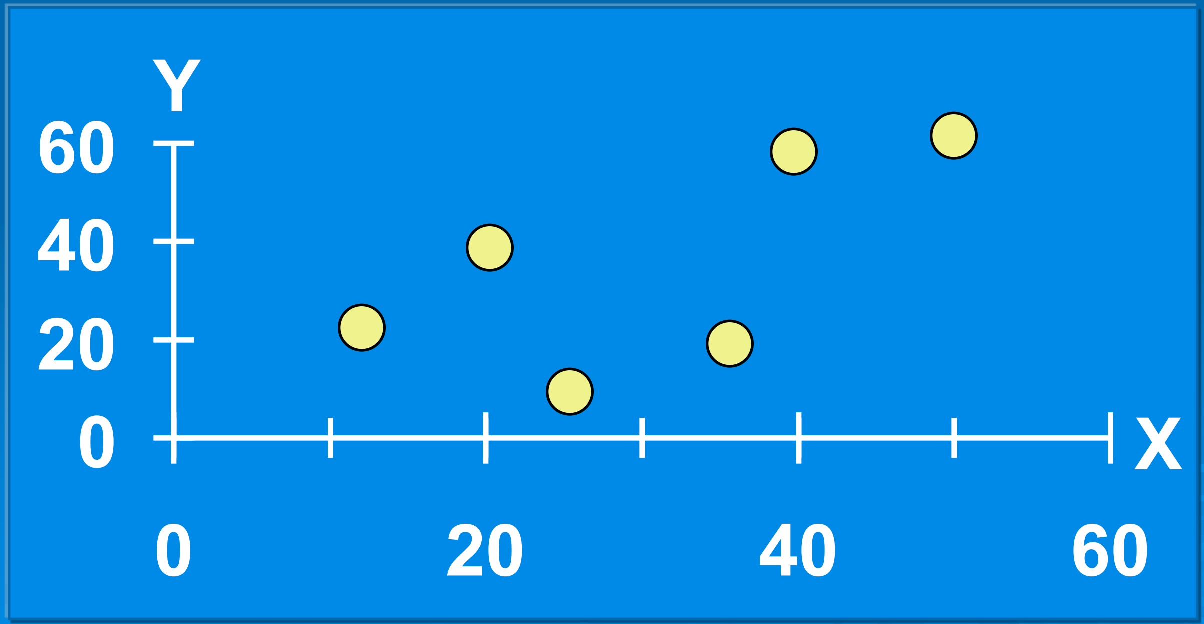
Sample Linear Regression Model



Estimating Parameters: Least Squares Method

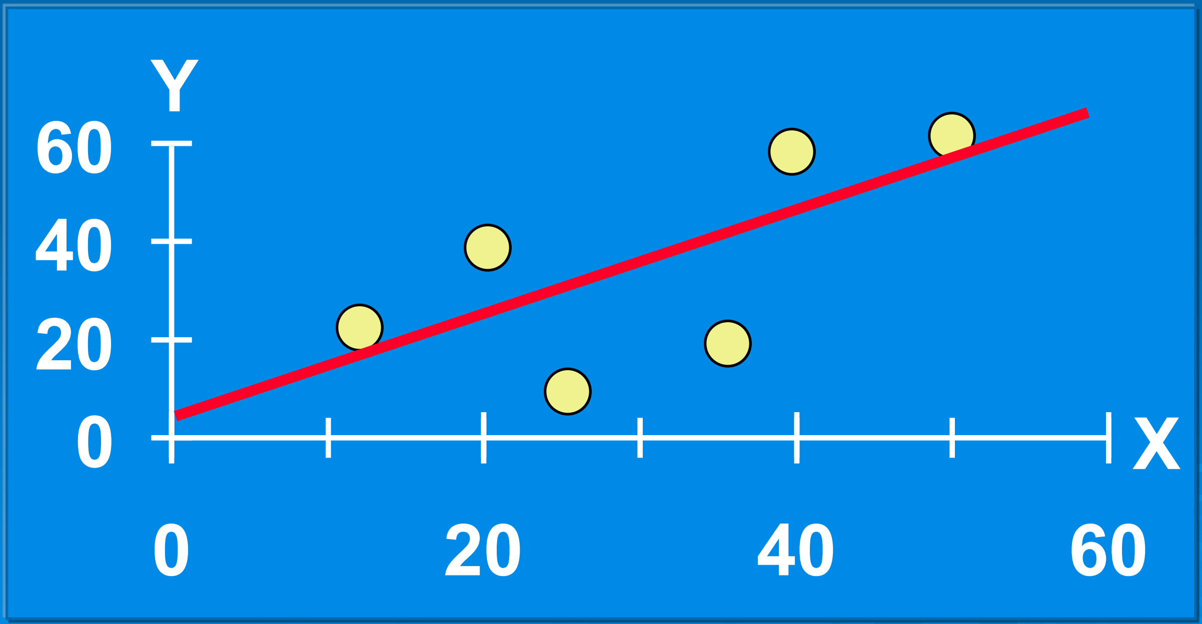
Scatter plot

- 1. Plot of All (X_i, Y_i) Pairs
- 2. Suggests How Well Model Will Fit



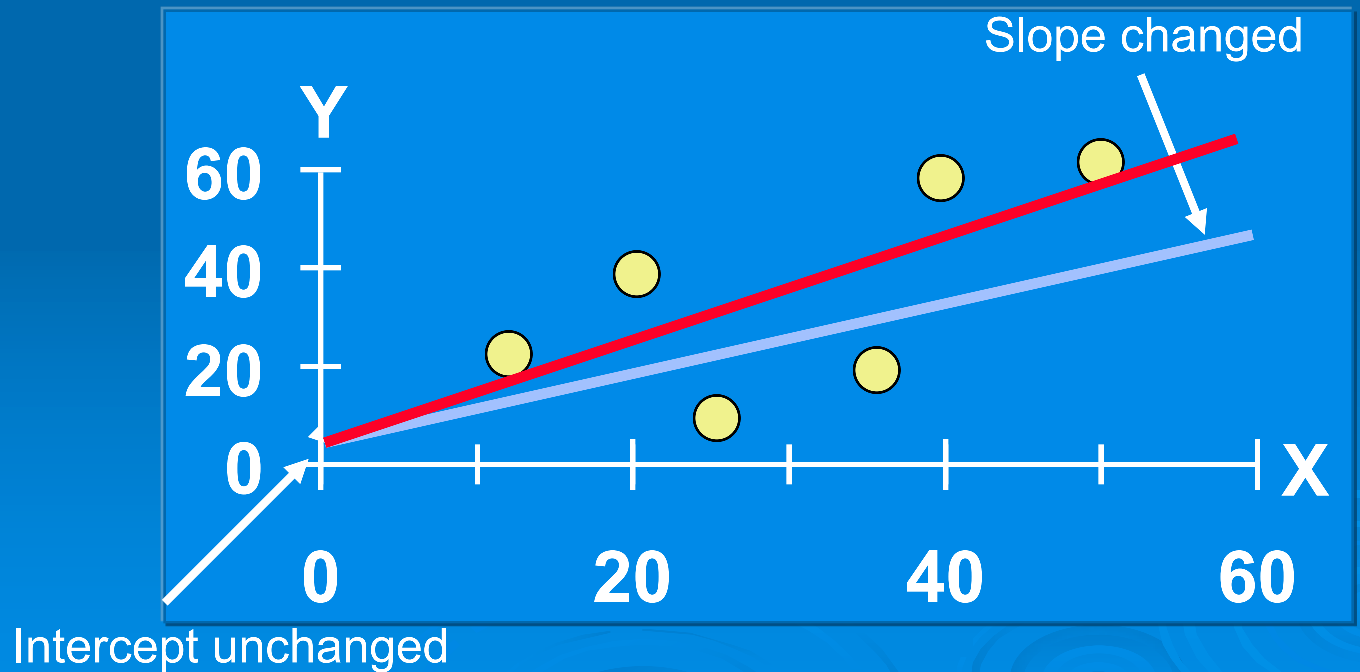
Thinking Challenge

How would you draw a line through the points? How do you determine which line 'fits best'?



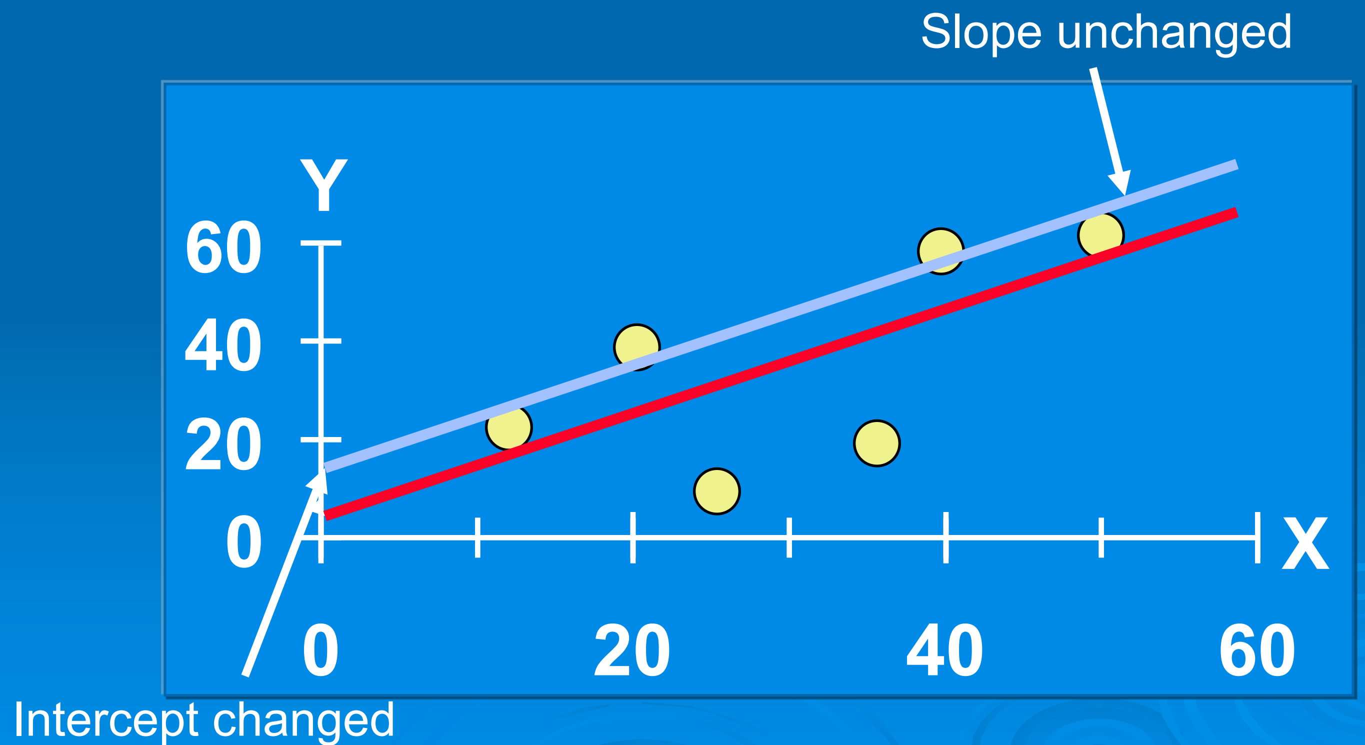
Thinking Challenge

How would you draw a line through the points? How do you determine which line 'fits best'?



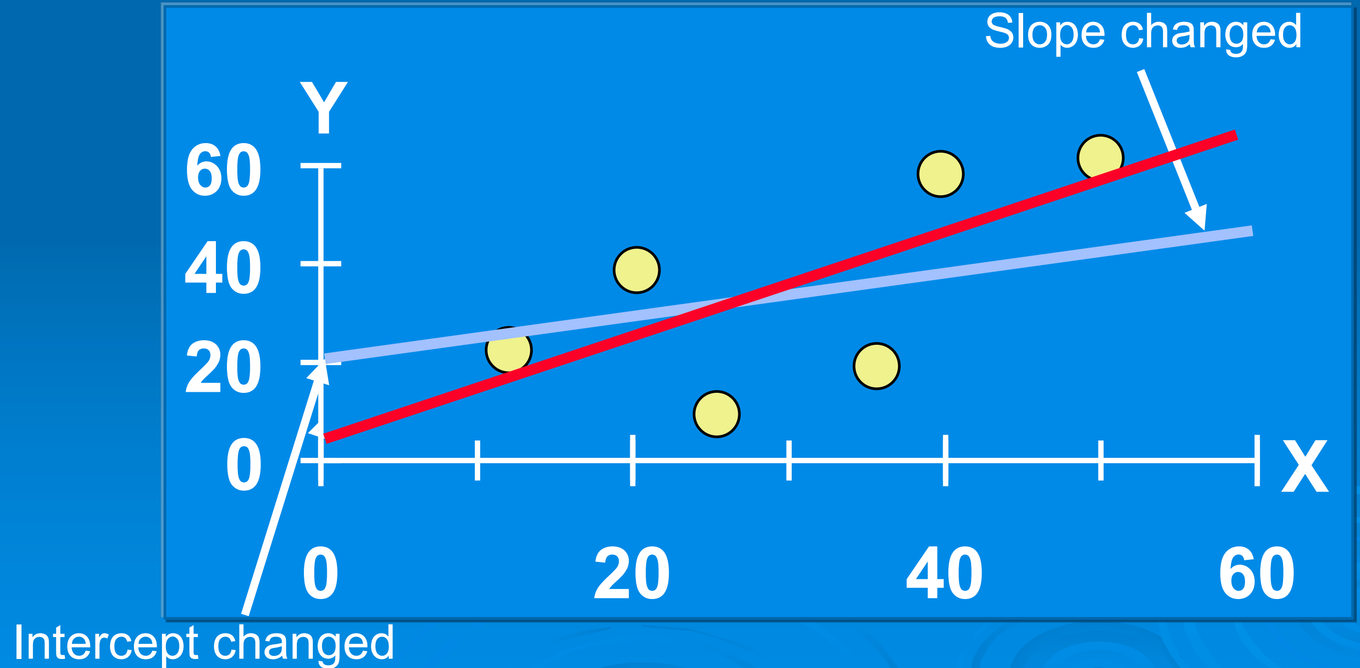
Thinking Challenge

How would you draw a line through the points? How do you determine which line 'fits best'?



Thinking Challenge

How would you draw a line through the points? How do you determine which line 'fits best'?



Least Squares

- 1. 'Best Fit' Means Difference Between Actual Y Values & Predicted Y Values Are a Minimum. *But* Positive Differences Off-Set Negative ones

Least Squares

- 1. 'Best Fit' Means Difference Between Actual Y Values & Predicted Y Values is a Minimum. *But* Positive Differences Off-Set Negative ones. So square errors!

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

Least Squares

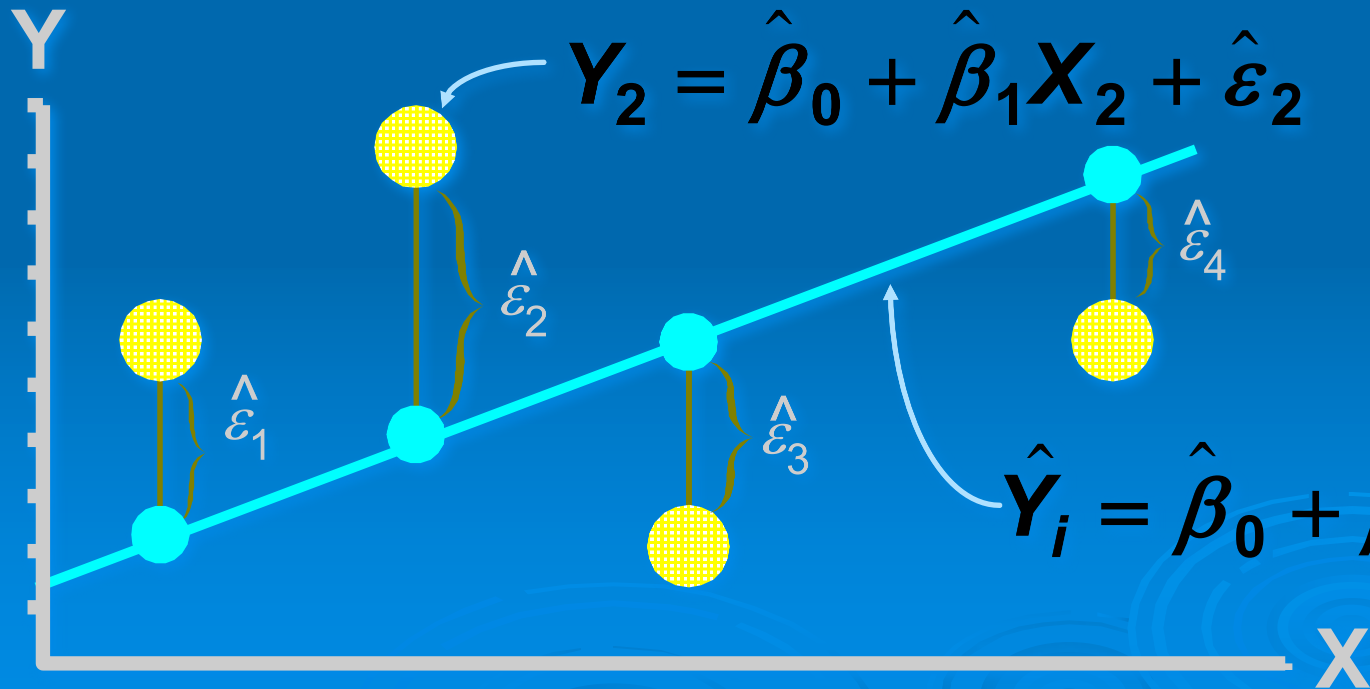
- 1. 'Best Fit' Means Difference Between Actual Y Values & Predicted Y Values Are a Minimum. *But* Positive Differences Off-Set Negative. So square errors!

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

- 2. LS Minimizes the Sum of the Squared Differences (errors) (SSE)

Least Squares Graphically

LS minimizes $\sum_{i=1}^{n}\hat{\varepsilon}_i^2 = \hat{\varepsilon}_1^2 + \hat{\varepsilon}_2^2 + \hat{\varepsilon}_3^2 + \hat{\varepsilon}_4^2$



$$Y_2 = \hat{\beta}_0 + \hat{\beta}_1 X_2 + \hat{\varepsilon}_2$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Coefficient Equations

- Prediction equation

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- Sample slope

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

- Sample Y - intercept

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Derivation of Parameters (1)

- Least Squares (L-S):
  Minimize squared error

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i \right)^2$$

$$0 = \frac{\partial \sum \varepsilon_i^2}{\partial \beta_0} = \frac{\partial \sum \left(y_i - \beta_0 - \beta_1 x_i \right)^2}{\partial \beta_0}$$

$$= -2\left(n\bar{y} - n\beta_0 - n\beta_1 \bar{x} \right)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Derivation of Parameters (1)

- Least Squares (L-S):

  Minimize squared error

$$0 = \frac{\partial \sum \varepsilon_i^2}{\partial \beta_1} = \frac{\partial \sum \left(y_i - \beta_0 - \beta_1 x_i\right)^2}{\partial \beta_1}$$

$$= -2 \sum x_i \left(y_i - \beta_0 - \beta_1 x_i\right)$$

$$= -2 \sum x_i \left(y_i - \overline{y} + \beta_1 \overline{x} - \beta_1 x_i\right)$$

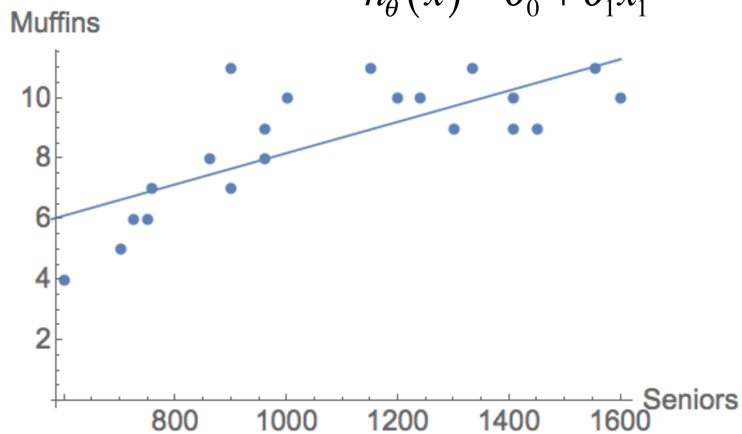
$$\beta_1 \sum x_i \left(x_i - \overline{x}\right) = \sum x_i \left(y_i - \overline{y}\right)$$

$$\beta_1 \sum \left(x_i - \overline{x}\right)\left(x_i - \overline{x}\right) = \sum \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)$$

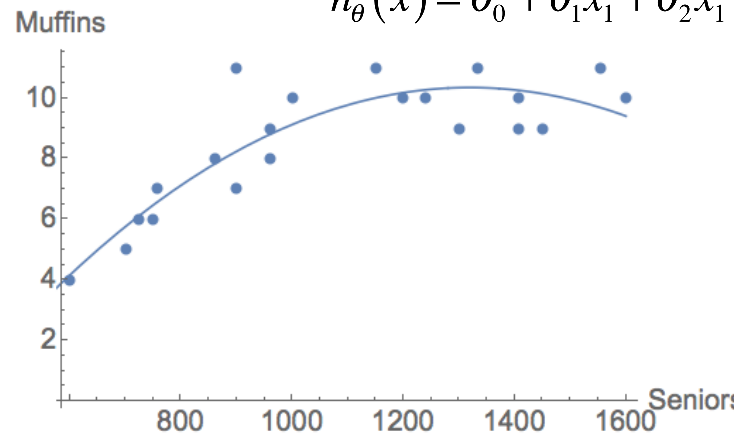
$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

Polynomial Regression

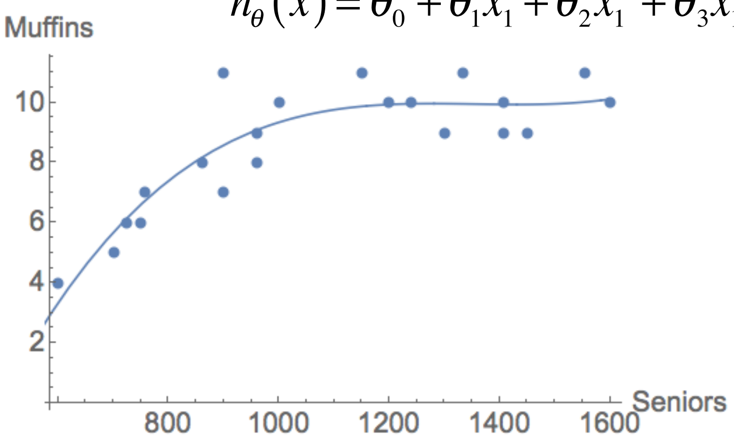
$$h_\theta(x) = \theta_0 + \theta_1 x_1$$



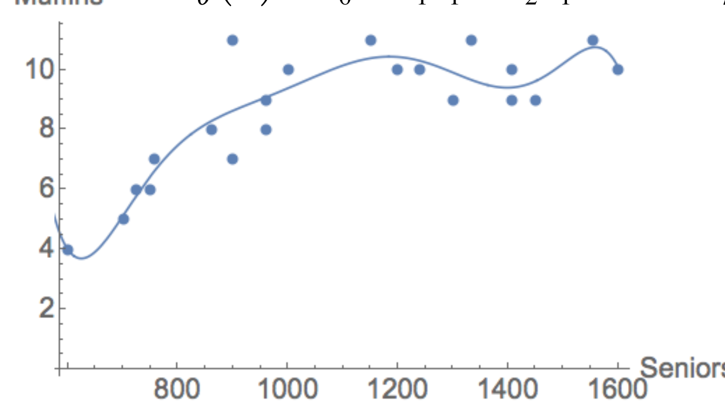
$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2$$



$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^3$$



$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \ldots + \theta_7 x_1^7$$

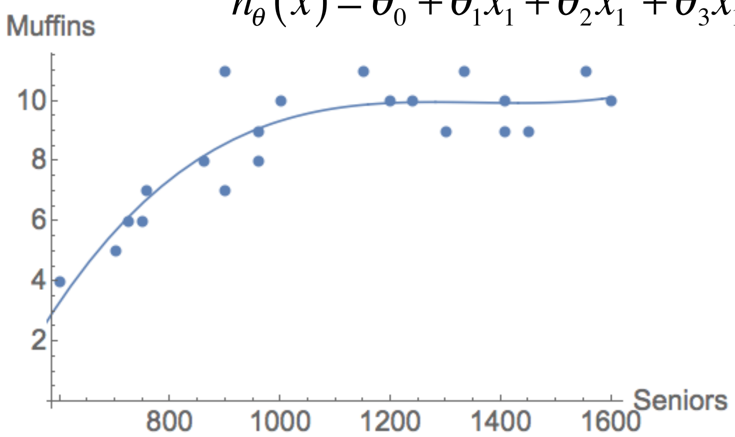


How To Prevent Overfitting

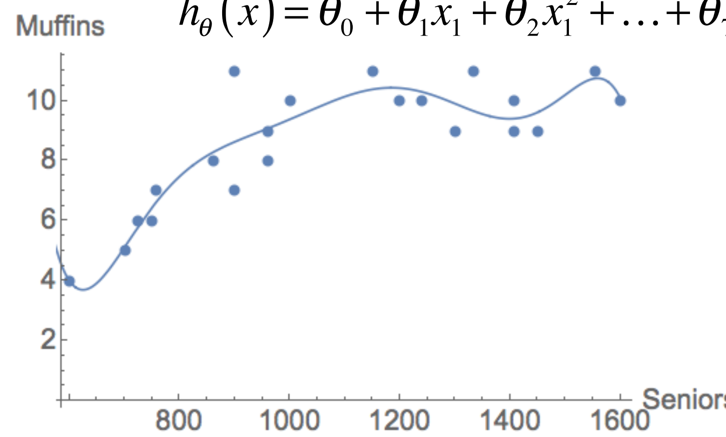
- Adjust features
  - Reduce number of polynomials
  - Reduce number of features
- Regularization
  - Keep all the features, but reduce their impact
  - Works well when we have a lot of features, each of which contributes a little

Regularization: Intuition

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^3$



$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \ldots + \theta_7 x_1^7$



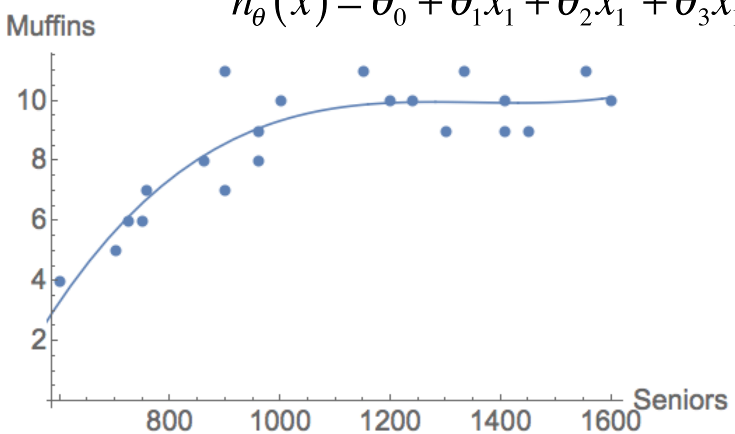
$$\min_\theta J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$+1000\theta_4 + 1000\theta_5 + 1000\theta_6 + 1000\theta_7$$

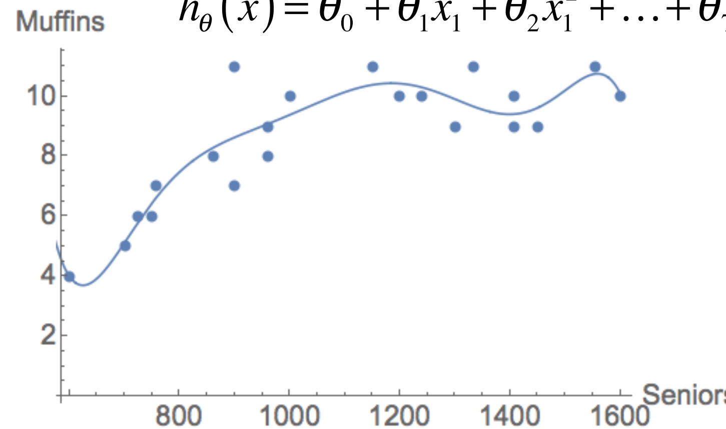
$$\theta_4, \theta_5, \theta_6, \theta_7 \to 0$$

Regularization: Intuition

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^3$$



$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \ldots + \theta_7 x_1^7$$



$$\min_\theta J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 + \underbrace{\lambda \sum_{j=1}^{n} \theta_j^2}_{\text{Regularization term}}$$

Works ok for reducing the impact of polynomials (better to reduce nb of polynomials directly)
Works great for a bag of equally important features.