Section 3E

How Numbers Deceive:
Polygraphs, Mammograms, and
More

Better in Each Case, but Worse Overall

It is possible for a set of data to give different results in each of several groups than it does when the groups are taken together.

If you look carefully, it is because the overall results were divided into unequally sized groups.

This is an example of **Simpson's paradox** which arises surprisingly often and is one way in which numbers can deceive unless they are examined with great care.

Example: The following table shows the number of hits (H), number of at-bats (AB), and batting average (AVG = H/AB) for major leaguers Derek Jeter and David Justice in 1995 and 1996.

	1995	1996
Jeter	12 H, 48 AB, AVG = .250	183 H, 582 AB, AVG = .310
Justice	104 H, 411 AB, AVG = .253	45 H, 140 AB, AVG = .321

Which player had the higher batting average in both 1995 and 1996?

Justice had higher batting averages in both 1995 & 1996

Compute the batting average for each player for the two years combined. Which is highest?

Jeter = (12 + 183)/(48 + 582) = 195/630 = .310

Justice = (104 + 45)/(411 + 140) = 149/551 = .270

Jeter had the higher combined batting average.

Does a Positive Mammogram Mean Cancer?

True positives are positive mammograms where the tumor is malignant.

False negatives are negative mammograms where the tumor is malignant.

True negatives are negative mammograms where the tumor is benign.

False positives are positive mammograms where the tumor is benign.

Example: Using the table below, answer the questions.

	Malignant	Benign	Total
Positive	90	990	1080
Negative	10	8910	8920
Total	100	9900	10,000

Suppose a patient has a positive mammogram. What is the chance that she really has cancer?

90/1080 = 0.083 = 8.3%

What is the chance of a positive mammogram, given that the patient has cancer?

90/100 = .90 = 90%

Suppose a patient has a negative mammogram. What is the chance that she actually does have cancer?

10/8920 = 0.001 = 0.1%

Example: Suppose that a polygraph is 85% accurate. The 2000 employees of a company are given a polygraph test during which they are asked whether they use drugs. All of them deny drug use, when, in fact, 1% of the employees actually use drugs. Assume that anyone whom the polygraph operator finds untruthful is accused of lying.

	Users	Non-users	Total
Test Finds Employee Lying	17	297	314
Test Finds Employee Truthful	3	1683	1686
Total	20	1980	2000

How many employees were accused of lying? Of these, how many were actually lying and how many were telling the truth? What percentage of those accused of lying were falsely accused?

314 were accused of lying. 17 of those were actually lying and 297 were not. 297/314 = .946 = 94.6% were falsely accused of lying.

How many employees are found truthful? Of these, how many were actually truthful? What percentage of those found truthful really were truthful?

1686 were found truthful. 1683 were actually truthful and 3 were not. 1683/1686 = .998 = 99.8% were found truthful and really were.