# Mathematics 1000, Winter 2008

## Lecture 8

Sheng Zhang

February 4, 2007

# Today's Topics

# Properties of $r$

- Always between $-1$ and 1

- $r > 0 \longrightarrow$ positive relation
- $r < 0 \longrightarrow$ negative relation

- Close to 1 or close to $-1 \longrightarrow$ strong correlation
- Close to 0 $\longrightarrow$ weak correlation

# More properties of $r$

- Does not depend on units used
- Can be affected by outliers

# More properties of $r$

- Does not depend on units used
- Can be affected by outliers

# Caution

Think about whether the value of $r$ calculated by your calculator makes sense in view of the picture as well as the context.

# Warning

The book writes the equation of the regression line as

$$y = a + bx$$

Many calculators use $y = ax + b$

These are inconsistent! The meanings of $a$ and $b$ are reversed.

If you are used to $y = mx + b$ then your $b$ is the same as the calculator's. It is not the book's $b$.

# The Least-Squares Regression Line

This is the line that ‘best’ fits the data in a specific, technical sense.

The slope of this line is positive when the the correlation is positive.

But the size of the slope depends on the scale used for the parameters.

The size of the correlation $r$ tells you how well the straight regression line fits the data.

# Example: Beer and Blood Alcohol Concentration

$r = .89$, $m = .018$

beers.pdf

# Relation between slope and correlation

$$m = r\frac{s_y}{s_x}$$

$m =$ slope
$r =$ correlation
$s_x =$ standard deviation for $x$
$s_y =$ standard deviation for $y$

If $s_y$ is much larger than $s_x$, then $m$ can be big even if $r$ is small.

If $s_y$ is much smaller than $s_x$, then $m$ can be small even if $r$ is large.

# Remember

There will be such a least squares regression line, whether or not it has any significance!

When it is meaningful, we can use the regression line to make predictions.

Interpretation is critical.

# Example

speedfuel2.pdf

# Interpreting Correlation and Regression

The regression line can be strongly influenced by a small number of outliers, sometimes even by just one outlier.

To minimize the sum of squares, when there is a clump of points and one outlier, the trend line will join that outlier to the middle of the clump.

# Outliers Example 1

exp1.pdf

# Outliers Example 1, with regression line

exp1L.pdf

# Outliers Example 2

exp2.pdf

# Outliers Example 2, with regression line

exp2L.pdf

# Outliers Example 3

exp3.pdf

# Outliers Example 3, with regression line

exp3L.pdf

# Marriage-Divorce Rates, All States

divorceall.pdf

# Marriage-Divorce Rates, w/o Hawaii and Nevada

divorcenovegas.pdf

# Causal relations

We have a scatterplot and a correlation. What can we conclude from this?

If the correlation is close to 0 or is strongly influenced by outliers, then the correct answer is “not much.”

We’ll focus now on situations where the correlation is high and is not influenced by outliers.

# Be careful!

This stage can be especially tricky and contentious.

People seem to have a tendency to see a cause when there may be only a high correlation.

# Example: Pocket change

There is a statistical correlation between an individual's height and the amount of money in that person's pocket.

Why do taller people tend to carry more money in their pockets?

Are they less afraid of being robbed?

Or is there some other factor at work?

# Example: Pocket change

There is a statistical correlation between an individual's height and the amount of money in that person's pocket.

Why do taller people tend to carry more money in their pockets?

Are they less afraid of being robbed?

Or is there some other factor at work?

# Example: Pocket change

There is a statistical correlation between an individual's height and the amount of money in that person's pocket.

Why do taller people tend to carry more money in their pockets?

Are they less afraid of being robbed?

Or is there some other factor at work?

# Example: Pocket change

There is a statistical correlation between an individual's height and the amount of money in that person's pocket.

Why do taller people tend to carry more money in their pockets?

Are they less afraid of being robbed?

Or is there some other factor at work?

# The underlying cause

There are two types of people.

One type tends to be taller and carry money in their pockets.

The other type tends to be shorter and carry money in their purses.

# The underlying cause

There are two types of people.

One type tends to be taller and carry money in their pockets.

The other type tends to be shorter and carry money in their purses.

# The underlying cause

There are two types of people.

One type tends to be taller and carry money in their pockets.

The other type tends to be shorter and carry money in their purses.

# Some slightly more subtle examples

From the book:

- Children who listen to Mozart tend to score higher in tests of verbal skills.
- Countries where people spend more time watching television have higher life expectancies.
- People who drink large amounts of diet soda are heavier for their height than people who do not.

# Example: a cancer scare

A 1979 study found a strong correlation between exposure to strong electromagnetic fields and incidence of childhood leukemia.

This was initially interpreted as evidence that these fields posed a health hazard.

More extensive investigations have indicated that “although some of the correlations remain we [researchers at the University of Texas] now do not think that this type of radiation causes cancer.”

`www.utexas.edu/courses/bio301d/Topics/EMF/Text.html`

# A tough call

Determining the direction of the causal relationship is sometimes tricky.

Health and wealth are correlated, but which causes which?

# Global warming

Is there a link between human activities, such as burning fossil fuels (e.g., gasoline), and increasing global temperatures?

This has been a highly contentious question.

Last week, the IPCC, the Intergovernmental Panel on Climate Change, issued a report that concluded there is almost certainly such a causal link.

ipcc22.pdf

For most people, this should be a definitive answer to the question.

It is not difficult to find people with apparently good credentials who dispute that conclusion.

marshall.pdf

That the oil industry lobbying group is offering generous payments to people who will make such statements is of course rather suspicious.

bribe2.pdf

Unfortunately, sorting through sophisticated, but possibly dishonest, claims is beyond the current scope of this course.

The best we can hope for this semester is to be able to spot some of the cruder fakes.

# Looking ahead

Chapter 7 concerns how to gather data.

If you want to know something about all of the states, you can just list them all.

But if you want to know something about all of the students at Wayne State University, it is probably impractical to ask every one of them.

# Sampling techniques

We'll be looking at sampling techniques, ways to look at a relatively small number of cases in order to get information about the whole.

We'll see where the 'margin of error' comes from in survey results and polls.