Put Options (NYT September 26, 2001)

Put options on stocks give buyers the right to sell stock at a specified price during a certain time. They rise in value if the underlying stock falls below the strike price.

The value of puts on airlines stocks soared Sept. 17 when U.S. stock and options markets reopened after a four-day closure as airline stocks slid as much as 40 percent.

American Airlines at $32 prior to attack. Suppose a terrorist buys a put option (at say $5 per share) to have the right to sell at $25. The price after the attack was at $16. Put option is now more valuable.

Outliers

Outliers are data that are not compatible with the bulk of the data.

They show up in graphical displays as detached or stray points.

Sometimes they indicate errors in data input. Experts estimate that roughly 5% of all data entered is in error.

Sometimes they are the most important data points.
The point:

The outlier is a very informative point.

This is an example of a 'good' outlier.

It is the most interesting point in the data set.

However, it is interesting only because of the pattern of the rest of the data.

Next

An example in which the outlier is bad.

It will cause the correlation to be lower than it should be.

It will obscure the strength of the relationship.
The point:

The point is that one point can interfere with the proper interpretation of the data. Plots look completely different for example.

Removing only one point can dramatically change the value of the correlation coefficient.

These are examples of **bad outliers** and we generally remove them from the data set.

They probably were errors in data input.

In these examples the outlier resulted in too low correlation.

Next:

Another bad outlier.

This time it will create a false impression of strong relationship.

The correlation will appear to be larger than it probably really is.
The point:

In this case the outlier has increased, not decreased, the correlation.

The outlier has dominated the data and caused it to appear that there is some correlation when there really is none.

The Moral:

There can be good outliers: Election fraud. We use them to identify important parts of the data. Or in analyzing put options for extreme cases.

More often the outliers are bad. They can depress the correlation and make you think the relationship is weaker than it really is.

They can increase the correlation and make it appear that the relationship is stronger than it really is.

IMPORTANT: Always look at a scatter plot as well as compute the correlation.

Another problem:

Sometimes we see strong relationship in absurd examples.

Two seemingly unrelated variables have a high correlation.

This signals the presence of a third variable that is highly correlated with the other two.

(Confounding)

How can we have such high correlation between shoe size and vocabulary?

Easy: Both increase with age and hence age is a hidden variable.

Age is positively correlated with both shoe size and with vocabulary.