**Lecture #11**

**LINEAR REGRESSION**

Often in science or engineering – based on applicable domain knowledge – we expect variable X and Y to be linearly correlated. The following scatter diagram gives an intuitive understanding of such a relationship:

Y

X

Each point in the scatter diagram represents one pair of (X,Y) values. We are looking for an possible linear relationship between X and Y, that is:

Y’ = mX + C where Y' represents the orange line.

Question: Which straight line best represents the presumed linear relationship?

Answer: The one which minimizes the sum of squares of errors: Si [y’(xi) – yi]2

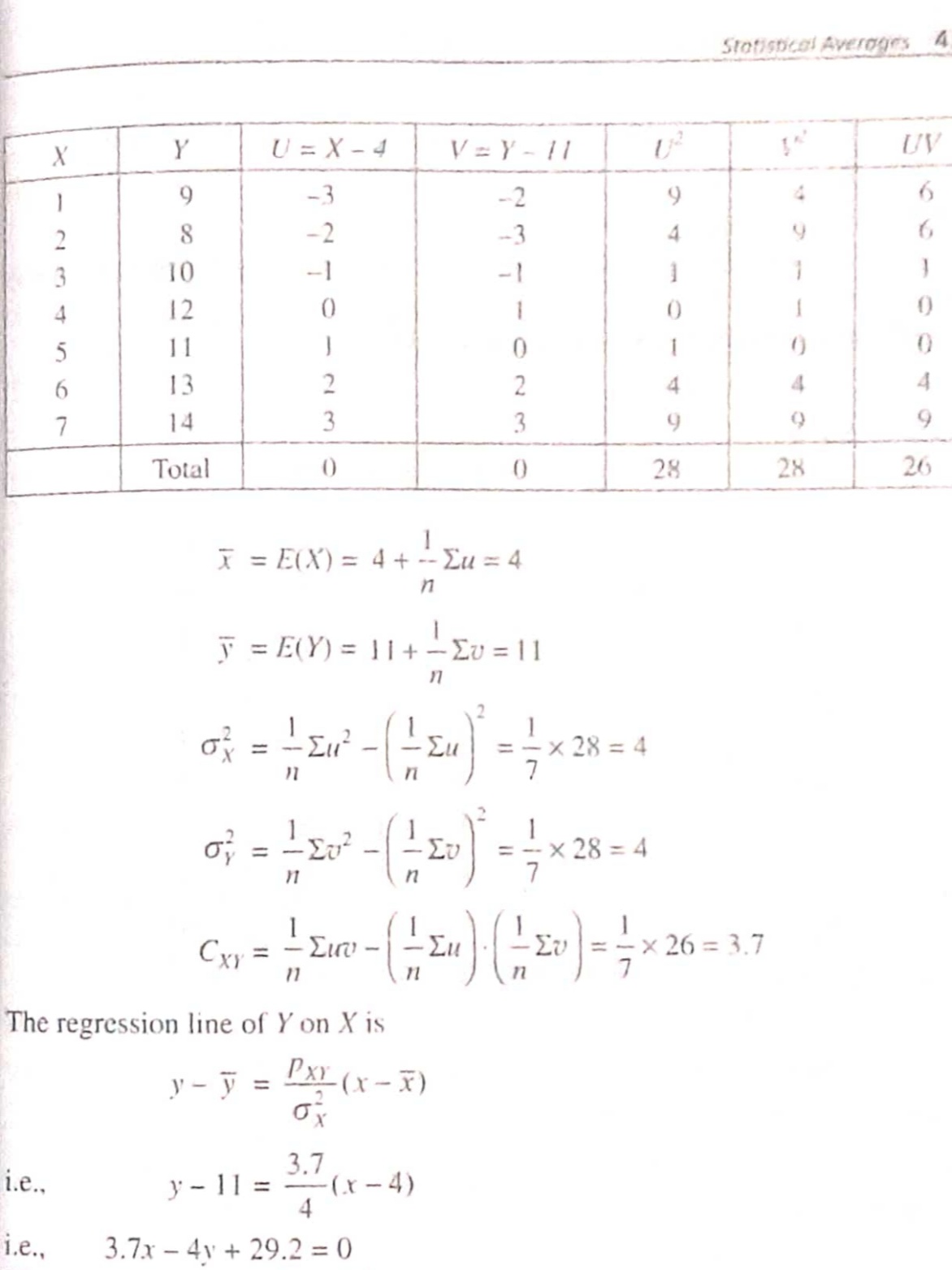
This is known as the least squares fit of the straight line to the data points. Using very basic calculus, the equation of the line can be derived as:

y' - mY = [Cov(X,Y)/sX2] \* [x – mX] this is the “linear regression of Y on X"

Note that the “direction of causality” cannot be obtained from this data. We could just as easily have found the regression of X on Y. Does X determine Y, or *vice versa*? Can we tell from the given data alone? No.

Solved example from Ref. #2 Obtain the regression line of Y on X from the X, Y pair-wise data listed in the first two columns.





**TRIALS, SAMPLING, LEARNING**



How can we determine that a newly acquired coin, or a pair of dice, or a carton of 100 pouches of milk powder, is “good", “fair" or “acceptable”?

Of course we must:

(a) clearly define the relevant criteria, and

(b) carry out repeated trials [in the first two cases], and “sample" a few pouches [in the third case]. The trials are also samples, in a certain sense.

Knowledge of (a) is part of the indispensable domain knowledge – or, in simple language, knowledge about the *real world*. The only way in which we can learn about the real world is through *sampling* of relevant data and *generalization*.

Whatever we infer from (b) is a form of learning, and it allows us to make certain decisions, future predictions or projections. But all this must happen even in the presence of a degree of uncertainty!



*Deterministic* versus *probabilistic* models of the real world. Of course, in this course, we are considering probabilistic models.

So far, we have been making the following type of Inferences:

F( x ; m ,s .... )

Predict x

*But how did we come to learn F( x ; m ,s .... ) in the first place?*

Estimate q

xi

F( x ; q .... )

where parameter q is unknown.

The process depicted in the second diagram is called *parameter estimation*, which is also essentially a form of learning.

So far, we have assumed that F(X), F(Y) F(X,Y) *et cetera* have already been learnt. In general, we follow the sequence of *learning* followed by *application*.

Some simple inferences can be made on the basis of only m and s, without even knowing anything else about F(X).

Tchebycheff inequality

Let X be any RV with mean m and variance s2. Then the following inequality holds:

Prob[ |X–m| > c ] < s2/c2, where c > 0.

Note that two factors are affecting – *in opposite sense!* – the probability of finding a “remote value": (i) variance s2, and (ii) distance c from mean.

[Think about what will be the units of c.]

This agrees with “common sense”, and the inequality gives a very useful bound.

Simple example: A food item is packaged with mean weight 500 gm, and standard deviation 2 gm. Find a bound on the probability that the weight in gm will be outside the interval [490..510].

The interval is given by |m - x| < 10, so c = 10 gm; s2 = 4 gm2. By Tchebycheff inequality, an upper bound on the required probability is 4/100 = 0.04.

Note that we obtained this upper bound without knowing the full probability distribution!

Solved example from Ref. #2

An RV X has mean m = 12 and variance s2 = 9, with an otherwise unknown probability distribution. Find a lower bound on Prob( 6 < X < 18 ).

Proceeding as in the previous example, with c = 6 [why?], an upper bound on X being outside the given interval is found to be 9/62 = ¼. Therefore a lower bound on the required probability is ¾.