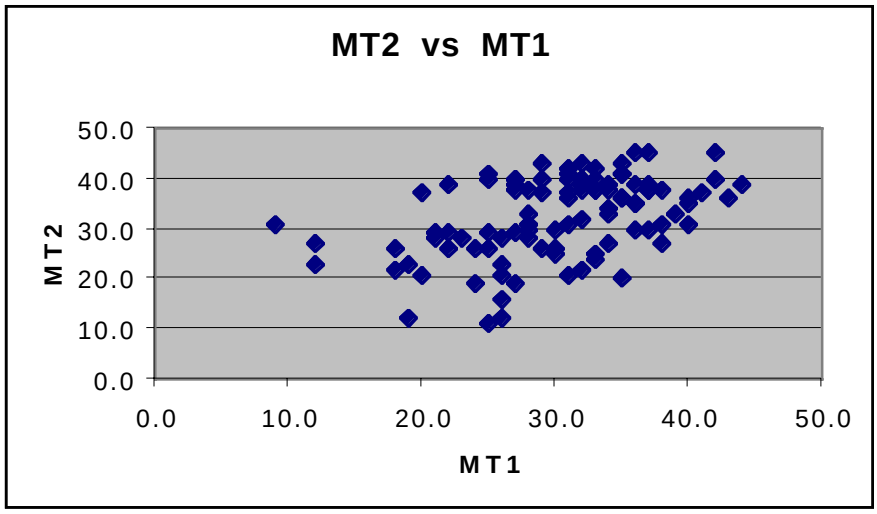
STAT 100 Chance and Data Analysis Nov. 13, 2002

Today: Re-mark procedure
MT2 results
Intro to Linear Regression (new topic – not on course outline)
Intro to Statistical Optimization (and Traveling Salesman Problem)

Re-mark procedure:

The procedure is to write down, on the front of your test or on an additional piece of paper, the reason you think you deserve more moarks than you got. Not just "Remark Q1" but what it is that, in view of the posted answer, you think deserves more marks. Robin Insley will then direct the paper to the TA that marked that question and you will be able to pick it up in the SW once it is re-marked. If you are not satisfied with the TA response and/or mark change, then see me.

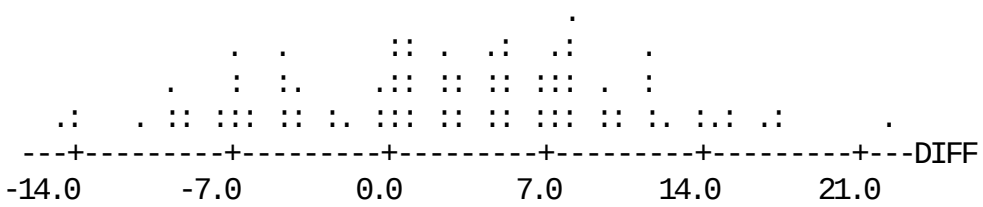
MT2 Results



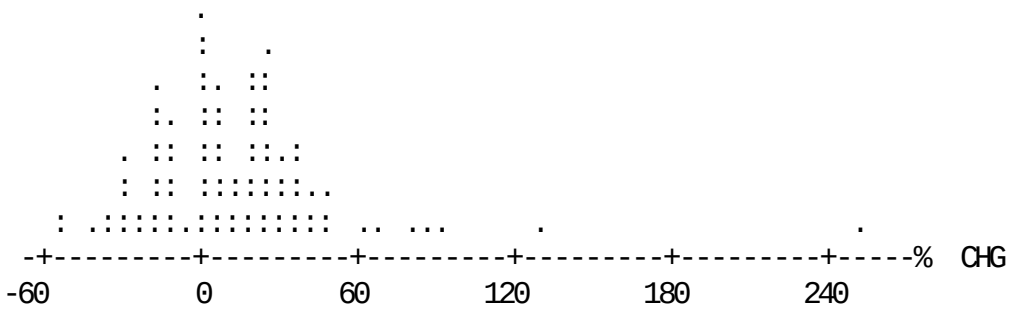
Correlation is +0.5

	MT1	MT2
Mean	30	32
SD	7	8
Q1	26	27
Q2	31	34
Q3	35	39
80^{th}	37	40

The difference DIFF=MT2-MT1 for those that did both:



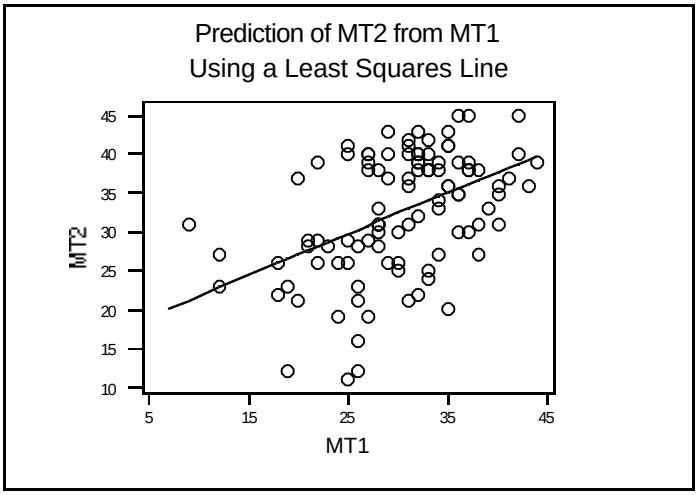
Another way to look at it is the percentage change 100*(MT2-MT1)/MT1



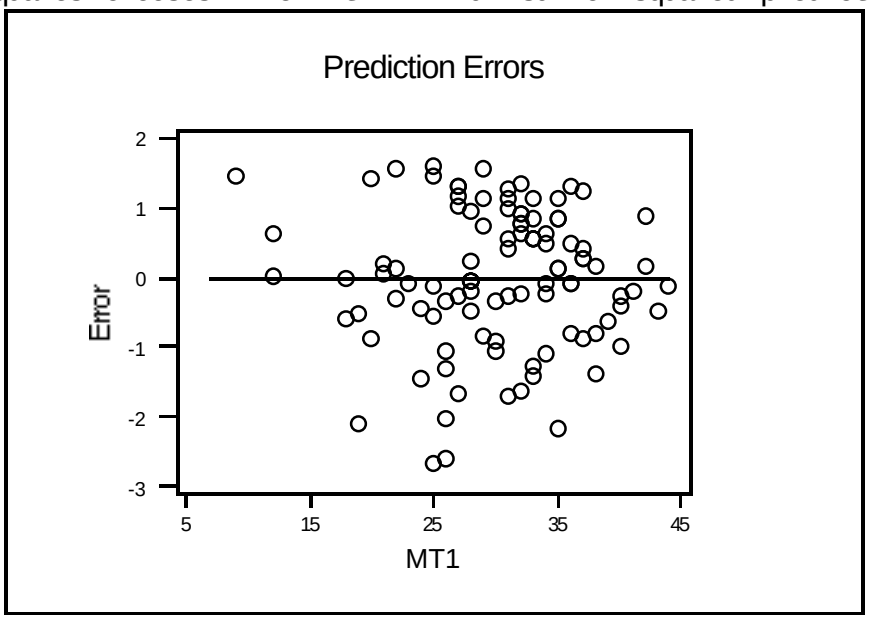
Linear Regression

A way to predict one variable from another variable.

The "Least Squares Line" or the "Prediction Line" or the "Regression Line" is the line that minimizes the sum of squared errors.

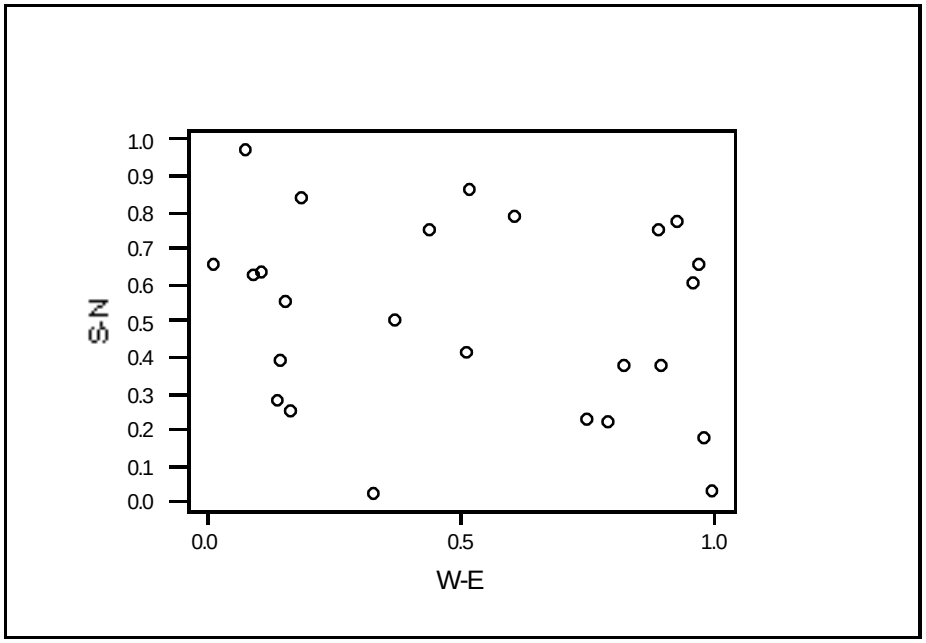


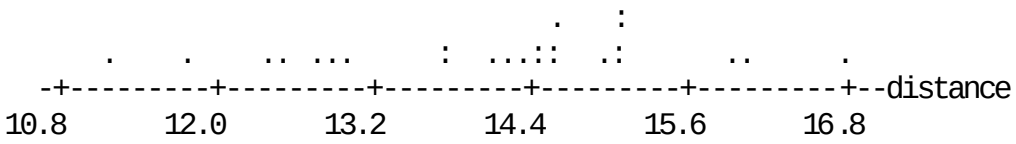
Least Squares chooses line with minimum sum of squared prediction errors



Introduction to Optimization

Read the article on Optimization in Tanur pp 241-248.

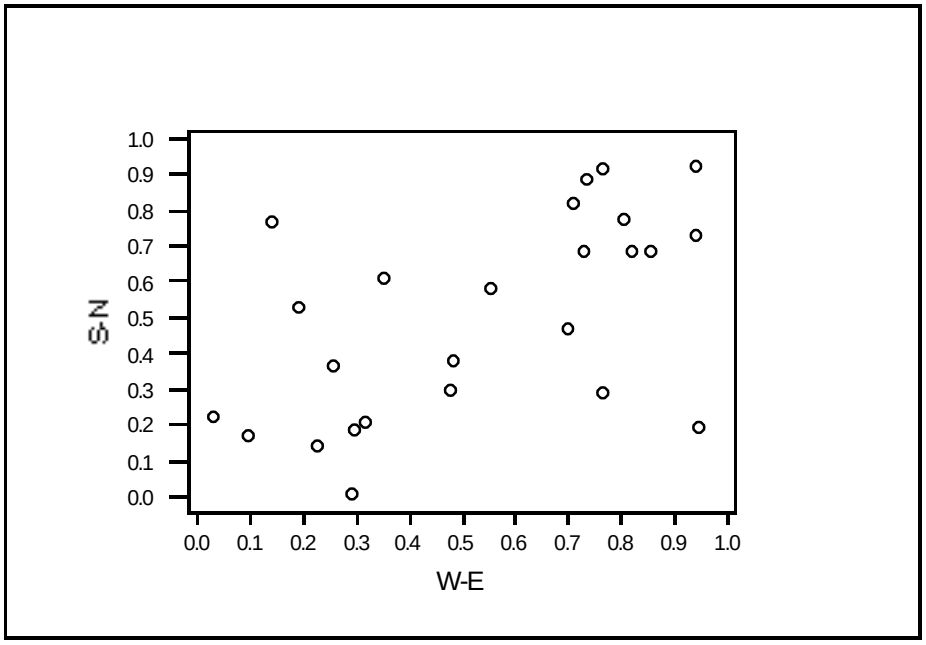


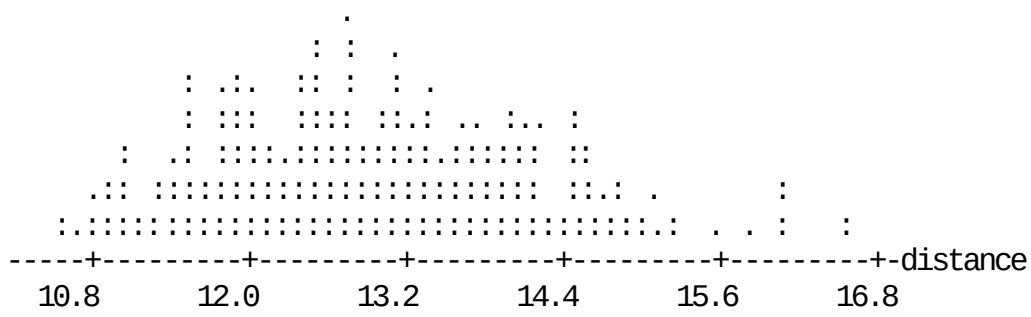


Best Route in 25 random tries:

Hor	Vert
0.98	0.07
0.23	0.75
0.39	0.14
0.56	0.15
0.63	0.09
0.25	0.17
0.75	0.89
0.38	0.82
0.38	0.89
0.42	0.51
0.50	0.37
0.28	0.14
0.61	0.95
0.65	0.97
0.78	0.93
0.84	0.19
0.63	0.11
0.66	0.01
0.75	0.44
0.22	0.79
0.86	0.51
0.79	0.60
0.03	0.99
0.17	0.98
0.02	0.33
0.98	0.07

Another example with 25 points but 250 random paths

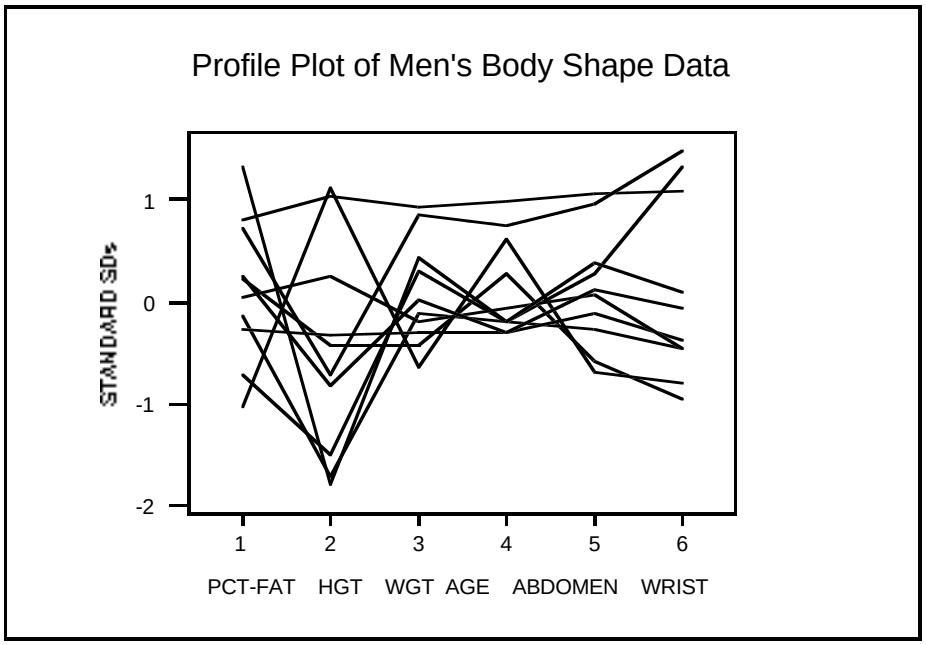




Big Picture ???

In situations where there are too many options to try in attempting to select the best one, random tries can find a nearly-best solution.

Another example: What is the best way to sort the variables for the following plot? What is the objective?



This plot is called a Profile Plot. Note that the data for each variable must be standardized so that all variables relate to the same scale. For example, here is the raw and standardized data for five men (not in the above plot, but typical),

AGE	WGT	HGT	s-AGE	s-WGT	s-HGT
39.00	234.75	74.50	-0.47	1.90	1.19
53.00	224.50	77.75	0.64	1.55	2.08
60.00	157.75	67.50	1.20	-0.72	-0.72
23.00	154.25	67.75	-1.74	-0.84	-0.65
47.00	184.25	74.50	0.17	0.18	1.19