

# Correlation between TV, physician and World's Life Expectancy



**BY**  
**JOAN MULIAWAN**



“There is a little **correlation** between the circumstances of people’s lives and how happy they are”

Dennis Prager



## Outline

- What is correlation?
- Correlation coefficient  $r$
- Classification
- Sample Data



## What is Correlation?

- Indicates the strength and direction of a linear relationship between two random variables.
- In general statistical usage, correlation or co-relation refers to the departure of two variables from independence.



## Correlation coefficient $r$

- Given two variables,  $X$  and  $Y$ , the correlation coefficient is a real value  $r \in [-1, 1]$ .

$$r = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum (x_i - \bar{X})^2 \sum (y_i - \bar{Y})^2}}$$

Also known as  
Pearson's  $r$



## Classification

- Classification is perhaps the most familiar and most popular data mining technique.
- Classification accuracy is usually calculated by determining the percentage of tuples placed in the correct class.



## Sample Data

Source :

\_The World Almanac  
and Book of Facts  
1993\_ (1993), New  
York: Pharos Books.

Country	People/TV	People/ physician	Life Expectancy
Argentina	4	370	70.5
Bangladesh	315	6166	53.5
Brazil	4	684	65
Canada	1.7	449	76.5
China	8	643	70
Colombia	5.6	1551	71
Egypt	15	616	60.5
Ethiopia	503	36660	51.5
France	2.6	403	78
Germany	2.6	346	76
India	44	2471	57.5
Indonesia	24	7427	61
Iran	23	2992	64.5



## Sample Data

Country	People/TV	People/ physician	Life Expectancy
South Africa	11	1340	64
Spain	2.6	275	78.5
Sudan	23	12550	53
Taiwan	3.2	965	75
Tanzania	496	25229	52.5
Thailand	11	4883	68.5
Turkey	5	1189	70
Ukraine	3	226	70.5
UK	3	611	76
United States	1.3	404	75.5
Venezuela	5.6	576	74.5
Vietnam	29	3096	65
Zaire	1000	23193	54





## Sample Data

Country	People/TV	People/ physician	Life Expectancy
Italy	3.8	233	78.5
Japan	1.8	609	79
Kenya	96	7615	61
North Korea	90	370	70
South Korea	4.9	1066	70
Mexico	6.6	600	72
Morocco	21	4873	64.5
Myanmar	592	3485	54.5
Pakistan	73	2364	56.5
Peru	14	1016	64.5
Philippines	8.8	1062	64.5
Poland	3.9	480	73
Romania	6	559	72
Russia	3.2	259	69



## Sample Data

Life expectancy class:

- 57 and below = low
- 57.5 – 69.5 = medium
- 70 and above = high

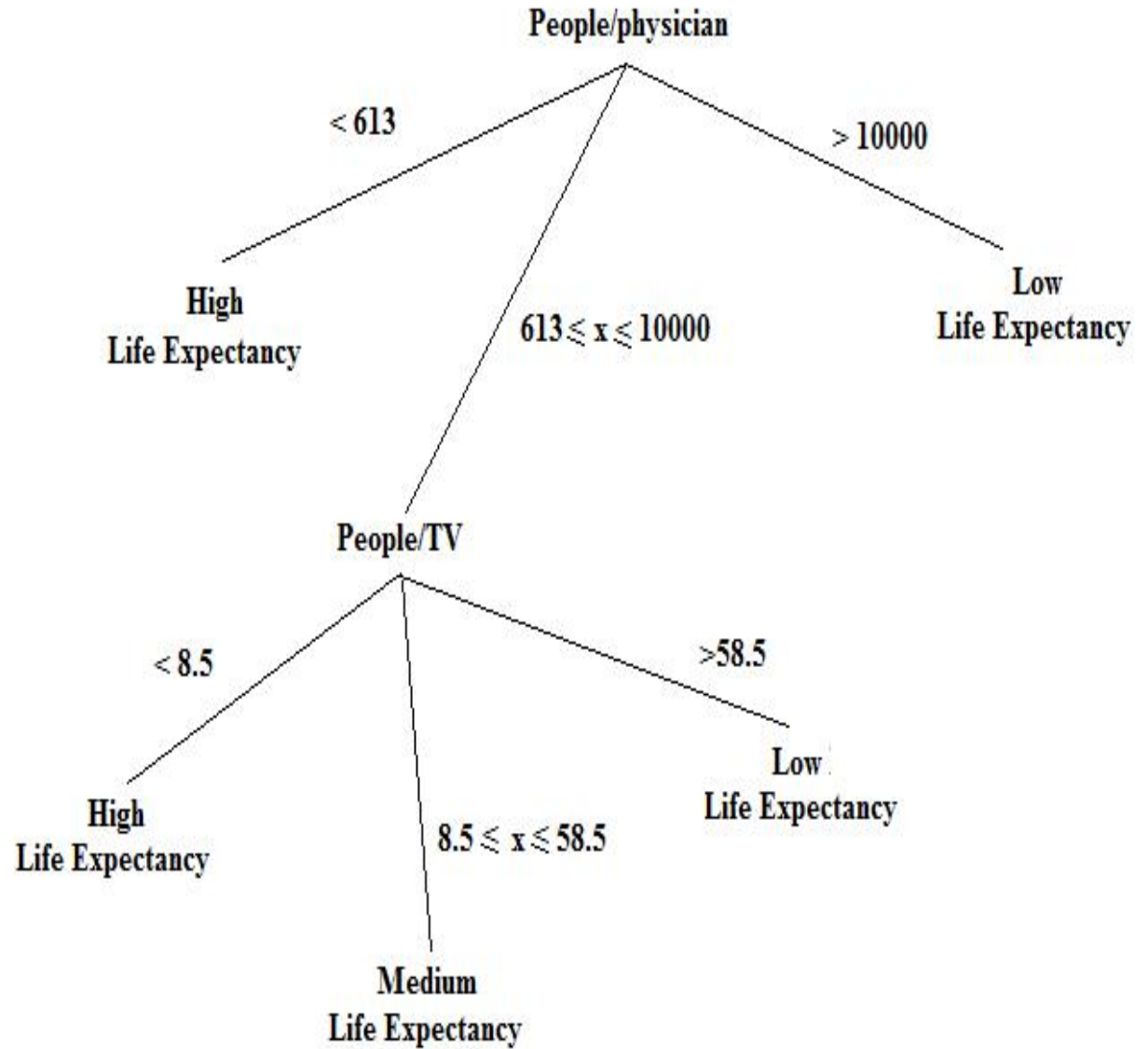
Correlation between people/TV and life expectancy is -0.623944.

Correlation between people/physician and life expectancy is -0.666.



Since people/physician attribute is having a higher correlation, I choose this over people/TV as the first splitting attribute.

# Decision Tree





## Sample Data

Country	People/TV	People/physician	Life Expectancy
Argentina	4	370	70.5
Bangladesh	315	6166	53.5
Brazil	4	684	65
Canada	1.7	449	76.5
China	8	643	70
Colombia	5.6	1551	71
Egypt	15	616	60.5
Ethiopia	503	36660	51.5
France	2.6	403	78
Germany	2.6	346	76
India	44	2471	57.5
Indonesia	24	7427	61
Iran	23	2992	64.5



## Facts

- On the one hand, Sweden, the Netherlands and Belgium all have very high life expectancies and are also considered very egalitarian.
- By comparison, Japan is less egalitarian but has a higher life expectancy.
- The United States, United Kingdom and Canada have far more income inequality than the above countries, as well as lower life expectancies.
- However, life expectancies in the U.S., U.K. and Canada are about the same, even though the U.K. and Canada have national health insurance dedicated to equal care for all and the U.S. does not.



## Source

- National Science for Policy Analysis. Retrieved May 6, 2008, from <http://www.ncpa.org/ba/ba328/ba328.html>
- Wikipedia. Retrieved May 6, 2008, from <http://en.wikipedia.org/wiki/Correlation>
- Dunham, Margaret. (2003) Data Mining: Introductory and Advance Topics. New Jersey: Pearson Education Inc.
- Retrieved May 6, 2008 from <http://thinkexist.com/quotes/with/keyword/correlation/>
- The World Almanac and Book of Facts 1993 (1993), New York: Pharos Books.



## Summary

- What is correlation?
- Correlation coefficient  $r$
- Classification
- Sample Data





**QUESTION ??**