STAT 100 **Notes** **Jan 19, 2010**

Assignment #3 (Due 4 pm Tuesday January 26)

1. With reference to the article "To Catch a Thief", describe how the telephone company can use computing methods to detect cell phone fraud. (Max 100 words).
2. With reference to the article "Assuring Product Reliability", explain why the reliability of the system is the product of the reliabilities of the components of the system. (Max 50 words)
3. With reference to the article "Advertising as a Engineering Science", why were the variables SUBJECT and DAY-OF-WEEK used to design the email allocation to registrants, whereas the information about AGE and DATE-OF-REGISTRATION was not used for the design? (Max 50 words).

Note: To obtain full marks, you must not exceed word limits.

STAT 100 pieces: Lectures, Notes, Readings, Assignments, and Stat Workshop.

Schedule for Week 3:

0. Preliminaries: Averages, and Moving Averages
1. Statistical Weather Forecasting: Wilks, pp 171-181
2. Randomness in the Stock Market: Clearly & Sharpe, pp 359-372
3. Some Technical Definitions and Statistics Jargon

0. Preliminaries

Recent Temperature Maximums for Vancouver, Jan 1-18, 2010

9,9,8,6,8,6,5,7,11,12,13,13,10,9,10,9,10,11 Celsius degrees

(go to vanc.temp.ppt)

Moving Average of Order 3

8.7,7.7,7.3,6.7,6.3,6.0,7.7,10.0,12.0,12.7,12.0,10.7,9.7,9.3,9.7,10.0

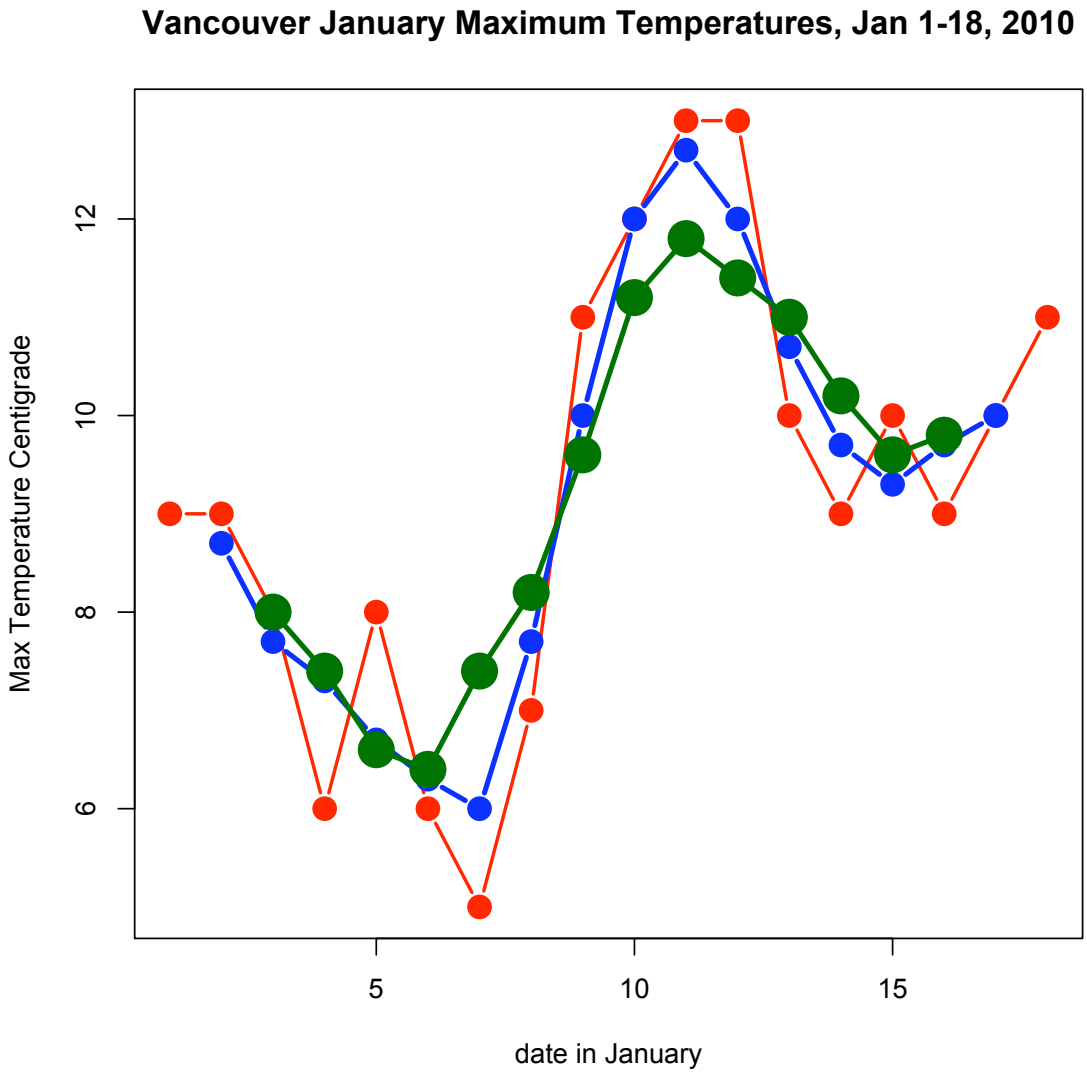
Moving Average of Order 5

8,7.4,6.6,6.4,7.4,8.2,9.6,11.2,11.8,11.4,11.0,10.2,9.6,9.8

These data appear in the graph below:

Red is the original data (18 points)
Blue is order 3 moving average (16 points)
Green is order 5 moving average (14 points)

Higher Order means more smoothing – cuts off peaks and valleys.



Q: Does the apparent trend help to forecast the future temperatures? Might meteorology (physics of air flows, water currents, etc) help? Hint: Remember the random walk, and how the apparent trends did not help to predict future changes?

Q: Would this technique of moving averages provide the seasonal trend apparent in the fuel consumption data?

1. **Statistical Weather Forecasting: Wilks, pp 171-181**

Edward Norton Lorenz (May 23, 1917 - April 16, 2008)
Introduced
"The butterfly effect 1963"
Physics of meteorology is very complex.
Computer advance offered hope of solution and better forecasts, but
Chaos Models of Lorenz showed that deterministic solution practically impossible.
He showed that there could be deterministic models with chaotic outcomes.
Sensitivity to initial conditions was great, and initial conditions impossible to pin down. Instead use a probability distribution of initial conditions and generate distribution of possible outcomes.
Using this as a probability distribution: "dynamic models"

Purely statistical forecast (no physics) : OK for < 6 hours or >10 days (Miller reference)

Recent Books of General Interest about Probability and Statistics (not part of STAT 100)

Mlodinow, L (2008) *The Drunkard's Walk*. Vintage Books. New York.

Levitt, S.B. and Dubner, S.J. (2005*) Freakonomics.* Harper Collins. Toronto.

Levitt, S.B. and Dubner, S.J. (2009*) Super Freakonomics.* Harper Collins. Toronto.

Rosenthal, J.S. (2005) *Struck by Lightning* Harper Perennial. Toronto.

Gardner, D. (2008) *Risk* McClelland and Stewart. Toronto.

Tanur, J.M. et al (eds) (1989) *Statistics: A Guide to the Unknown*. Third Edition. Duxbury. Belmont.

Taleb, N. N. (2008) *Fooled by Randomness: The Hidden Role of Chance in the Markets and Life,* 2nd Edition. Random House.

Gladwell, Malcolm (2008) *Outliers: The story of Success*. Little, Brown & Co.

Randomness in the Stock Market: Clearly & Sharpe, pp 359-372

1d | 2d | 5d | 1m | 3m | 6m | 1y | 3y | 5y | 10y | **20y** | Max

1d | 2d | 5d | 1m | 3m | 6m | **1y** | 3y | 5y | 10y | 20y | Max

Quotes delayed 15 minutes

1d | 2d | 5d | 1m | 3m | 6m | **1y** | 3y | 5y | 10y | 20y | Max

Comments in article:

P 359. Explanations in retrospect are plausible and suggest that such knowledge would help to predict the market over the short term. This is what market players like to think. But the randomness of the market suggests this predictability is an illusion.

Pp 362-364. Methods suggested for short term forecasting are: moving average, regression line, autoregressive model. Latter seems best although it is not very different from the prediction "tomorrow's price is the same as today's price" – ie. like a random walk.

P 365. Random Walk – steps need not be +1 and -1. Perhaps normally distributed with mean 0. (Bell curve).

P 366. Counterintuitive that random walk can produce apparent trends (but not trends useful for prediction).

P 368. Long Term Trends – need inside info or else external explanatory variables like GNP.

P 369-370. Stock analysts and advice newsletters usually worse than passive investment (one that invests in the whole market (or most of it) all the time).

Technical Items Required for Future Discussions: (more notes in preparation)

Standard Deviation (SD) – a measure of data spread or "variability"

Normal Distribution - a model for the relative frequency of values with a certain mean and SD.

Standard Normal Distribution – a Normal Distribution with Mean 0 and SD = 1

Regression – a way to predict one variable from another variable (or variables)

These will be discussed in detail on Thursday, Jan 21.

KLW 2010/01/19