Required for next week: February 13, 1996

- "Should pregnant women move? Linking risks for birth defects with proximity to toxic waster sites", S.A. Geschwind. *Chance* **5**(3-4), 40–46 (1992). Come prepared to discuss this article and ask questions about the parts you didn't understand.

- Please pick up a copy of *Tainted Truth* by Cynthia Crossen; it's at the bookstore under SCI199Y. We'll start discussing it after reading week.

Quebec referendum: addendum

On January 27, 1996, a story on the front page of the *Globe and Mail* said "Most Quebeckers expect separation within 10 years". The article summarized results from a poll conducted by Leger and Leger, for the Journal de Montreál and the Globe and Mail. In a continuation of the article on p.8, the proportion polled indicating support for the "yes" and "no" sides for all the polls taken by Leger and Leger since last July is plotted. We saw most of this graph before: two points added were a post-referendum poll conducted in November, 1995, and the poll discussed in this article.

The first paragraph says "three in four Quebeckers believe the province will become a sovereign country some day and about 60 per cent expect that the change will occur within 10 years". The poll was reported carefully, though, and on p.8 we find "When asked 'Do you believe that Quebec will become a sovereign country within...', 2.9% replied within 1 year, 20.2% said within 2 or 3 years, 24.2% answered within 4 or 5 years, 14.5% predicted it would be within 6 to 10 years and 12% thought it would take more than 10 years. Only 21.9% thought Quebec would never become sovereign."

What do you think about the phrasing of the question?

A series of questions were asked on whether Quebeckers feel they get a good financial deal from the rest of the country. Here's an example: "Does Quebec receive or not receive less than its share of federal government spending in the provinces?". Result: 50% said less; 37% said not less.

Statistics Canada data indicates that in fact in Quebec Ottawa raises $4,107 per capita and spends $4,286.

Regression

In the simplest version of regression, we have just two variables, usually called y and x. Examples are given in the Paulos handout of last week, and in the Chance article for this week. Sometimes this is called "simple linear regression".

The least squares line that is fitted to a graph of y against x is meant to summarize the relationship between y and x. In particular, if this line has a slope of zero,

then there is no association between the two variables. Since there is always some variability in the measurements, the statistical question of interest is whether or not the slope is 'significantly different from zero', usually assessed by a hypothesis test, as described in the notes of November 21 and December 5. If the slope is greater than (less than) zero, then this is evidence that an increase in the x variable is associated with an increase (decrease) in the y variable. (This is equivalent to the assertion that x and y are positively (negatively) correlated.)

This type of analysis can be generalized in several ways. In the marathon runners example, least squares was used to fit a cubic polynomial instead of a straight line. There was a reference to the cubic coefficient being 'marginally significant' for the Osaka races, meaning that the estimate for the cubic term was just barely larger than zero. A second very common generalization is to use least squares to summarize the relationship between a single y variable, such as health status, and several different x variables, such as treatment, age, sex, prior history of disease, and so on. This usually goes under the name 'multiple regression', because there are multiple x variables that are potentially associated with the variable y. Although it is almost impossible to plot a picture of multiple regression, it is basically a fairly straightforward extension of simple linear regression. There are a number of important technical aspects to both these generalizations, though: there is a 3rd year half course (STA302) on regression.

In the Globe and Mail this week

- "Quebec tops in venture capital", Jan. 30, A8. A study that is part of a five year, $3.5-million research project at UBC on entrepreneurship in Canada notes that although QAuebec has one-quarter of the country's population and produces 23% of its goods and services, it is the home of 42% of venture-financed companies that were studied. Ontario, meanwhile, has 38 % of the population, produces 40% of Canada's output and has only 22% of the venture-backed companies.

- "Violent crime may be on upswing", Feb. 1, A2. "As of yesterday, Metro Toronto had recorde eight homicides during January, double last year's toll and second only to the 10 killings in January in 1991. That year the city went on to record an unprecedented 88 homicides, compared with its average of about 60. Last year's toll was 57." Do you think it's reasonable to make predictions based on January's data? Is it clear that 88 is considerably larger than 60? There's been a lot of discussion in the media about whether the crime rates are going up or down, and why.

- "Study gives baby boomer unhealthy grades", Jan. 31, A6. As one of the culprits, couldn't resist.