University of Illinois | **Final Exam** | Department of Economics
Spring 2013 | **Economics 574** | Professor Roger Koenker

The ultimate objective of field courses is (presumably) to prepare you to digest new material in future research. This year's final consists of three parts: the first is very abstract, the second is very empirical, the third is more conceptual. The exam is due friday, May 10 at 5pm. Feel free to email or stop by my office if you feel that there are points that need clarification; I'll post corrections or clarifications on the course webpage, if necessary.

The first question is simply a very brief hint about a completely new topic that I find especially fascinating, but didn't have time (or more the point, sufficient knowledge of) to pursue in the class. The other two questions were motivated by emails that I received from R-help, the first was earlier suggested as a possible paper topic for 574, but was a bit too simplistic to make a full paper, the second question is more open-ended, and I'm just looking for some organized thoughts on the central questions.

1. In a landmark 1781 paper the French mathematician and civil engineer Gaspard Monge explored the following problem of "optimal mass transportation:" Suppose you have a pile of sand and an excavated hole you wish to fill with the sand; the cost of moving the sand from point x in the pile to point y in the hole is proportional to the distance between x and y. What is the optimal way to accomplish the task of filling the hole? Interest in this problem was revived by Kantorovich in the 1940's and it has developed rapidly in recent years. Even in its original form the problem is very challenging, despite its apparent simplicity. The rapid recent development of this subject seems to have a wide variety of applications in statistics. In economics it is particularly relevant to the matching literature.

   In one dimension it has the following probabilistic interpretation: Suppose P and Q are probability measures on $(\mathbb{R}, \mathcal{B})$ with respective marginal distributions F and G, and joint distribution H. Let

   $$R(P, Q) = \inf \mathbb{E}|X - Y|$$

   where the inf is taken over all random variables X and Y with distributions P and Q respectively. Recall that one possible construction of such an X and Y is to take $Z \sim U[0, 1]$ and set $X = F^{-1}(Z)$ and $Y = G^{-1}(Z)$, so X and Y are comonotonic and then

   $$\mathbb{E}|X - Y| = \int_0^1 |F^{-1}(u) - G^{-1}(u)| du$$

   Provided that $\mathbb{E}|X - Y| < \infty$, it can be further shown that for any $X \sim F$ and $Y \sim G$,

   $$\mathbb{E}|X - Y| \geq \int_0^1 |F^{-1}(u) - G^{-1}(u)| du = \int_{-\infty}^{\infty} |F(x) - G(x)| dx$$

   The equality on the right hand side "follows from elementary geometric considerations" in the words of an early paper in this literature.

   (a) Draw a picture (with caption) to illustrate these "considerations."

(b) The inequality in the above expression requires a bit more work, but basically boils down to the Frechét bound, $H(x, y) \leq \min\{F(x), G(y)\}$. Thus

$$R(P, Q) = \int_{-\infty}^{\infty} |F(x) - G(x)| dx$$

In this form R is sometimes referred to as Mallows metric for probability measures. Try to explain briefly some intuition for why the inf is attained by this pair of comonotonic random variables.

2. The following question is directly quoted from a recent R-help suggestion to make a tutorial on median regression in R based on a recent paper by Robert Vanderbei. The suggestion was made by a J.C. Nash, presumably not to be confused with J.F. Nash. I don't endorse the comment at the end about median regression being overlooked, on the contrary I'd say it is rather underlooked.

> In the recent SIAM Review, vol 54, No 3, pp 597-606, Robert Vanderbei does a nice analysis of daily temperature data. This uses publicly available data. A version of the paper is available at
>
> `http://arxiv.org/pdf/1209.0624`
>
> and there is a presentation at
>
> `http://www.princeton.edu/~rvdb/tex/talks/GERAD/LocalWarming.pdf`
>
> This would make a nice case for a vignette showing how to do such an analysis in R, possibly as a project for a senior undergraduate or perhaps even at the Master's level if some tools were developed, since Vanderbei presents a good argument for using Least Absolute Deviations regression (LAD). Generally LAD is overlooked by statisticians, maybe because the tools are unfamiliar.

This can be done quite easily in R with the `rq( )` from the `quantreg` package. You can either use the data that Vanderbei used, or try to use some counterpart for Champaign-Urbana.

3. Again quoting from an email message I received:

> I'm very interested in how increases for funding for Early Childhood education may have differential effects on the outcomes of children on a distribution. I think that Quantile Regression is an excellent tool for answering this question but I'm unsure on a lot of the details. As I've gotten more into the literature regarding Quantile Regression, I've realized that my theoretical math skills are not where they need to be to be able to answer the questions that I want to answer.
>
> I've read this paper, `http://www.sbe.org.br/dated/ebe28/pdf/41.pdf`, about unconditional quantile regression, but am quickly realizing that I'm not willing to take their word that they've solved the problem that they can predict unconditional quantiles based on changes of x using their response function without knowing the underlying math. I also trust your judgement as the father of quantile regression far more than I trust these guys but ideally I'd like to get to the theoretical foundations to be able to make these judgements on my own.

> While I love my current job, I can't find the time to dedicate to this intellectual endeavor. I was wondering what you think about unconditional quantile regression, but also what you would recommend to someone who wants to contribute meaningfully to science but doesn't trust anything he doesn't understand.

Try to write a polite answer to this question. My response, which was very terse, suggested that he might look at:

`http://www.econ.uiuc.edu/~roger/courses/574/readings/dreg.pdf`

and the references cited there.