ESD.86
Models, Data, Inference for Socio-Technical Systems
Massachusetts Institute of Technology
Engineering Systems Division
Problem Set #6
Issued: Wednesday 11 April, 2007. Due: Monday 23 April, 2007.

1. The *t*-distribution and the *t*-test. Consider the paper "Student", 1908, "The Probable Error of a Mean," *Biometrika* **6**:1-25. Regarding this paper, R. A. Fisher wrote "The study of the exact distributions of statistics commences in 1908 with Student's paper ... Once the true nature of the problem was indicated, a large number of sampling problems were within reach of a mathematical solution..." This suggests how important the paper is and why some time spent with the ideas in it will be useful.

a) In a paragraph, summarize the motivation for the paper and what it accomplished. In so far as possible, do so without assuming the readers knows any technical jargon whatsoever. Imagine, for example, the reader is a very bright high school senior who might be expected to know concepts from algebra, geometry, and calculus, but not yet be exposed to concepts from probability and statistics.

b) Assume z is a standard normal random variable and u is χ^2 with k degrees of freedom. Assume z and u are independent. Sketch out a derivation of the probability density function $f_x(x)$ of the random variable x defined as

$$ x = \frac{z}{\sqrt{u/k}} $$

You may find it useful to know that when you make a change of variables so that $x=x(u,v)$ and $y=y(u,v)$ then $f_{xy}(x,y) = f_{uv}(x,y) \, | \, \mathbf{J} \, |$ where $\mathbf{J}$ is the Jacobian matrix of the transformation

$$ J = \begin{bmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{bmatrix} $$

c) Student's "Illustration *I*" (p. 238-9) concerns data from 10 patients on the effect of two soporifics. Analyze this data using both a *t*-test and a randomization (permutation) test. Comment in a few sentences on the results obtained.

2. Regression. Find an application of regression in the technical literature. Preferably, the application will be from the past 5 years and will be in your field of interest.

a) Provide a complete citation.

b) Summarize the application in less than one page of text and graphics.

c) What conclusions were drawn based on the regression. Discuss their validity. Were appropriate diagnostics performed?

d) Were the results communicated well? If there are opportunities for improvement, discuss them.