Problem Set #6

1. ***The t-distribution and the t-test***

a)

b)

Normal pdf: $g(z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \forall z \in \Re$

Chi-squared pdf: $h(u) = \dfrac{1}{\Gamma\left(\dfrac{k}{2}\right)} \left(\dfrac{1}{2}\right)^{\frac{k}{2}} u^{\frac{k}{2}-1} e^{-\frac{u}{2}}, \forall u \in \Re^+$

Since Z and U are independent, $f(z,u) = g(z).h(u)$

Thus, $f(z,u) = \dfrac{u^{\frac{k}{2}-1} e^{-\frac{z^2-u}{2}}}{\sqrt{2\pi}.\Gamma\left(\dfrac{k}{2}\right)} \left(\dfrac{1}{2}\right)^{\frac{k}{2}}, \forall u \geq 0$

Change of variables: $f_{x,y}(x,y) = f_{z,u}(x,y)|J|$, where J is the determinant of the Jacobian matrix.

Here, $|J| = \begin{vmatrix} \dfrac{1}{\sqrt{\dfrac{u}{k}}} & -\dfrac{z}{2}\sqrt{\dfrac{k}{u}} \\ 0 & 1 \end{vmatrix} = \sqrt{\dfrac{k}{y}}$

Therefore, $f(x,y) = \sqrt{\dfrac{y}{k}} \left(\dfrac{1}{\sqrt{2\pi}.\Gamma\left(\dfrac{k}{2}\right)}\right) \left(\dfrac{1}{2}\right)^{\frac{k}{2}} .y^{\frac{k}{2}-1} e^{-\frac{y}{2}.(x^2+1)}, \forall y \geq 0$

We integrate now over the entire y domain to obtain:

$f(x) = \int_0^{+\infty} f(x,y).dy = \left[\dfrac{1}{\sqrt{2\pi}.\Gamma\left(\dfrac{k}{2}\right)}\right] \left(\dfrac{1}{2}\right)^{\frac{k}{2}} . \int_0^{+\infty} \sqrt{\dfrac{y}{k}} .y^{\frac{k}{2}-1} e^{-\frac{y}{2}(x^2+1)}$

The result of this integral has already been demonstrated (cf. statistics textbooks):

$f(x) = \dfrac{\Gamma\left(\dfrac{k+1}{2}\right)}{\Gamma\left(\dfrac{k}{2}\right)} \dfrac{1}{\sqrt{k\pi}\left(1 + \dfrac{x^2}{k}\right)^{\frac{k+1}{2}}}$

This is the pdf of the Student's t-distribution.

c)

2. *Regression*

a)

The papers that have been selected are:

- Badhwar G.D. & Anz-Meador P.D. 'Determination of the area and mass distribution of orbital debris fragments'. *Earth, Moon and Planets*. Vol.45, No.1, pp.29-51. 1989.
- Finkler M. 'Healthcare cost differences in the 1990s: The influence of metropolitan area marketplace dynamics'. *Research in Healthcare Financial Management*. Vol.10, No.1, pp.27-42. 2005.
- Keating E.G. & Dixon M. *Investigating optimal replacement of aging Air Force systems*. Santa Monica: RAND. 2004.
- Hendricks K.B. & Singhal V.R. 'The effects of supply chain glitches on shareholder value'. *Journal of Operation Management*. Vol.21, No.5, pp.501-522. 2003.
- Hsieh M. & Magee C. 'Standards as independent artifacts: Application to prediction of promotion for Internet standards'. *ICCS'06: International Conference on Complex Systems*. 2006.
- Hua S.Y. & Wemmerlov U. 'Product change intensity, product advantage, and market performance: An empirical investigation of the PC industry'. *Journal of Product Innovation Management*. Issue 23, pp.316-329. 2006.
- Murata A. & Furukawa N. 'Relationships among display features, eye movement characteristics, and reaction time in visual search'. *Human Factors*. Vol.47, No.3, pp.598-612. Fall 2005.
- Li S. & Lin B. 'Accessing information sharing and information quality in supply chain management'. *Decision Support Systems*. Issue 42, pp.1641-1656. 2006.
- Shoemaker P.J.H. & Waid C.C. 'An experimental comparison of different approaches to determining weights in additive utility models'. *Management Science*. Vol.28, No.2. February 1982.
- Powell W.W., Koput K.W. & Smith-Doerr L. 'Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology'. *Administrative Science Quarterly*. Vol.41, No.1, pp.116-145. March 1996.
- Tso G.K. & Yau K.K. 'Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks'. *PROBASTAT 2006: Fifth International Conference on Mathematical Statistics*. Smolenice Castle: June 2006.
- Van Belle D.A. 'Bureaucratic responsiveness to the news media: Comparing the influence of The New York Times and network television news coverage on US foreign aid allocations'. *Political Communication*. Vol.20, No.3, pp.263-285. 2003.
- Weiss M.B.H. & Sirbu M. 'Technological choice in voluntary standards committees: An empirical analysis'. *Economics of Innovation and New Technology*. Issue 1, pp.111-133. 1990.

b)

Here is the summary of one paper (Badhwar & Anz-Meador, 1989), by Matt Richards (Courtesy of Matt Richards. Used with permission.):

> Understanding the LEO orbital debris environment is critical for conducting hazard analyses for LEO spacecraft. A critical element in modeling the LEO orbital debris environment is the loss rate of debris due to atmospheric drag. Calculating the effects of atmospheric drag requires knowledge of the area-to-mass ratio of debris fragments.

However, given that debris originates from anti-satellite tests and breakups of spacecraft, precise data on debris mass, size, and shape is not available. Inverting the technique used to determine atmospheric density from the observed decay of satellites with know area-to-mass ratios, Badhwar and Anz-Meador develop a methodology for inferring the area-to-mass ratio of debris fragments based upon observations of their orbital elements.

First, observed radar cross-section (RCS) values are correlated to the effective area of space objects by comparing RCS measurements from the Eglin radar operating at 70 cm wavelength to 196 satellites of known shape, size, and mass. Knowledge of these 196 satellites was used to calibrate the calculated mass to the observed mass and to develop a relationship between the mean RCS and the effective area of the object. Figure 2 from the paper (below) plots the effective (projected surface) area versus the mean of the RCS for the objects with known geometry as well as the best-fit least-squares line to a power law fit.

Figure 2 from paper removed due to copyright restrictions.

Upon establishing a power law relationship between mass and effective area, the correlations were applied to 2600 debris fragments. By dividing the RCS of the debris by the area-to-mass ratio (calculated using the orbital elements), Figure 6 (below) plots the calculated mass against the effective area.

Figure 6 from paper removed due to copyright restrictions.

c)

The discussion and improvement suggestions for the paper (by the same student) follow:

Based on the data shown in Figure 2 and in Figure 6 (shown on the previous page), the authors propose a power law relation for effective area to RCS ($A\_eff=k[RCS]^{.76}$) and a power law relation, $m=kA^{(1.86)}$, for relating debris mean mass to mean effective area. Although valid for capturing general trends, I would not characterize the validity of these relationships as precise. The authors note in the conclusion that their overall technique is valid to only about 30%.

In the text describing Figure 2, the authors do a good job of performing diagnostics on the least-squares fit. The proportion of variance accounted for is computed with an R-square calculation of .578, indicating only a moderately good fit. Furthermore, the authors do a good job discussing problems with the regression's validity, notably the assumption that all objects were tumbling in random directions (when observed by the Eglin radar). In actuality, these objects would be expected to become gravity gradient stabilized in a given orientation, presenting a non-random face to the radar beam.

In the discussion of Figure 6, the authors do a less-than-stellar job diagnosing the least-squares fit. With the best fit line intended to inform a hazard analysis model, the plot of mean debris mass and the mean effective area is the critical output of the paper. However, there is no discussion of the proportion of variance accounted for nor a discussion of whether the data satisfies the assumptions required for inferences to be discussed. For example, Figure 6 clearly exhibits heteroscedastic data with a larger spread of calculated masses towards the lower-end of effective areas. Given that effective area is derived from the RCS, it is understandable that there would be more noise in the data for smaller pieces of orbital debris. However, the authors do not discuss why a linear regression was used despite a lack of homoscedastic data. The spread in values is attributed to a variety of reasons, but only increased heterogeneity among smaller objects (e.g., objects with larger masses are expected to be largely aluminum while smaller objects can be aluminum or less dense plastics or insulation materials) is discussed. To address the need for normally distributed data, Figure 7 does a good job showing that the log of the mass is fairly Gaussian.

d)

The results were communicated adequately but there is room for improvement. Technically, it would be better if quantitative measures were calculated for assessing the validity of the linear regression model. This would entail calculating the R-squared value for Figure 6. Additionally, given the large number of data in Figure 6, it would be possible to conduct a Chi-squared goodness-of-fit test for each slice of the calculated mass / effective area data. Rather than simply showing a three-dimensional representation of the data to show normality, the Chi-squared test would provide a quantitative measure of the validity of the linear regression. In terms of presentation, the overall model was communicated adequately. An overview flowchart mapping the calculations in the introduction would have facilitated understanding.