Show all work. This assignment is due on 11/30.

1. Describe a test that would enable you to determine whether the appropriate underlying distribution for a duration model was a Weibull versus an exponential. (20 points)

2. One way to write the negative binomial distribution is
\[
\binom{r+y-1}{y} p^{r-1} q^y
\]
where \( 0 < p \leq 1, \ q = 1 - p, \ r > 0, \ E[y] = \frac{r q}{p}, \) and \( \text{var}[y] = \frac{r q}{p^2}. \) Show that this is equivalent to the expression for the negative binomial we discussed in class for \( p = 1/\sigma^2 \) and \( r = \lambda/ (\sigma^2 - 1). \) (20 points)

3. The file duration.txt (available from the usual place) is a data set where the first column of duration.txt contains the time at which a failure occurred while the remaining columns contain explanatory variables \((x_1 \text{ and } x_2)\) for each observation.

   (a) Using the Kaplan-Meier estimator of the hazard function, plot the hazard function against time. What type of underlying distribution do you think would be appropriate for this data? (10 points)

   (b) Estimate a weibull duration model using this data. Interpret your results. If you have trouble estimating the model try using the likelihood transformation described by Greene, pp. 991-992. (10 points)

   (c) Do some specification checks to see if the choice of distribution is appropriate. (10 points)

4. The file count.txt (you know where to get it) contains event count data. The first column of count.txt contains the dependent variable while the remaining columns contain explanatory variables \((x_1 \text{ and } x_2)\) for each observation.

   (a) Estimate a Poisson regression model. Report and interpret your results. (15 points)

   (b) Check to see if the Poisson model is appropriate for these data. If it is not, estimate a model that is appropriate and explain why it is appropriate. (15 points)