

STA 312f23 Assignment Seven¹

The paper and pencil part of this assignment is not to be handed in. It is practice for Quiz 7 on November 3d. The R part may be handed in as part of the quiz. **Bring hard copy of your printout to the quiz.** Do not write anything on your printout in advance except possibly your name and student number.

1. You have done these exercises before. It's just a warm-up. For the Weibull distribution (see formula sheet), show that
 - (a) $E(T^k) = \frac{\Gamma(1+\frac{k}{\alpha})}{\lambda^k}$
 - (b) Median = $\frac{[\log(2)]^{1/\alpha}}{\lambda}$
 - (c) $h(t) = \alpha\lambda^\alpha t^{\alpha-1}$
2. Let T have a standard exponential distribution (that is, $\lambda = 1$), and let $Z = \log(T)$. Derive the density of Z . We will call this the *standard Gumbel*, even though the term is also applied to $-\log(T)$.
3. Let Z have the standard Gumbel density of Question 2.
 - (a) Find the moment-generating function of Z , $M_z(t) = E(e^{Zt})$.
 - (b) Differentiating the moment-generating function and setting $t = 0$, show that $E(Z) = \Gamma'(1)$ (derivative of the gamma function).
 - (c) Find the median of Z .
 - (d) Find the mode of Z .
4. Let Z have the standard Gumbel density of Question 2, and let $Y = \sigma Z + \mu$, where $\sigma > 0$.
 - (a) Find the density of Y .
 - (b) Find the expected value of Y .
 - (c) Given that the variance of the standard Gumbel is $\frac{\pi^2}{6}$ (you don't have to prove this), what is the variance of Y ?
 - (d) Find the median of Y .
 - (e) Find the mode of Y .

¹This assignment was prepared by [Jerry Brunner](#), Department of Mathematical and Computational Sciences, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](#). Use any part of it as you like and share the result freely. The L^AT_EX source code is available from the course website: <http://www.utstat.toronto.edu/brunner/oldclass/312f23>

5. Let $T \sim \text{Weibull}(\alpha, \lambda)$, and $Y = \log(T)$.
- First, find the density of Y .
 - Now re-parameterize, meaning express the parameters in a different, equivalent way. Let $\alpha = \frac{1}{\sigma}$ and $\lambda = e^{-\mu}$. Write the resulting density of Y .
 - If necessary, re-write the density of Y so that it has the form of the Gumbel density of Question 4.

The point of all this is that if you believe the distribution of a set of failure time data could be Weibull (a popular choice), you can log-transform the data and apply a Gumbel model. The Gumbel distribution may be preferable because the parameters μ and σ are easy to interpret.

6. High School History classes from across Ontario are randomly assigned to either a discovery-oriented or a memory-oriented curriculum in Canadian history. At the end of the year, the students are given a standardized test and the median score of each class is recorded. Please consider a regression model with these variables:

- X_1 Equals 1 if the class uses the discovery-oriented curriculum, and equals 0 the class it uses the memory-oriented curriculum.
- X_2 Average parents' education for the classroom
- X_3 Average parents' income for the classroom
- X_4 Number of university History courses taken by the teacher
- X_5 Teacher's final cumulative university grade point average
- Y Class median score on the standardized history test.

The full regression model has $E[Y|\mathbf{x}] = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5$. For each question below, give the null hypothesis in terms of β values. Also give $E[Y|\mathbf{x}]$ for the *restricted* (reduced) model you would use to answer each question. Don't re-number the variables.

- If you control for parents' education and income and for teacher's university background, does curriculum type affect test scores?
- Allowing for parents' education and income and for curriculum type, is teacher's university background (two variables) related to students' test performance?
- Correcting for teacher's university background and for curriculum type, are parents' education and income (considered simultaneously) related to students' test performance?
- Adjusting for curriculum type, teacher's university background and parents' education, is parents' income related to students' test performance?

7. In a study comparing the effectiveness of different exercise programmes, volunteers were randomly assigned to one of three exercise programmes (A , B , C) or put on a waiting list and told to work out on their own. Aerobic capacity is the body's ability to process oxygen. Aerobic capacity was measured before and after 6 months of participation in the program (or 6 months of being on the waiting list). The response variable was improvement in aerobic capacity. The explanatory variables were age (a covariate) and treatment group. Note that treatment group is a variable with four categories; wait list is one of them.

Consider a regression model with an intercept, and no interaction between age and treatment group.

- Make a table showing how you would set up indicator dummy variables for treatment group. Make Waiting List the reference category
 - Write the regression model. Please use x for age, and make its regression coefficient β_1 .
 - In terms of β values, what null hypothesis would you test to find out whether, allowing for age, the three exercise programmes differ in their effectiveness?
 - Write the null hypothesis for the preceding question as $H_0 : \mathbf{L}\beta = \mathbf{0}$. Just give the \mathbf{L} matrix.
 - In terms of β values, what null hypothesis would you test to find out whether Programme B was better than the waiting list?
 - In terms of β values, what null hypothesis would you test to find out whether Programmes A and B differ in their effectiveness?
 - Suppose you wanted to estimate the difference in average benefit between programmes A and C for a 27 year old participant. Give your answer in terms of $\hat{\beta}$ values.
 - Is it safe to assume that age is independent of the other explanatory variables? Answer Yes or No and briefly explain.
8. Pigs are routinely given large doses of antibiotics even when they show no signs of illness, to protect their health under unsanitary conditions. Pigs were randomly assigned to one of three antibiotic drugs. Dressed weight (weight of the pig after slaughter and removal of head, intestines and skin) was the response variable. The explanatory variables are Drug type, Mother's live adult weight and Father's live adult weight.

Data are in the file [pigweight.data.txt](#). You can get a copy with

```
oink = read.table("http://www.utstat.toronto.edu/~brunner/data/legal/pigweight.data.txt").
```

- Write the regression equation for the full model, including ϵ_i .
- Make a table with one row for every drug, and with columns showing how the dummy variables were defined. Make another column giving $E(y|\mathbf{x})$ for each drug.
- Fit the model (meaning estimate the parameters), and display the results of `summary`. What proportion of variation in the response variables is explained by the explanatory variables?
- Predict the dressed weight of a pig getting Drug 2, whose mother weighed 140 pounds, and whose father weighed 185 pounds. Your answer is a single number.

- (e) This parallel planes regression model specifies that the differences in expected weight for the different drug treatments are the same for every possible combination of mother's weight and father's weight. Give a 95% confidence interval for the difference in expected weight between drug treatments 2 and 3. The final answer is a pair of numbers, a lower confidence limit and an upper confidence limit. There is an easy way and a less easy way. The lecture slides illustrate both ways.
- (f) In symbols, give the null hypotheses you would test to answer the following questions. Your answers are statements involving the β values from your regression equation.
- i. Allowing for mother's weight and father's weight, does type of drug have an effect on the expected weight of a pig?
 - ii. Controlling for mother's weight and father's weight, which drug helps the average pig gain more weight, Drug 1 or Drug 2?
 - iii. Adjusting for mother's weight and father's weight, which drug helps the average pig gain more weight, Drug 1 or Drug 3?
 - iv. Correcting for mother's weight and father's weight, which drug helps the average pig gain more weight, Drug 2 or Drug 3?
- (g) For each of the questions below, give the value of the t or F statistic (a number from your printout), and indicate whether or not you reject the null hypothesis. The numbers may or may not be part of the default output from `summary`.
- i. Controlling for mother's weight and father's weight, does type of drug have an effect on the expected weight of a pig?
 - ii. Adjusting for mother's weight and father's weight, which drug helps the average pig gain more weight, Drug 1 or Drug 2?
 - iii. Allowing for mother's weight and father's weight, which drug helps the average pig gain more weight, Drug 1 or Drug 3?
 - iv. Controlling for mother's weight and father's weight, which drug helps the average pig gain more weight, Drug 2 or Drug 3?
 - v. Allowing for which drug they were given, does expected weight of a pig increase faster as a function of the mother's weight, or does it increase faster as a function of the father's weight?
- (h) We can assume that farmers want their pigs to weigh a lot. In plain, non-statistical language, can you offer some advice to a farmer based on these data? Remember, the farmer must be able to understand your answer or it is worthless.

Please bring your printout to the quiz. **Your printout should show *all* R input and output, and *only* R input and output.** Do not write anything on your printouts except your name and student number.