Math 5364 Homework 28

1. The data set `cows.txt` contains milk production values for 300 (hypothetical) cows, 100 from the Andrews Farm, 100 from the Bailey farm, and 100 from the Carter farm.

   (a) Import the data into SAS, and find the average milk production, stratified by farm. Also, obtain a histogram and qqplot of the milk production values at each farm.

   (b) Perform an ANOVA to test whether the average milk production at the three farms is the same.

   (c) Test whether the average milk production is the same using `PROC GLM`.

2. Let $U(a, b)$ denote a uniform distribution on the interval $[a, b]$ and $N(\mu, \sigma^2)$ denote a normal distribution with mean μ and variance σ^2. Let $X_{i1} \sim U(0, 100)$, $X_{i2} \sim U(30, 70)$, and $\epsilon_i \sim N(0, 1)$, for $i = 1, \ldots, 1000$. Also, suppose X_{i3} is a categorical variable taking the values "A", "B", and "C" with probabilities 0.5, 0.35, and 0.15, respectively. Finally, assume that all of the random variables X_{ij} and ϵ_i are statistically independent, and define

$$Y_i = 150 + 8X_{i1} + 6X_{i2} + 0.25X_{i2}^2 - 7X_{i1}X_{i2} + 5I(X_{i3} = \text{"B"}) + 10I(X_{i3} = \text{"C"}) + \epsilon_i.$$

   Recall that I is the indicator function, e.g., $I(X_{i3} = \text{"B"}) = 1$ if $X_{i3} = \text{"B"}$, and $I(X_{i3} = \text{"B"}) = 0$ otherwise.

   (a) Use SAS to simulate values of all random variables described above.

   (b) Verify that X_{i1}, X_{i2}, and ϵ_i have the distributions given above by plotting histograms for these variables.

   (c) Verify that the observed frequencies of the different levels of X_{i3} are approximately equal to those stated in the problem.

   (d) Fit the given regression equation to your simulated data, and verify that the estimated coefficients agree with those stated in the problem.