# Economics 217

## Homework #2

**Due Tuesday, February 5th**

### Problem 1

The normal distribution has the distribution function

$$f\left(y_i;\mu_i\right)=\frac{1}{\left(2\pi\sigma^2\right)^{1/2}}\exp\left[-\frac{1}{2\sigma^2}\left(y_i-\mu_i\right)^2\right]$$

Using this function, we will do some very simple examples of maximum likelihood.

**a.** Suppose that we have a sample of individuals, indexed from $i=1...n$, and the outcome variable for each individual is written as $y_i$. Write down and simplify the log-likelihood function for the normal distribution. (5 points).

**b.** Suppose that we restrict the mean of the normal distribution to be the same for all individuals $i$ ($\mu_i=\mu$). Please manipulate the log-likelihood function to solve for $\widehat{\mu}$, i.e., the maximum likelihood estimate for $\mu$. (10 points).

### Problem 2

For this question we will use the loess function in R to study monthly fluctuations in real wages.

$$f=\left(\exp(x)-2\right)^3$$

Obviously we can find this by hand, but your job is to find this using Newton-Raphon. For this question, you may find the following "snippet" of code helpful.

```
NR(x)<-function(x){
counter<-i
while(some condition is met){
Do something with x
counter<-counter+1
}
}
```

Basically, you're writing a function to execute Newton-Raphson, which is an iterative procedure while some condition is met.

**a.** Starting from an initial value of zero, please find the zero of $f=\left(\exp(x)-2\right)^3$ using Newton-Raphson.

**b.** Suppose that the function is now $f = (\exp(x) - 2)^n$. Please run your same Newton-Raphson procedure, starting from an initial value of zero, for $n = 2...8$. Please report how the number of iterations required to find the solution change with $n$.

### Problem 3

In this question, we will study wage outcomes and demographics. To do so, we would like to estimate the impact of education, gender, ethnicity, and age on the real wage. Please use an empirical model that (1) constrains predicted wages to be positive, and (2) restricts the ORG data to California residents in the year 2013.

**a.** The "wage gap" is typically referred to as the difference in wages between two groups, holding other characteristics fixed. Please estimate the male-female wage gap and comment on its significance. If possible, please quantify any difference in wages.

**b.** Does ethnicity play a role in wage outcomes, holding other variables fixed? Please test this hypothesis at the 95% level.

**c.** Please generate predictions for the dependent variable (on the scale of the dependent variable), and use a density plot to compare these predictions to the actual data on real wages. Please label your plot clearly.

### Problem 4

For this question we will use the loess and gam functions in R to study the relationship between real average hourly wages and hours worked (using the Org dataset).

**a.** Using loess, please estimate the relationship between the log real wage, $rw$ and hours worked. On a second plot, please use log of hours worked. Please interpret your figures. (10 points)

**b.** Using gam, please estimate the relationship between the log real wage and log hours worked controlling for education and age. Please provide confidence intervals for your estimates on the Figure, and interpret your results. (10 points)

**c.** Back to using loess, repeat the same exercise as in (2a), but please write a cross-validation procedure to find the optimal degree of smoothing (span, in the function). You may not use a "canned" package from the R library to run the cross-validation. Please plot your optimal figure, as well as provide results as to why you chose the degree of smoothing that you did. (10 points)