

1. COMPUTATIONAL ASSIGNMENT 4

1.1. **Question 1.** You want to find out how long it takes the processor in your computer to execute a `for()` statement in JAVA. The task is complicated by the fact that, despite documentation that says the function `threadData.getCurrentThreadCpuTime()` returns CPU time used by the current thread in nanoseconds, when you look at what is actually returned the resolution appears to be only milliseconds (i.e., the last six digits are always zeroes), for example

```
CPU time: 450000000
```

In addition, there are one-time overhead costs involved in loading, initializing, and (possibly) interpreting your program. These are almost certainly far more costly than executing a single `for` loop.

You decide on a strategy to execute the `for` loop many times, varying the number of executions, and record the reported CPU time.

The JAVA program looks like this: (you don't have to run it yourself, the data is provided below)

```
import java.lang.management.*;
public class timing{
    public static void main( String args[] )
    {
ThreadMXBean threadData = ManagementFactory.getThreadMXBean( );
        long n = 5000;
        for ( long j = 1; j < n*1000000; j++);
        System.out.println( "CPU time: " + n + " " +
            threadData.getCurrentThreadCpuTime( ));
    }
}
```

The results of running it with varying values of n are as follows: (the six zeroes at the end of each time value are not shown; the data can be interpreted as `for` loops in units of millions and CPU time in milliseconds)

n	time/1000000
25	320
40	450
100	1010
150	1470
250	2390
400	3770
600	5630
1000	9320
3000	27780
5000	46220

a) Use a simple linear regression to fit a model of the form

$$y = \beta_1 + \beta_2 x + e$$

the the data given. What are the estimates of the coefficients, $\hat{\beta}_1$ and $\hat{\beta}_2$?

R hint: You can manually enter short lists of data and create a data frame with code like the following abbreviated example:

```
> x=c(25,40,100,150)
> y=c(320,450,1010,1470)
> df<-data.frame(x,y)
> str(df)
```

Once you have the data frame, use `attach` to make the column names visible, then run

```
summary(lm(y ~ x))
```

b) How would you interpret the estimated value of β_2 , the slope of the regression line?

c) How would you interpret the estimated value of β_1 , the intercept?

1.2. **Question 2.** You are in charge of forecasting growth in sales for a small company. You have eleven years of data to go by,

year	sales (thousands)
2000	1210
2001	1276
2002	1295
2003	1489
2004	1592
2005	1692
2006	1821
2007	1964
2008	2091
2009	2218
2010	2370

The owner feels that the business has been growing at a steady rate, meaning the percentage increase in business is the same year over year. Your job is to estimate the (constant) annual percentage increase using the eleven data points.

hint: A constant percentage growth leads to an exponential growth model:

$$y(t) = \theta_1 e^{\theta_2 t}$$

where θ_2 represents the growth factor each year. This equation is not linear, but

$$\ln y = \ln(\theta_1) + \ln(\theta_2) \cdot t$$

is a linear equation.

a) Use this fact to set up and run a simple regression model of the log of sales against time. (you can adjust the time scale any way you want, say, to start at zero).

b) What do the parameter estimates from the regression represent?

c) Based on this model, what is the estimated annual growth rate, expressed as a percentage?

1.3. **Question 3.** Use the 2009 EPA mileage data to compare highway mileage for cars and trucks using a one-way ANOVA (the column heading for the classification variable is `car.truck`). Once you have loaded the EPA data as in the foils, you can obtain a data frame with only highway data using the statements:

```
epahighway<- subset(epa, C.H=="H")
attach(epahighway)
```

If you use the `lm()` function rather than `aov` the parameter estimates will display.

a) What is the estimated difference in highway gas mileage between cars and trucks?

b) Is the difference significant at the $\alpha = .05$ level?

1.4. **Question 4.** A truck dealer feels that it is unfair to single out trucks as getting worse gas mileage than cars because on average trucks weigh more than cars and if you compared a truck to a car of the same weight, you would not see any difference.

a) Test the dealer's assumption using the R statement

```
aggregate(epahighway$etw, by=list(car.truck), FUN=mean)
```

to display the average vehicle weight for cars and trucks in the data. What is the difference in the mean weights of cars and trucks in the data?

b) Rerun the analysis of problem 3) adding vehicle weight (`etw`) as a continuous predictor to adjust for the effect of the weight difference. What is the estimated difference between cars and trucks when weight is taken into consideration?

c) Is the difference between cars and trucks significant at the $\alpha = .05$ level?