Name:

Stats in practice #6

Regression

# Timothy and Jimothy

Brothers Timothy and Jimothy each run an ice cream stand in the city in which they live.

* Timothy lives in San Diego, California where the temperature is moderate, with a usual range of about 50 to 90 degrees.
* Jimothy lives in Phoenix, Arizona where the temperature can be more extreme, with a usual range of about 50 to 120 degrees.

The brothers are well aware that temperature has an effect on their sales (hotter day means more ice cream sold). They would like to develop regression models to predict sales based on temperature. There are two datasets on the course webpage which contain sales and temperature data for a random sample of 60 days over the last 5 years. *Note: This is not a real dataset, I made it up. (:*

# Timothy’s Ice Cream Stand

## Loading the data.

### Load the data into R and save the temperature and sales

### load(‘http://math.unm.edu/~knrumsey/Timothy\_sales.rda’)

### temp <- Timothy\_sales$temp

### sales <- Timothy\_sales$sales

## Hand calculations

### Consider the simple linear regression equation:

### Which variables (temperature and sales) do *x* and *y* correspond to?

### Using R, calculate the mean and variance of *x.*

### Using R, calculate the mean and variance of *y.*

### Using R, calculate the correlation of *x* and *y.*

### Using the equations given in class, calculate the slope and y-intercept of the least squares regression line.

## Regression using R

### This problem is much easier to solve using R. We can find the equation for the least squares regression line by typing:

### fit <- lm(sales~temp)

### Type “fit” into the console to see your results. Do the estimates for y-intercept and slope match what you got by hand? (It should)

### Make a scatterplot of temperature vs sales, and add the least squares regression line to the plot. Fill in the “?” with an appropriate value in the code below.

### plot(temp, sales, pch=?, main=’Timothy’)

### abline(fit, lty=3, col=?)

### Create a residual plot by typing:

### plot(temp, resid(fit), pch=?, main=’Residual Plot for Timothy’)

### abline(h=0, lty=3, col=?)

### Do the assumptions of linear regression appear to be met?

### Use your regression line to estimate Timothy’s sales when the temperature is 80 degrees.

### **Copy and paste your scatterplot (with regression line) from part ii. below. Also attach your residual plot.**

# Jimothy’s Ice Cream Stand

## Loading the data.

### Load the data into R and save the temperature and sales

### load(‘http://math.unm.edu/~knrumsey/Jimothy\_sales.rda’)

### temp <- Jimothy\_sales$temp

### sales <- Jimothy\_sales$sales

## Simple linear regression

### Repeat the analysis you did in section 2C (parts i. through iii.) for Jimothy’s data. Explain why simple linear regression may not be appropriate for this dataset.

### *Clarification: This next part is somewhat “beyond the scope of this class”. You will never need to know this for a test, but it’s useful to be aware that we can do this (easily) using R.*

### Suppose now we want to find the least squares polynomial equation:

### The equations for *a*, *b, c* and *d* are now more complicated, but R can handle this easily!

### fit\_poly <- lm(sales~temp+I(temp^2)+I(temp^3))

### Create a scatterplot of temperature vs sales for Jimothy by typing:

### plot(temp, resid(fit), pch=?, main=’Residual Plot for Timothy’)

### Add the least squares polynomial by typing:

### abcd <- coefficients(fit\_poly)

### curve(abcd[1]+abcd[2]\*x+abcd[3]\*x^2+abcd[4]\*x^3, add=TRUE, col=?)

### Create a new residual plot for Jimothy. Use the same code as in 2Cii. but change “fit” to “fit\_poly” and don’t forget to adjust the title of the plot.

### How does the polynomial regression equation compare to the simple linear regression equation? Which one gives a better fit? **Copy and paste the scatterplot (with regression polynomial) and the residual plot on the next page.**