STAT 420 Spring 2014

Homework 2: Due February 12 by 7:00pm

Exercise 1

DO NOT use a computer for this problem.

An employee claims that drinking beer has no effect on the amount of time it takes for him to perform a particular task. The following data show how many seconds he took to perform the task after consuming various quantities of beer, measured in ounces:

Beer consumption (x)	0	12	24	36	48	60	72
Task time (y)	62	50	59	74	59	83	68

Consider the simple linear regression model: $y_i = b_0 + b_1 x_i + e_i$ with $e_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$.

(a) Find the equation of the least-squares regression line.

(b) Calculate the fitted values $\hat{y}_i$.

(c) Calculate the residuals $\hat{e}_i$. Does the sum of the residuals equal zero?

(d) Give an estimate for σ, the standard deviation of the observations about the true regression line?

(e) What proportion of observed variation in time needed to perform the task is explained by a straight-line relationship with the amount of beer consumed?

(f) How much time would you expect the employee to need to perform the task after consuming 144 ounces of beer.

(g) The company statistician said that the prediction obtained in part (f) should be used "with extra caution". Why do you think he said this?

Exercise 2

DO use a computer for this problem. Use the data from Exercise 1.

Use a computer to find the equation of the least-squares regression line.

Create a scatterplot and add the least-squares regression line to it.

Exercise 3

Sometimes it is known in advance that the least-squares regression line must go through the origin, i.e., the regression model is of the form: $y_i = bx_i + e_i$ with $e_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$.

In this case, finding the least-squares lines involves findings the value $\hat{b}$ that minimizes

$$f(b) = \sum_{i=1}^{n} (y_i - bx_i)^2$$

Use the derivative of f with respect to b to derive the formula for the slope of the least-squares regression line in this case.

Exercise 4

It has been proposed that the brightness (measured in some unit of color) for a commercial product is proportional to the time it is in a certain chemical reaction during the production process, or

$y_i = bx_i + e_i \quad \text{with } e_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$

where y_i measures brightness, x_i measure time, and b is a parameter. The following data on x and y are available:

x	1.0	1.2	1.4	1.6	1.8	2.0
y	3.4	7.0	6.0	9.0	8.0	11.0

(a) Find the least-squares estimate $\hat{b}$. Hint: use answer from Exercise 3.

(b) Calculate the fitted values $\hat{y}_i$.

(c) Calculate the residuals $\hat{e}_i$. Does the sum of the residuals equal zero?

(d) Use a computer to find $\hat{b}$. To fit a model without the intercept, use

```
lm(y ~ 0 + x)
```

(e) Create a scatterplot and add the least-squares regression line $\hat{y} = \hat{b}x$ to it.