Stat 5102 (Geyer) Spring 2012
Homework Assignment 10
Due Wednesday, April 18, 2012

Solve each problem. Explain your reasoning. No credit for answers with no explanation. If the problem is a proof, then you need words as well as formulas. Explain why your formulas follow one from another.

10-1. The URL

`http://www.stat.umn.edu/geyer/5102/data/prob10-1.txt`

contains six variables named `x1`, `x2`, `x3`, `x4`, `x5`, and `y`.

(a) Perform a simple linear regression of `y` on `x1`.

(b) Perform a hypothesis test of whether the regression coefficient for `x1` is significantly different from zero.

(c) Perform a hypothesis test of whether the correlation coefficient of `x1` and `y` is significantly different from zero.

(d) Calculate a 95% confidence interval for the true population regression coefficient for `x1`.

10-2. The URL

`http://www.stat.umn.edu/geyer/5102/data/prob10-1.txt`

contains six variables named `x1`, `x2`, `x3`, `x4`, `x5`, and `y`.

(a) Perform a quadratic regression of `y` on `x1`, that is, the mean of `y` given `x1` is a quadratic function of `x1`.

(b) Perform a hypothesis test of whether the regression coefficient for `x1^2` is significantly different from zero.

(c) Calculate a 95% confidence interval for the true population regression coefficient for `x1^2`.

10-3. The URL

`http://www.stat.umn.edu/geyer/5102/data/prob10-1.txt`

contains six variables named `x1`, `x2`, `x3`, `x4`, `x5`, and `y`.

(a) Perform a linear regression of `y` on `x1` and `x2`, that is, the mean of `y` given `x1` and `x2` is a linear function of these two variables.

(b) Perform a hypothesis test of whether the regression coefficient for `x2` is significantly different from zero.

(c) Calculate a 95% confidence interval for the true population regression coefficient for `x2`.

10-4. The URL

`http://www.stat.umn.edu/geyer/5102/data/prob10-1.txt`

contains six variables named `x1`, `x2`, `x3`, `x4`, `x5`, and `y`.

(a) Perform a linear regression of `y` on `x1` and `x2`, that is, the mean of `y` given `x1` and `x2` is a linear function of these two variables.

(b) Perform a quadratic regression of `y` on `x1` and `x2`, that is, the mean of `y` given `x1` and `x2` is a general quadratic function of these two variables containing terms `x1^2`, `x1 * x2`, and `x2^2`.

(c) Perform a hypothesis test of whether the second model fits the data better than the first.

10-5. The URL

`http://www.stat.umn.edu/geyer/5102/data/prob10-5.txt`

contains two variables named `x` and `y`. This problem involves Fourier series. We will fit models of the form

$$E(Y) = \beta_1 + \sum_{i=1}^{k} \big(\beta_{2k} \sin(kx) + \beta_{2k+1} \cos(kx) \big)$$

for various values of k. We call the model for k the Fourier series with terms up to frequency k.

(a) Make a scatter plot of these two variables.

(b) Fit the Fourier series with terms up to frequency 1.

(c) Fit the Fourier series with terms up to frequency 2.

(d) Fit the Fourier series with terms up to frequency 3.

(e) Perform a hypothesis test of whether the second model fits the data better than the first.

(f) Perform a hypothesis test of whether the third model fits the data better than the second.

(g) Make a scatter plot of the data and add the regression functions for all three models to the plot.

10-6. The URL

`http://www.stat.umn.edu/geyer/5102/data/prob10-1.txt`

contains six variables named `x1`, `x2`, `x3`, `x4`, `x5`, and `y`.

(a) Perform a simple linear regression of `y` on `x1`.

(b) Perform a multiple linear regression of `y` on `x1`, `x2`, `x3`, `x4`, and `x5`.

(c) Calculate a 95% confidence interval for the mean of the first individual using only `x1`.

(d) Calculate a 95% confidence interval for the mean of the first individual using all five predictor variables.

(e) Same as (c) except prediction interval.

(f) Same as (d) except prediction interval.