STAT 420
2GR, 2UG

Spring 2017

Homework 3

(Due Friday, February 17, by 4:00 p.m.)

Please submit your assignment *on paper*, following the Guidelines for Homework Write-Ups and Submissions. Please include your name (with your last name underlined), and your NetID at the top of the first page. (Even if correct, answers might not receive credit if they are too difficult to read. No credit will be given without supporting work.)

1. **Do** use a computer (i.e. R / R Studio) for this problem, and include any relevant R code and output used to achieve the solution.

(a) Write your own function that computes the second moment of a discrete random variables. The input should be a matrix with two columns, with the possible values of the random variable in the first column, and the corresponding probabilities in the second column. The function should be called your first name initial plus your last name. For example, mine would be called `mwang`:

```
> mwang = function(X) { ••• }
```

(b) Use your function from part (a) on

(i) `X1 = cbind( c(2,4,6,8), c(0.1,0.2,0.4,0.3) )`

(ii) `X2 = cbind( c(0,3,6,9,12), c(0.20,0.10,0.30,0.15,0.25) )`

2. **Do** use a computer (i.e. R / R Studio) for this problem, and include any relevant R code and output used to achieve the solution.

Nancy claims that drinking beer has no effect on her running speed. The data on the right show how many seconds she took to run from her home to the grocery store where she usually goes for shopping, after consuming various amounts of beer, measured in ounces:

Beer Amount in ounces (x)	0	12	24	36	48	60	72
Running Time in seconds (y)	141	127	141	163	145	179	161

Consider the model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where ε_i's are i.i.d. $N(0, \sigma^2)$.

Please include a printout and mark (circle or highlight) the relevant answers.

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

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

(c) Calculate the residuals e_i. Does the sum of the residuals equal zero or not?

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

(e) Calculate the multiple R-squared and the adjusted R-squared.

(f) Use the t-test to perform the significance of the regression test at a 10% significance level. That is, test $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$ at a 10% significance level. Mark the value of the test statistic and the p-value.

(g) Use the F-test to perform the significance of the regression test at 5% significance level. That is, test $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$ at a 5% significance level. Mark the value of the test and the p-value.

(h) Construct a 95% confidence interval for β_1.

(i) Construct a 95% confidence interval for β_0.

(j) Construct a 90% confidence interval for the average time Nancy needs to run from her home to this grocery store after consuming 136 ounces of beer.

(k) Construct 90% prediction interval for the time Nancy needs to run from her home to this grocery store after consuming 136 ounces of beer.

(l) Create a scatterplot and add the least-square regression line to it.

3. You are given a random sample of 6 nights for a move revenue at Savoy 16 in Champaign, IL. The goal is to identify a relationship between the number of people who went to see this movie and the revenue earned in hundreds of dollars (including ticket sales, food and drinks).

Number of patrons, in hundreds (x)	0.7	0.5	0.9	1.1	1.5	1.3
Revenue earned, in hundreds of dollars (y)	7.0	4.8	12.0	14.0	18.4	14.9

Consider the model $Y_i = \beta x_i + \varepsilon_i$, where ε_i's are i.i.d. $N(0, \sigma^2)$.

[Note: For questions (a), (b), and (c), Do NOT use a computer but a calculator is allowed.]

(a) Find the least-squares estimate $\hat{\beta}$. You may round the result for $\hat{\beta}$ to one decimal place so that the computation in the following (b) and (c) based on the estimated $\hat{\beta}$ can be simpler or easier.

Hint: Use the answer or result in **Problem 1 of Homework 2**.

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

(c) Calculate the residuals e_i. Does the sum of the residuals equal zero or not?

(d) Use a computer (i.e. R / R Studio) to find $\hat{\beta}$. To fit a model without the *y*-intercept, use

```
> lm(y ~ x + 0)        or        > lm(y ~ x - 1)
```

Create a scatterplot and add the least-squares regression line $\hat{y} = \hat{\beta} x$ to it. (Include any relevant R code and output used to achieve the solution.)