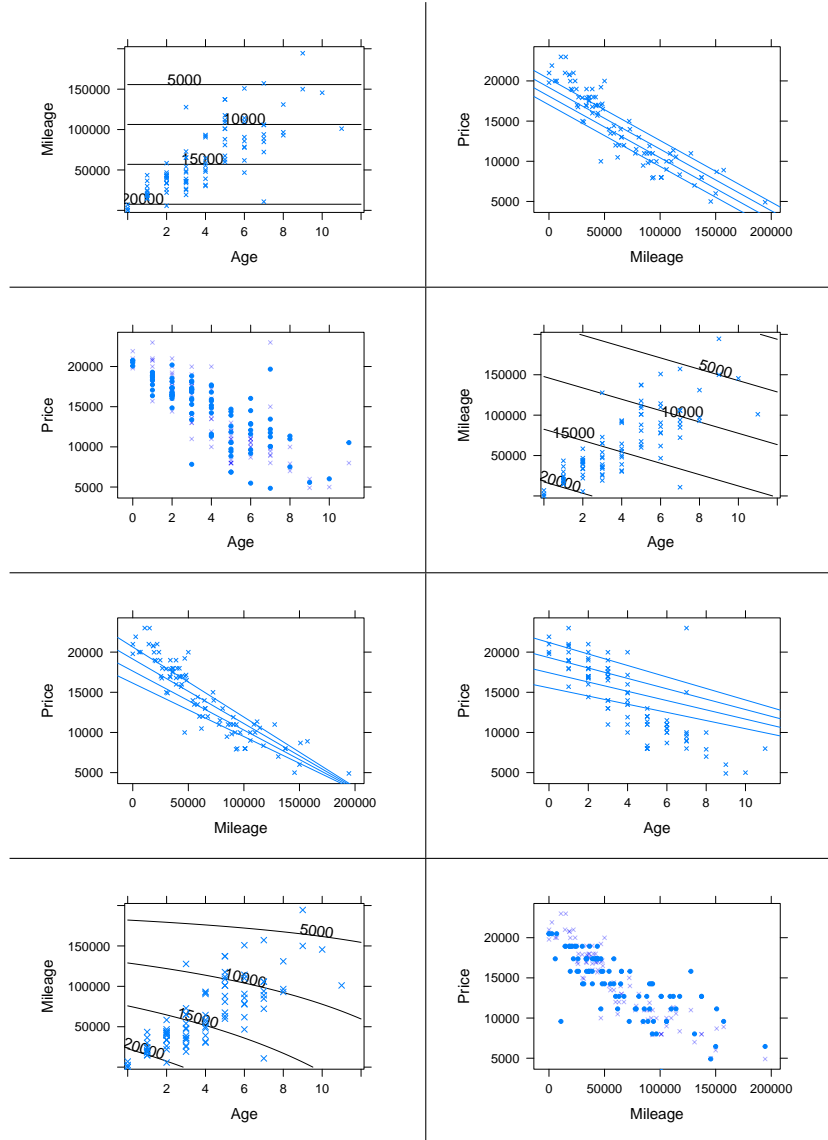


# Statistics Problems

Project MOSAIC Exercises

October 20, 2013

**Problem 1** Here are several graphs for used-car prices, mileage, and age. The actual cases are shown by x, the model values either as lines or solid dots.



Alongside each graph, write down the model design that corresponds to the model values. You can choose from among these models.

Price  $\sim$  Mileage      Price  $\sim$  Age      Price  $\sim$  Age+Mileage  
 Price  $\sim$  Age\*Mileage      Mileage  $\sim$  Price      Price  $\sim$  1

Hint: All the models you need are found among these six, but some of the six may be unrelated to the graphs.

**Problem Info:**

- Argument to

**Sexpr** `{knit_child(text='http://dtkaplan.github.io/MOSAIC-Exercises/Statistics/Rnw-Latex/car-model-graphs.Rnw')}`

**Problem 2** Data on 500 adults, including weight (kg), height (cm), waist size (cm), and sex. Here are the first five cases from a dataframe named `body`:

case	height	weight	waist	sex
1	178	72	87	M
2	166	91	109	F
3	174	80	95	F
4	180	79	93	M
5	168	55	70	F

... and so on.

A model was fitted to this data with the following model design:

```
mod1 = lm(weight ~ height * sex, data = body)
```

- Here are names of various model vectors from `mod1`. Beneath each name, write the first five values of the corresponding explanatory vector.

case	Intercept	height	sexM	height:sexM
1				
2				
3				
4				
5				

- The coefficients found from fitting the model were

Intercept	height	sexM	height:sexM
-1.6	0.37	3.6	0.20

Find:

- the model weight of the person in case 1 (Include units. Write a numerical formula, but it is not necessary to simplify the answer.)
- the model weight of the person in case 2 (Include units. Write a numerical formula, but it is not necessary to simplify the answer.)

3. The coefficient `height:sexM` is positive, not negative or zero. What does this tell you about the relation between height and weight for the people in the database?

**Problem Info:**

- Argument to  
`Sexpr` `{knit.child(text='http://dtkaplan.github.io/MOSAIC-Exercises/Statistics/Rnw-Latex/DF-F2013-001.Rnw')}`

**Problem 3** Several different models were made of wage (\$/hour) as a function of sex and marital status.

Here's one model's coefficient, fitted to the CPS85 data:

##	(Intercept)	sexM	marriedSingle
##	7.6838	3.1923	0.5759
##	sexM:marriedSingle		
##	-3.0972		

1. Fill in the model values for the wages for workers in each category:

Model Values		
	Married	Single
M		
F		

Model design: wage  $\sim$  sex \* married

2. Here's a table of model values from a possibly different model design fitted to other data. Write down the model design.

	Married	Single
M	10.25	9.12
F	10.25	9.12

Model design:

3. Here's another such table. Write down the model design.

	Married	Single
M	11.00	10.00
F	10.50	9.50

Model design:

### Problem Info:

- Argument to `Sexp` `{knit_child(text='http://dtkaplan.github.io/MOSAIC-Exercises/Statistics/Rnw-Latex/DF-F2013-002.Rnw')}`

**Problem 4** You are working for the Fall Foliage study group at Deciduous University. DecidU has assembled a large catalog of historical records of the duration and intensity of the fall color season. It looks like this:

Year	Intensity	Duration	Rainfall	Temperature
1962	5	15	3.6	50
1963	2	22	1.2	56
1964	1	12	5.8	38

... and so on up to the present.

Intensity is on a scale of 0 to 5 with 5 meaning “great beauty” and 0 meaning, “Blah, Let’s move to Florida.” Duration is the length of the season from 80% of peak leaf cover to 10%. Rainfall is in inches of rain in August and September. Temperature is the lowest early morning temperature in August.

Your job is to build statistical models that can help to answer researchers’ questions based on the data in the catalog.

Here are several models:

1. Intensity  $\sim$  Year
2. Intensity  $\sim$  Rainfall
3. Intensity  $\sim$  Rainfall \* Temperature
4. Intensity  $\sim$  Rainfall + Year + Temperature

These are some of the research questions posed by your colleagues. For each question, circle the model forms that are potentially useful. If there is are model forms that are useless in answering the research question question, draw an X through it.

- Does the Intensity of the foliage season depend on Rainfall?  
mod1 mod2 mod3 mod4
- Has the Intensity of the season lessened over the years (perhaps due to global warming)?  
mod1 mod2 mod3 mod4
- Has the Intensity of the season lessened systematically over the years, even adjusting for the effects of Temperature.  
mod1 mod2 mod3 mod4
- Has the Intensity peaked in the 1980s and is now declining?  
mod1 mod2 mod3 mod4
- Is it Rainfall or Temperature that has the largest effect on Intensity?  
mod1 mod2 mod3 mod4

- Does the effect of Rainfall on Intensity depend on the Temperature?  
mod1 mod2 mod3 mod4
- Does the effect of Temperature on Intensity depend on Rainfall?  
mod1 mod2 mod3 mod4

**Problem Info:**

- Argument to  
**Sexpr** `{kmit.child(text='http://dtkaplan.github.io/MOSAIC-Exercises/Statistics/Rnw-Latex/Folliage-1.Rnw')}`



**Problem 5** Here are several different statistics that can be calculated on a quantitative variable.

For each, indicate whether the statistic indicates something about variation in the variable, and whether or not the statistic is robust to outliers.

1. mean	variation: yes or no	robust: yes or no
2. median	variation: yes or no	robust: yes or no
3. IQR	variation: yes or no	robust: yes or no
4. standard deviation	variation: yes or no	robust: yes or no
5. variance	variation: yes or no	robust: yes or no

**Problem Info:**

- Argument to  
**Sexpr** `{knit_child(text='http://dtkaplan.github.io/MOSAIC-Exercises/Statistics/Rnw-Latex/variation-201.Rnw')}`