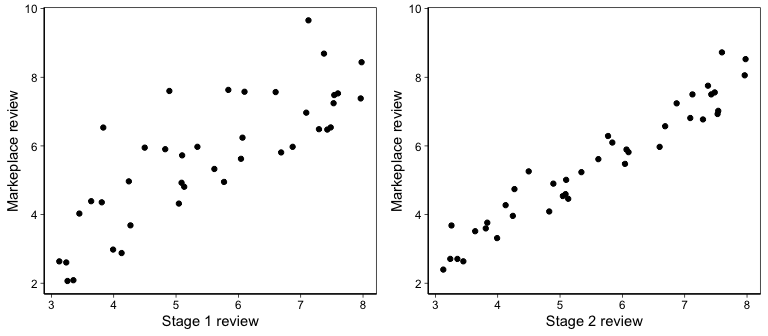
Problem set 6

## Section 8.1 problems

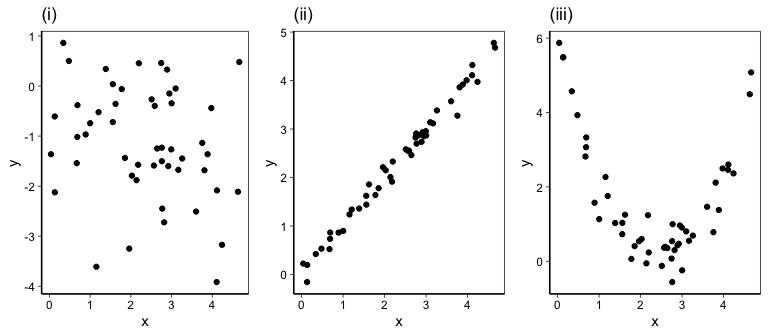
### Problem 1: Focus groups and product reviews.

The two scatterplots below show the relationship between Stage 1 product testing reviews, Stage 2 product testing reviews, and Markeplace reviews recorded during development for many objects sold by a major online retailer. Stage 1 typically occurs a year or more before the product goes to market, Stage 2 a few months before the product goes to market, and the Marketplace review is from the final product.



1. Based on these graphs, which of the two testing stage reviews has the strongest correlation with the markeplace review? Explain.
2. Can you think of a reason why the correlation between the stage you chose in part (a) and the marketplace review is higher?
3. Suppose you are interested in testing the hypothesis that the average reviews were consistent across the three time points. Would these data violate the independence assumption necessary to perform an ANOVA?

### Problem 2: By eye.



1. For each of the plots above, identify the strength of the relationship (e.g. weak, moderate, or strong) in the data and whether fitting a linear model would be reasonable.
2. For the plots in which you identified a linear model as being reasonable, suppose you were asked to estimate an intercept and slope for each. Which plot would you be more confident about making these estimates for? Put another way, which plot would you have less uncertainty about?

### Problem 3: Correlation.

What would the correlation be between the annual rainfall in Durham and annual rainfall in Beaufort if Durham always got

1. 20 inches more rain than Beaufort?
2. half as much rain as Beaufort?
3. 15% more rain than Beaufort?

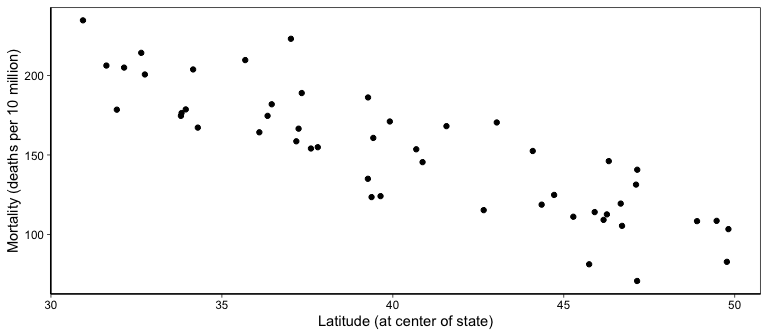
## Section 8.2 problems

### Problem 4: Over-under.

Suppose we fit a regression line to predict the price of a decorative rug based on its area. For a particular rug, we predict the price to be 128 US dollars. The rug’s residual is 35 US dollars. Did we over or under estimate the price of the rug? Explain your reasoning.

### Problem 5: Skin cancer mortality and latitude.

The following regression output is for predicting annual skin cancer mortality per ten million from state latitude.



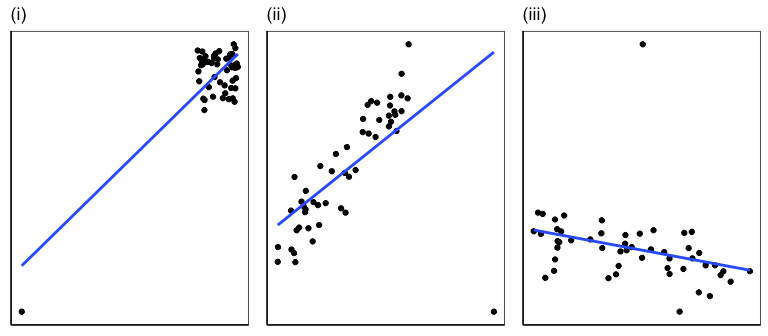
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 393.583 21.143 18.616 0  
## latitude -5.980 0.519 -11.525 0

We also know 39.33, 73.5%, and 72.9%.

1. Write out the linear model.
2. Interpret the intercept.
3. Interpret the slope.
4. Interpret .
5. Calculate the correlation coefficient.

## Section 8.3 problems

### Problem 6: Outliers



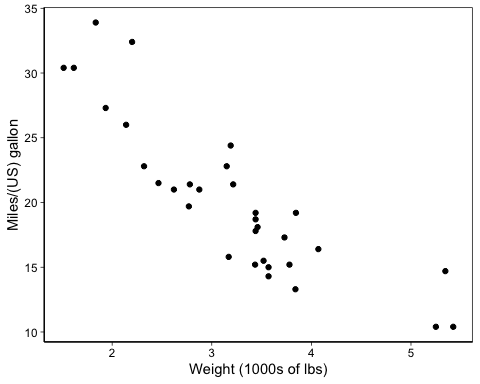
1. Identify the outliers in the scatterplots shown below, and determine what type of outliers they are. Explain your reasoning.
2. Order the three plots from smallest to largest apparent leverage of each of the outliers shown.

## Section 8.4 problems

In the following exercises, visually check the conditions for fitting a least squares regression line. However, you do not need to report these conditions in your solutions.

### Problem 7: Car weight and fuel efficiency

The scatterplot below summarizes the relationship between weight (in 1000s of pounds) and fuel efficiency (in Miles/(US) gallon, i.e. mpg) for 32 automobiles for which information was provided in 1974 Motor Trend US magazine. Summary output of the least squares fit for predicting mpg from weight is also provided in the table.



## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 37.285 1.878 19.858 0  
## mtcars$wt -5.344 0.559 -9.559 0

1. Is there strong evidence that heavier cars are less fuel efficient? State the hypotheses and include any information used to conduct the test.
2. Write the equation of the regression line for predicting mpg from car weight.
3. Interpret the slope and intercept in the context of the application.
4. Given that 0.75, what is the correlation between weight and mpg in this data set?
5. You see an ad for a 1970s car that weighs 4000 lbs. What would you predict its mpg to be? How reliable is this prediction?
6. You see an ad for a different 1970s car that weighs 6000 lbs. Would it be wise to use the same linear model to predict its mpg? Why or why not?

### Problem 8:

Is the amount spent on advertising in a given year predictive of how much a store will earn in sales? A local department store tracks its financial expenditures and profits for 24 years. The management board calculated the regression of net sales against advertising spending (both in thousands of US dollars). The estimated regression line is

1. What is the predicted net sales for a year in which five thousand dollars is spent on advertising?
2. The standard error for the coefficient of advertising is 5, which is associated with df = 22. Does the model provide strong evidence that advertising is significantly associated with sales?