

# HOMWORK №14

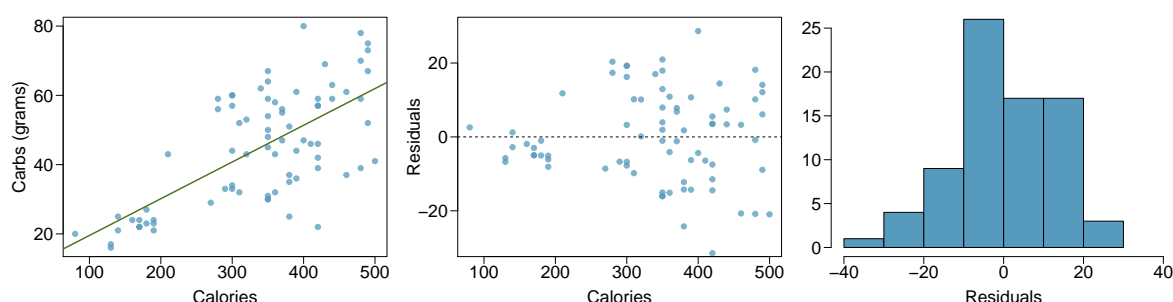
Math 107, Spring 2016

Due: June 3 by 4:00 pm

This is an optional homework assignment, so if you do not submit the assignment it will not adversely impact your grade. If you choose to submit the assignment, it will replace your lowest homework grade, regardless of whether your grade on this assignment is higher. When calculating your final homework grade, I will still drop the lowest two scores.

## Problem 1

The scatterplot below shows the relationship between the number of calories and amount of carbohydrates (in grams) Starbucks food menu items contain. Since Starbucks only lists the number of calories on the display items, we are interested in predicting the amount of carbs a menu item has based on its calorie content.



- Describe the relationship between number of calories and amount of carbohydrates (in grams) that Starbucks food menu items contain.
- In this scenario, what are the explanatory and response variables?
- Why might we want to fit a regression line to these data?
- Do these data meet the conditions required for fitting a least squares line?

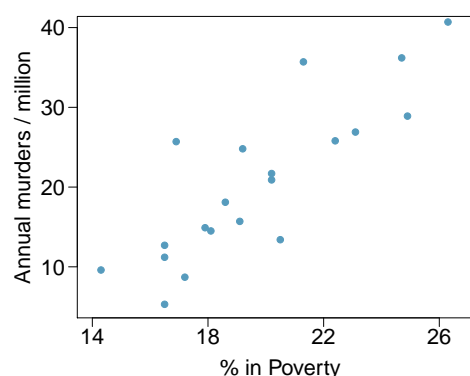
## Problem 2

The following regression output is for predicting annual murders per million from percentage living in poverty in a random sample of 20 metropolitan areas.

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-29.901	7.789	-3.839	0.001
poverty%	2.559	0.390	6.562	0.000

$s = 5.512$        $R^2 = 70.52\%$        $R^2_{adj} = 68.89\%$

- Write out the linear model.
- Interpret the intercept.
- Interpret the slope.
- Interpret  $R^2$ .
- Is the poverty percentage a significant predictor of the murder rate? You should be sure you address all five steps in your hypothesis test.
- Calculate a 95% confidence interval for the slope of poverty percentage, and interpret it in context of the problem.



## Problem 3

The movie Moneyball focuses on the “quest for the secret of success in baseball.” It follows a low-budget team, the Oakland Athletics, in the early 2000s, who believed that underused statistics, such as a player’s ability to get on base, actually better predicts the ability to score runs than typical statistics like homeruns, RBIs (runs batted in), and batting average. In fact, obtaining players who excelled in these underused statistics turned out to be much more affordable for the team. In this problem we’ll be looking at data from all 30 Major League Baseball teams and examining the linear relationship between runs scored in a season and a number of other player statistics. Our aim will be to summarize these relationships both graphically and numerically in order to find which variable, if any, best helps us predict a team’s runs scored in a season.

- (a) Create a scatterplot and describe the apparent relationship between runs and at\_bats. Be sure to include this plot in your submission.
- (b) Fit a simple linear regression model and report the least squares equation. This can be done in R using the command:

```
model <- lm(runs ~ at_bats, data = mlb)
summary(model)
```

Include this summary in your submission.

- (c) What is the value of  $R^2$ ? Provide an interpretation of  $R^2$  in the context of the problem.
- (d) Calculate a 95% confidence interval for the slope,  $\beta_1$ , and provide an interpretation of the confidence interval in the context of the problem.
- (e) In 2012, the Baltimore Orioles had 5,560 at bats. Predict the number of runs for the Orioles this year.
- (f)  $R^2$  can be used to select the “best model”, assuming that all the necessary assumptions are satisfied for each model. Below is a table of the  $R^2$  values for the 9 other simple linear regression models. We will assume that the necessary assumptions for inference are upheld in all models. Of all ten variables (the 9 below and at\_bats), which seems to be the best predictor of runs? Justify your answer. (Note: the three newer variables are the statistics used by the author of Moneyball to predict a team’s success.)

Predictor	$R^2$
hits	0.6419
homeruns	0.6266
bat_avg	0.2223
strikeouts	0.0292
stolen_bases	0.0138
wins	0.4364
new_onbase	0.3799
new_slug	0.7445
new_obs	0.6572