**Regression**

**Reading:**

*Required:* <https://www.datacamp.com/community/tutorials/linear-regression-R>

*Optional:* Chapter 9 in Statistical Methods: <https://pubs.usgs.gov/tm/04/a03/tm4a3.pdf>

Chapter 23 in R for Data Science: <https://r4ds.had.co.nz/model-basics.html>

**Exercises:**

This week we will recreate the exercise from the primary reading. We’ll start this in class. Your job now is to create a summary version of the reading that you’ll turn into your 1-pager for HW that describes steps for linear regression.

**Tutorials:**

There is a tutorial files available in our Posit that follows along the required reading and some other examples - <https://posit.cloud/spaces/115479/content/5575956>

**Homework:**

The homework this week will consist of 2 linear regression examples and a cheatsheet (as described above).

## Part 1.

Create a short 1-page cheatsheet for linear regression.

## Part 2.

In Biotechnology, we think alot about kinetics and reactors. The concentration of bacteria in a batch reactor is monitored after adding an inhibitor.

time [hours] <- 4 8 12 16 20 24

cbacteria [CFU] <- 1600 1320 1000 890 650 560

Here CFU is in units of colony forming units. The form of the equation describing how the concentration of bacteria in a solution of an inhibitor decreases with time is

Where *C(t)* is concentration of bacteria at time t, *C0* is t the initial concentration of bacteria, *k* is the rate constant of the inhibitor, and *t* is the time.

* Linearize this equation and fit a linear model to the data.
  + Your choice of model must not include negative concentrations (impossible) and concentration always decreases with time.
* Describe how the resulting model meets the tests [(see tutorial video](https://drive.google.com/open?id=1nzF-cY5uFtCs2Ee1HsaWCYfiTGk38dwk&authuser=wrightrc%40vt.edu&usp=drive_fs), and this week's lecture recordings).
* Estimate the concentration at the beginning (time = 0) and the time when cbacteria reaches 200 CFU.
* Explicitly include a solution summary below the problem statement on your R-markdown, where the remainder of your markdown file backs up your work. For example…

“The above data was transformed to a linear form using an xxxxx transformation. The resulting linear model had an adjusted R2 of 0.xxx, and both the intercept and slope were significant (|t|>>2). Leverage was not detected for any observations based on both Cook's D and DFFITS metrics. Residual analysis showed that the variance of the residuals was largely consistent, no bias was detected, and the model residuals followed the normality assumption.

The resulting model formulation in untransformed space takes the following form:

(where you replace the constants *C0* and *k* with their values from your model)

The concentration at the beginning of the experiment is 3434 CFUs (range 2000-4000 using an alpha of 95%). To reach a concentration of 200CFUs, the predicted time ranges from cc - bb hours, with a best-guess estimate of yy hours.”

## Part 3

In Watershed Science, we often need to design the size of BMPs and reservoirs on incoming streamflow. When long-term streamflow is not available, one might use long-term precipitation to estimate streamflow. Suppose we have the following data:

Precip [cm/yr] <- 88.9 108.5 104.1 139.7 127 94 116.8 99.1

Streamflow [m^3/s] <- 14.6 16.7 15.3 23.2 19.5 16.1 18.1 16.6

Fit a linear regression to the data using the workflow approach, and describe how the resulting model meets the tests. Estimate streamflow for a precipitation value of 120 cm, including uncertainty. Explicitly include a solution summary below the problem statement on your R-markdown, where the remainder of your markdown file backs up your work.