

STAT GU4223/GR5223 Spring 2019

HW is due in class **at the beginning** of the lecture on the due date. Late HW is not accepted. See syllabus if you want to submit HW early. All submitted HW (except codes/plots) needs to be hand-written unless you have a documented disability, verified through disability services, that prevents this; or unless otherwise stated in the HW.

Make sure you make a copy of your HW before submitting so that you can compare it with the solutions once the solutions are posted. Show all work to get full credit.

TA Office Hours & Location: See Courseworks announcement.

Statistics Dept. MA Help Room Information: <http://stat.columbia.edu/ma-programs/m-a-help-room/>

You can find the codes and datasets that you will need for most of the HW problems on one of the resource pages for the book:

<https://www.karlin.mff.cuni.cz/~hlavka/sms2/index.html>

<https://github.com/quantlet/MVA>

<https://github.com/QuantLet/SMS2>

If you use code/reference from these or other sites/resources make sure to cite them appropriately in your HW separately for each problem.

Make also sure to include a print-out copy of the code you are using in the HW separately for each problem.

If you are working on the HW together with your classmates, you can submit one HW for a group of up to 3 students as long as the HW is accompanied by contribution statements clearly stating each student's contribution to the HW, in this case the students submitting the HW as a group will receive the same HW grade. Keep in mind that for the exams and the course each student is responsible from the full assigned set of HW problems.

Solve the following exercises from the course textbook, together with their additional follow-up questions following the + signs below. Selected problems will be graded.

HW1 (out: 02/17, Due: 02/26):

Note: HW 1 will be graded based on submission. You will get full credit for attempting more than 15 of the 20 problems below. Problem = indicated textbook exercise +additional follow-up questions listed after the + signs following the exercise. However, for the exams and the course you are responsible from the full assigned set of HW problems.

Ch1: #9 use Silverman's rule of thumb for bandwidth selection; #10 +re-create Figure 1.9; #11 +interpret the plot +highlight possible disadvantages of PCP; #12 +identify the discrete variables in the car data using the PCP you have drawn in #11; #17 +give five-number summaries of variables population and unemployment +in 17.1 draw for both variables population and unemployment +in 17.2 choose 20 data points +highlight advantages/ disadvantages/usefulness of using these four types of plots for this dataset; #18.

Ch 2: #2; #3; #4 you can use Quantlet SMSjordandec +check (2.2) and (2.3) on pg.58 of the textbook numerically +compute A^{-1} and A^2 using Jordan decomposition of A ; #5;

Ch3: #1; #2; #4 +is the covariance you computed an appropriate measure to judge the strength of linear association between the variables weight and mileage? would computing the correlation instead suffice to judge the strength of linear association between the variables weight and mileage?; #5; #8; #10; #15; #16 +also compute empirical covariance between variables X_2 and X_3 in new currency using empirical covariance between X_2 and X_3 in old currency and highlight the difference between the two computations (i.e going from old currency to new currency for X_1 and X_2 vs X_2 and X_3) + how would the computations of empirical correlations be affected under currency change?; #18 hint: show that H is idempotent and use result from Exercise 2.6; #19 take matrix $D^{-1/2}$ as defined on pg. 91 +what is the effect of multiplication from the left by the centering matrix H ? +what is the effect of multiplication from the right by the matrix $D^{-1/2}$.