Math 583 HW 1 2023 Due Wednesday, Aug. 30.

DO NOT place solutions side by side.

YOU ARE BEING GRADED FOR WORK NOT JUST THE FINAL ANSWER. As a rule of thumb, you should have some idea of what you were doing, even without the book. You are encouraged to form groups to discuss ideas and HW problems, but do not copy. Two pages, Problems A)-E).

A) Let $\boldsymbol{x}$ be a $p \times 1$ random vector with covariance matrix $\mathrm{Cov}(\boldsymbol{x})$. Let $\boldsymbol{A}$ be an $r \times p$ constant matrix and let $\boldsymbol{B}$ be a $q \times p$ constant matrix. Find $\mathrm{Cov}(\boldsymbol{Ax}, \boldsymbol{Bx})$ in terms of $\boldsymbol{A}, \boldsymbol{B}$ and $\mathrm{Cov}(\boldsymbol{x})$. (A proof is wanted.)

B) If X and Y are random variables, show that

$$\mathrm{Cov(X, Y)} = [\mathrm{Var(X + Y)} - \mathrm{Var(X - Y)}]/4.$$

C) Suppose that

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} \sim N_4 \left(\begin{pmatrix} 1 \\ 7 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 0 & 2 & 1 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & 3 & 1 \\ 1 & 0 & 1 & 5 \end{pmatrix} \right).$$

a) Find the distribution of X_3.

b) Find the distribution of $(X_1, X_4)^T$.

D) 2) Recall that if $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the conditional distribution of $\boldsymbol{X}_1$ given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is multivariate normal with mean $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$ and covariance $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

Let $\sigma_{12} = \mathrm{Cov(Y, X)}$ and suppose Y and X follow a bivariate normal distribution

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 134 \\ 96 \end{pmatrix}, \begin{pmatrix} 24.5 & 1.1 \\ 1.1 & 23.0 \end{pmatrix} \right).$$

a) Find $E(Y|X)$.

b) Find $\mathrm{Var}(Y|X)$.

Information about R: The Math lab is in Neckers 258. Login to one of these computers: the initial login can take 10 minutes. The computer on button is in the upper left corner while the monitor on button is in the lower right corner. You may need to press a computer keyboard key to get the login and password bars to appear.

The computer lab login is like logging into salukinet. If necessary, hit Ctrl, enter your AD\siu8... (dawg tag) and your password. Left click the lower left icon, or near the lower left icon, search for R, Word, etc. Left double click the `R` icon.

Hopefully, R is on most of the computers. R is free software that can be downloaded on your personal computer. See text (http://parker.ad.siu.edu/Olive/slch11.pdf).

R is available from the **CRAN** website (https://cran.r-project.org/). R is similar to *Splus*, but is free. R is very versatile since many people have contributed useful code, often as packages.

Downloading the book's files into R

Many of the homework problems use R functions contained in the book's website (http://parker.ad.siu.edu/Olive/slearnbk.htm) under the file name *slpack.txt*. The following two R commands can be copied and pasted into R from near the top of the file (http://parker.ad.siu.edu/Olive/hdrhw.txt). **Also use this file to copy and paste commands for problem E).**

Downloading the book's R functions *slpack.txt* and data files *sldata.txt* into R: the commands

```
source("http://parker.ad.siu.edu/Olive/slpack.txt")
source("http://parker.ad.siu.edu/Olive/sldata.txt")
```

can be used to download the R functions and data sets into R. Type *ls()*. Nearly 70 R functions from *slpack.txt* should appear. In R, enter the command *q()*. A window asking "*Save workspace image?*" will appear. Click on *No* to remove the functions from the computer (clicking on *Yes* saves the functions in R, but the functions and data are easily obtained with the source commands).

E 1.10. This problem uses some of the R commands at the end of Section 1.2.1. A problem with response and residual plots is that there can be a lot of black in the plot if the sample size n is large (more than a few thousand). A variant of the response plot for the additive error regression model $Y = m(\boldsymbol{x}) + e$ would plot the identity line, the two lines parallel to the identity line corresponding to the Olive (2017, Section 2.1) large sample $100(1-\delta)\%$ prediction intervals for Y_f that depends on $\hat{Y}_f$. Then plot points corresponding to training data cases that do not lie in their $100(1-\delta)\%$ PI. We will use $\delta = 0.01$, $n = 100000$, and $p = 8$.

a) Copy and paste the commands for this part into R. They make the usual response plot with a lot of black. Do not include the plot in *Word*.

b) Copy and paste the commands for this part into R. They make the response plot with the points within the pointwise 99% prediction interval bands omitted. Include this plot in *Word*. For example, left click on the plot and hit the *Ctrl* and *c* keys at the same time to make a copy. Then paste the plot into *Word*, e.g., get into *Word* and hit the *Ctrl* and *v* keys at the same time.