Math 583 HW 3 Fall 2017. Due Wednesday, Sept. 13.
Quiz 3 on Friday, Sept. 15 is similar to HW 3. Use 3 sheets of notes.

Problem numbers are from the Olive text.

A) 1.3. Suppose $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid $p \times 1$ random vectors from a multivariate t-distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with d degrees of freedom. Then $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{x}) = \dfrac{d}{d-2}\boldsymbol{\Sigma}$ for $d > 2$. Assuming $d > 2$, find the limiting distribution of $\sqrt{n}(\overline{\boldsymbol{x}} - \boldsymbol{c})$ for appropriate vector $\boldsymbol{c}$.

B) 2.1. Consider the Cushny and Peebles data set (see Staudte and Sheather 1990, p. 97) listed below. Find shorth(7). Show work.

```
0.0  0.8  1.0  1.2  1.3  1.3  1.4  1.8  2.4  4.6
```

See example done in class and Olive (2020, pp. 91-91).

C) 2.6. To find the sample median of a list of n numbers where n is odd, order the numbers from smallest to largest and the median is the middle ordered number. The sample median estimates the population median. Suppose the sample is $\{14, 3, 5, 12, 20, 10, 9\}$. Find the sample median for each of the three bootstrap samples listed below.
Sample 1: 9, 10, 9, 12, 5, 14, 3
Sample 2: 3, 9, 20, 10, 9, 5, 14
Sample 3: 14, 12, 10, 20, 3, 3, 5

Copy and paste the two source commands from near the top of *slrhw.txt* for the following *R* problems.

D) 2.9. a) Type the *R* command `predsim()` and paste the output into *Word*.

This program computes $\boldsymbol{x}_i \sim N_4(\boldsymbol{0}, diag(1,2,3,4))$ for $i = 1, ..., 100$ and $\boldsymbol{x}_f = \boldsymbol{x}_{101}$. One hundred such data sets are made, and ncvr, scvr, and mcvr count the number of times $\boldsymbol{x}_f$ was in the nonparametric, semiparametric, and parametric MVN 90% prediction regions. The volumes of the prediction regions are computed and voln, vols, and volm are the average ratio of the volume of the ith prediction region over that of the semiparametric region. Hence vols is always equal to 1. For multivariate normal data, these ratios should converge to 1 as $n \to \infty$.

b) Were the three coverages near 90%?

E) 2.10. Consider the multiple linear regression model $Y_i = \beta_1 + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} + e_i$ where $\boldsymbol{\beta} = (1, 1, 0, 0)^T$. The function `regbootsim2` bootstraps the regression model, finds bootstrap confidence intervals for β_i and a bootstrap confidence region for $(\beta_3, \beta_4)^T$ corresponding to the test $H_0 : \beta_3 = \beta_4 = 0$ versus H_A: not H_0. See Olive (2017, p. 82). The lengths of the CIs along with the proportion of times the CI for β_i contained β_i are given. The fifth interval gives the length of the interval $[0, D_{(c)}]$ where H_0 is rejected if $D_0 > D_{(c)}$ and the fifth "coverage" is the proportion of times the test fails to reject H_0. Since nominal 95% CIs were used and the nominal level of the test is 0.05 when H_0 is true, we want the coverages near 0.95. The CI lengths for the first 4 intervals should be near 0.392. The residual bootstrap is used. (See next page.)

Copy and paste the commands for this problem into *R*, and include the output in *Word*.