Exam 1 review. 7 sheets of notes and a calculator. Wednesday Feb. 14.

Types of problems.

1) Given a small data set, find $\overline{Y}$, S, MED(n) and MAD(n). See HW1 A and quiz 1. Recall that $\overline{Y} = \frac{\sum_{i=1}^n Y_i}{n}$ and the *sample variance*

$$\text{VAR}(n) = S^2 = S_n^2 = \frac{\sum_{i=1}^n (Y_i - \overline{Y})^2}{n-1} = \frac{\sum_{i=1}^n Y_i^2 - n(\overline{Y})^2}{n-1},$$

and the *sample standard deviation* (SD) $S = S_n = \sqrt{S_n^2}$.

If the data $Y_1, ..., Y_n$ is arranged in ascending order from smallest to largest and written as $Y_{(1)} \leq \cdots \leq Y_{(n)}$, then the $Y_{(i)}$'s are called the *order statistics.* The *sample median*

$$\text{MED}(n) = Y_{((n+1)/2)} \quad \text{if n is odd,}$$

$$\text{MED}(n) = \frac{Y_{(n/2)} + Y_{((n/2)+1)}}{2} \quad \text{if n is even.}$$

The notation MED(n) = MED($Y_1, ..., Y_n$) will also be used. To find the sample median, sort the data from smallest to largest and find the middle value or values.

The *sample median absolute deviation*

$$\text{MAD}(n) = \text{MED}(|Y_i - \text{MED}(n)|,\ i = 1, \ldots, n).$$

To find MAD(n), find $D_i = |Y_i - \text{MED}(n)|$, then find the sample median of the D_i by ordering them from smallest to largest and finding the middle value or values.

2) Know that a **scatterplot** of x versus y is used to visualize the conditional distribution of $y|x$. A **scatterplot matrix** is an array of scatterplots. It is used to examine the bivariate relationships of the p random variables. See HW1 D, Q1.

3) Suppose that all values of the variable w to be transformed are positive. The **log rule** says use $\log(w)$ if $\max(w_i)/\min(w_i) > 10$. Know how to use this rule. See Q1, HW1 E.

4) There are several guidelines for **choosing power transformations**. First, suppose you have a scatterplot of two variables $x_1^{\lambda_1}$ versus $x_2^{\lambda_2}$ where both $x_1 > 0$ and $x_2 > 0$. Also assume that the plotted points follow a nonlinear one to one function. The **ladder rule:** consider the **ladder of powers**

$$-1, -0.5, -1/3, 0, 1/3, 0.5, \quad \text{and} \quad 1.$$

To spread small values of the variable, make λ_i smaller. To spread large values of the variable, make λ_i larger. Know how to use this rule. See Q1.

5) The *population mean* of a random $p \times 1$ vector $\boldsymbol{x} = (x_1, ..., x_p)^T$ is $E(\boldsymbol{x}) = \boldsymbol{\mu} = (E(x_1), ..., E(x_p))^T$ and the $p \times p$ *population covariance matrix*
$\text{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}_{\boldsymbol{x}} = E(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x} - E(\boldsymbol{x}))^T = ((\sigma_{i,j}))$ where $\text{Cov}(x_i, x_j) = \sigma_{i,j}$. The $p \times p$ *population correlation matrix* $\text{Cor}(\boldsymbol{x}) = \boldsymbol{\rho} = ((\rho_{ij}))$. The *population covariance matrix* of $\boldsymbol{x}$ with $\boldsymbol{y}$ is $\text{Cov}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\Sigma}_{\boldsymbol{x},\boldsymbol{y}} = E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{y} - E(\boldsymbol{y}))^T]$.

6) If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are $p \times 1$ random vectors, $\boldsymbol{a}$ a conformable constant vector, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable constant matrices, then

$$E(\boldsymbol{X}+\boldsymbol{Y}) = E(\boldsymbol{X}) + E(\boldsymbol{Y}), \; E(\boldsymbol{a}+\boldsymbol{Y}) = \boldsymbol{a} + E(\boldsymbol{Y}), \; \& \; E(\boldsymbol{AXB}) = \boldsymbol{A}E(\boldsymbol{X})\boldsymbol{B}.$$

Also

$$\text{Cov}(\boldsymbol{a}+\boldsymbol{AX}) = \text{Cov}(\boldsymbol{AX}) = \boldsymbol{A}\text{Cov}(\boldsymbol{X})\boldsymbol{A}^T.$$

Note that $E(\boldsymbol{AY}) = \boldsymbol{A}E(\boldsymbol{Y})$ and $\text{Cov}(\boldsymbol{AY}) = \boldsymbol{A}\text{Cov}(\boldsymbol{Y})\boldsymbol{A}^T$. See HW1 B.

7) The $n \times p$ data matrix

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,p} \\ x_{2,1} & x_{2,2} & \dots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_p \end{bmatrix}$$

8) The **sample mean** or *sample mean vector*

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i = (\overline{x}_1, ..., \overline{x}_p)^T.$$

9) The **sample covariance matrix**

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T = ((S_{ij})).$$

10) The **sample correlation matrix** $\boldsymbol{R} = ((r_{ij}))$.

11) The spectral decomposition $\boldsymbol{A} = \sum_{i=1}^p \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^T = \lambda_1 \boldsymbol{e}_1 \boldsymbol{e}_1^T + \cdots + \lambda_p \boldsymbol{e}_p \boldsymbol{e}_p^T$.

12) Let $\boldsymbol{A} = \sum_{i=1}^p \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^T$ be a positive definite $p \times p$ symmetric matrix. Let $\boldsymbol{P} = [\boldsymbol{e}_1 \; \boldsymbol{e}_2 \; \cdots \; \boldsymbol{e}_p]$ be the $p \times p$ orthogonal matrix with ith column $\boldsymbol{e}_i$. Let $\boldsymbol{\Lambda}^{1/2} = \text{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_p})$. The *square root matrix* $\boldsymbol{A}^{1/2} = \boldsymbol{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}^T$ is a positive definite symmetric matrix such that $\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} = \boldsymbol{A}$.

13) The population squared Mahalanobis distance $D^2_{\boldsymbol{x}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$.

14) The sample squared Mahalanobis distance $D^2_{\boldsymbol{x}}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = (\boldsymbol{x} - \hat{\boldsymbol{\mu}})^T \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{x} - \hat{\boldsymbol{\mu}})$.

15) The *generalized sample variance* $= |\boldsymbol{S}| = \det(\boldsymbol{S})$.

16) The hyperellipsoid $\{\boldsymbol{x} | D^2_{\boldsymbol{x}} \le h^2\} = \{\boldsymbol{x} : (\boldsymbol{x} - \overline{\boldsymbol{x}})^T \boldsymbol{S}^{-1} (\boldsymbol{x} - \overline{\boldsymbol{x}}) \le h^2\}$ is centered at $\overline{\boldsymbol{x}}$ and has volume is

$$\frac{2\pi^{p/2}}{p\Gamma(p/2)} |\boldsymbol{S}|^{1/2} h^p.$$

Let $\boldsymbol{S}$ have eigenvalue eigenvector pairs $(\hat{\lambda}_i, \hat{\boldsymbol{e}}_i)$ where $\hat{\lambda}_1 \ge \cdots \ge \hat{\lambda}_p$. If $\overline{\boldsymbol{x}} = \boldsymbol{0}$, the axes are given by the eigenvectors $\hat{\boldsymbol{e}}_i$ where the half length in the direction of $\hat{\boldsymbol{e}}_i$ is $h\sqrt{\hat{\lambda}_i}$. Here $\hat{\boldsymbol{e}}_i^T \hat{\boldsymbol{e}}_j = 0$ for $i \ne j$ while $\hat{\boldsymbol{e}}_i^T \hat{\boldsymbol{e}}_i = 1$.

17) If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$.

18) If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and if $\boldsymbol{A}$ is a $q \times p$ matrix, then $\boldsymbol{AX} \sim N_q(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T)$. If $\boldsymbol{a}$ is a $p \times 1$ vector of constants, then $\boldsymbol{X} + \boldsymbol{a} \sim N_p(\boldsymbol{\mu} + \boldsymbol{a}, \boldsymbol{\Sigma})$. See Q2, HW2 E.

$$\text{Let} \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \ \text{and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

19) **All subsets of a MVN are MVN:** $(X_{k_1}, ..., X_{k_q})^T \sim N_q(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ where $\tilde{\boldsymbol{\mu}}_i = E(X_{k_i})$ and $\tilde{\boldsymbol{\Sigma}}_{ij} = \text{Cov}(X_{k_i}, X_{k_j})$. In particular, $\boldsymbol{X}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\boldsymbol{X}_2 \sim N_{p-q}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent iff $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$. See Q2, HW2 B.

20)
$$\text{Let} \ \begin{pmatrix} Y \\ X \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \begin{pmatrix} \sigma_Y^2 & \text{Cov}(Y, X) \\ \text{Cov}(X, Y) & \sigma_X^2 \end{pmatrix} \right).$$

Also recall that the *population correlation* between X and Y is given by

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{VAR}(X)}\sqrt{\text{VAR}(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

if $\sigma_X > 0$ and $\sigma_Y > 0$.

21) The conditional distribution of a MVN is MVN. 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 matrix $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. That is,

$$\boldsymbol{X}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2 \sim N_q(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

22) Notation:

$$\boldsymbol{X}_1 | \boldsymbol{X}_2 \sim N_q(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

23) Be able to compute the above quantities if X_1 and X_2 are scalars. See Q2, HW2 C.

24) A $p \times 1$ random vector $\boldsymbol{X}$ has an *elliptically contoured distribution,* if $\boldsymbol{X}$ has density

$$f(\boldsymbol{z}) = k_p|\boldsymbol{\Sigma}|^{-1/2} g[(\boldsymbol{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{z} - \boldsymbol{\mu})], \tag{1}$$

and we say $\boldsymbol{X}$ has an elliptically contoured $EC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ distribution. If the second moments exist, then

$$E(\boldsymbol{X}) = \boldsymbol{\mu} \tag{2}$$

and

$$\text{Cov}(\boldsymbol{X}) = c_X \boldsymbol{\Sigma} \tag{3}$$

for some constant $c_X > 0$.

25) The *population squared Mahalanobis distance*

$$U \equiv D^2 = D^2(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\boldsymbol{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}). \tag{4}$$

For elliptically contoured distributions, U has pdf

$$h(u) = \frac{\pi^{p/2}}{\Gamma(p/2)} k_p u^{p/2-1} g(u). \tag{5}$$

$U \sim \chi^2_p$ if $\boldsymbol{x}$ has a multivariate normal $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution.

26) The classical estimator $(\overline{\boldsymbol{x}}, \boldsymbol{S})$ of multivariate location and dispersion is the sample mean and sample covariance matrix where

$$\overline{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \quad \text{and} \quad \boldsymbol{S} = \frac{1}{\mathrm{n}-1} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\boldsymbol{x}_{\mathrm{i}} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{\mathrm{i}} - \overline{\boldsymbol{x}})^{\mathrm{T}}. \tag{6}$$

27) Let the $p \times 1$ column vector $T(\boldsymbol{W})$ be a multivariate location estimator, and let the $p \times p$ symmetric positive definite matrix $\boldsymbol{C}(\boldsymbol{W})$ be a dispersion estimator. Then the ith *squared sample Mahalanobis distance* is the scalar

$$D_i^2 = D_i^2(T(\boldsymbol{W}), \boldsymbol{C}(\boldsymbol{W})) = (\boldsymbol{x}_i - T(\boldsymbol{W}))^T \boldsymbol{C}^{-1}(\boldsymbol{W})(\boldsymbol{x}_i - T(\boldsymbol{W})) \tag{7}$$

for each observation $\boldsymbol{x}_i$. Notice that the Euclidean distance of $\boldsymbol{x}_i$ from the estimate of center $T(\boldsymbol{W})$ is $D_i(T(\boldsymbol{W}), \boldsymbol{I}_p)$. The classical Mahalanobis distance uses $(T, \boldsymbol{C}) = (\overline{\boldsymbol{x}}, \boldsymbol{S})$.

28) If p random variables come from an elliptically contoured distribution, then the subplots in the scatterplot matrix should be linear.

29) Let $\boldsymbol{X}_n$ be a sequence of random vectors with joint cdfs $F_n(\boldsymbol{x})$ and let $\boldsymbol{X}$ be a random vector with joint cdf $F(\boldsymbol{x})$.

a) $\boldsymbol{X}_n$ **converges in distribution** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$, if $F_n(\boldsymbol{x}) \to F(\boldsymbol{x})$ as $n \to \infty$ for all points $\boldsymbol{x}$ at which $F(\boldsymbol{x})$ is continuous. The distribution of $\boldsymbol{X}$ is the **limiting distribution** or **asymptotic distribution** of $\boldsymbol{X}_n$.

b) $\boldsymbol{X}_n$ **converges in probability** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$, if for every $\epsilon > 0$, $P(\|\boldsymbol{X}_n - \boldsymbol{X}\| > \epsilon) \to 0$ as $n \to \infty$.

30) Multivariate Central Limit Theorem (MCLT): If $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ are iid $k \times 1$ random vectors with $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}_{\boldsymbol{x}}$, then

$$\sqrt{n}(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{x}})$$

where the sample mean

$$\overline{\boldsymbol{X}}_n = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i.$$

31) Suppose $\sqrt{n}(T_n - \boldsymbol{\mu}) \xrightarrow{D} N_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. Let $\boldsymbol{A}$ be a $q \times p$ constant matrix. Then $\boldsymbol{A}\sqrt{n}(T_n - \boldsymbol{\mu}) = \sqrt{n}(\boldsymbol{A}T_n - \boldsymbol{A}\boldsymbol{\mu}) \xrightarrow{D} N_q(\boldsymbol{A}\boldsymbol{\theta}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T)$.

32) Suppose $\boldsymbol{A}$ is a conformable constant matrix and $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$. Then $\boldsymbol{A}\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{A}\boldsymbol{X}$.

33) Given a table of data $\boldsymbol{W}$ for variables $X_1, ..., X_p$, be able to find the **coordinatewise median** MED($\boldsymbol{W}$) and the **sample mean** $\overline{\boldsymbol{x}}$. If $\boldsymbol{x} = (X_1, X_2, ..., X_p)^T$ where X_j corresponds to the jth column of $\boldsymbol{W}$, then MED($\boldsymbol{W}$) $= (\text{MED}_{X_1}(n), ..., \text{MED}_{X_p}(n))^T$ where $\text{MED}_{X_j}(n) = \text{MED}(X_{j,1}, ..., X_{j,n})$ is the sample median of the data in the jth column. Similarly, $\overline{\boldsymbol{x}} = (\overline{X}_1, ..., \overline{X}_p)^T$ where $\overline{X}_j$ is the sample mean of the data in the jth column. See Q3.

34) A **DD plot** is a plot of classical vs robust Mahalanobis distances. The DD plot is used to check i) if the data is MVN (plotted points follow the identity line), ii) if the data is EC but not MVN (plotted points follow a line through the origin with slope $>$ 1), iii) if the data is not EC (plotted points do not follow a line through the origin) iv) if multivariate outliers are present (eg some plotted points are far from the bulk of the data or the plotted points follow two lines). See Q3.

35) Many practical "robust estimators" generate a sequence of K trial fits called *attractors*: $(T_1, \boldsymbol{C}_1), ..., (T_K, \boldsymbol{C}_K)$. Then the attractor $(T_A, \boldsymbol{C}_A)$ that minimizes some criterion is used to obtain the final estimator. One way to obtain attractors is to generate trial fits called *starts*, and then use the *concentration* technique. Let $(T_{-1,j}, \boldsymbol{C}_{-1,j})$ be the jth start and compute all n Mahalanobis distances $D_i(T_{-1,j}, \boldsymbol{C}_{-1,j})$. At the next iteration, the classical estimator $(T_{0,j}, \boldsymbol{C}_{0,j})$ is computed from the $c_n \approx n/2$ cases corresponding to the smallest distances. This iteration can be continued for k steps resulting in the sequence of estimators $(T_{-1,j}, \boldsymbol{C}_{-1,j}), (T_{0,j}, \boldsymbol{C}_{0,j}), ..., (T_{k,j}, \boldsymbol{C}_{k,j})$. Then $(T_{k,j}, \boldsymbol{C}_{k,j})$ is the jth attractor for $j = 1, ..., K$. Using $k = 10$ often works well, and the basic resampling algorithm is a special case $k = -1$ where the attractors are the starts.

36) The DGK estimator $(T_{DGK}, \boldsymbol{C}_{DGK})$ uses the classical estimator $(T_{-1,D}, \boldsymbol{C}_{-1,D}) = (\overline{\boldsymbol{x}}, \boldsymbol{S})$ as the only start.

37) The median ball (MB) estimator $(T_{MB}, \boldsymbol{C}_{MB})$ uses $(T_{-1,M}, \boldsymbol{C}_{-1,M}) = (\text{MED}(\boldsymbol{W}), \boldsymbol{I}_p)$ as the only start where MED($\boldsymbol{W}$) is the coordinatewise median. Hence $(T_{0,M}, \boldsymbol{C}_{0,M})$ is the classical estimator applied to the "half set" of data closest to MED($\boldsymbol{W}$) in Euclidean distance.

38) Elemental concentration algorithms use elemental starts: $(T_{-1,j}, \boldsymbol{C}_{-1,j}) = (\overline{\boldsymbol{x}}_j, \boldsymbol{S}_j)$ is the classical estimator applied to a randomly selected "elemental set" of $p + 1$ cases. If the $\boldsymbol{x}_i$ are iid with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$, then the starts $(\overline{\boldsymbol{x}}_j, \boldsymbol{S}_j)$ are identically distributed with $E(\overline{\boldsymbol{x}}_j) = E(\boldsymbol{x}_i)$ and $\text{Cov}(\overline{\boldsymbol{x}}_j) = \boldsymbol{\Sigma}_{\boldsymbol{x}}/(p + 1)$.

39) Let the "median ball" be the hypersphere containing the half set of data closest to MED($\boldsymbol{W}$) in Euclidean distance. The FCH estimator uses the MB attractor if the DGK location estimator $T_{DGK} = T_{k,D}$ is outside of the median ball, and the attractor with the smallest determinant, otherwise. Let $(T_A, \boldsymbol{C}_A)$ be the attractor used. Then the

estimator $(T_{FCH}, \boldsymbol{C}_{FCH})$ takes $T_{FCH} = T_A$ and

$$\boldsymbol{C}_{FCH} = \frac{\text{MED}(D_i^2(T_A, \boldsymbol{C}_A))}{\chi^2_{p,0.5}} \boldsymbol{C}_A \tag{8}$$

where $\chi^2_{p,0.5}$ is the 50th percentile of a chi–square distribution with p degrees of freedom. The RFCH estimator uses two standard "reweight for efficiency steps" while the RMVN estimator uses a modified method for reweighting.

40) For a large class of elliptically contoured distributions, FCH, RFCH and RMVN are $\sqrt{n}$ consistent estimators of $(\boldsymbol{\mu}, c_i\boldsymbol{\Sigma})$ for $c_1, c_2, c_3 > 0$ where $c_i = 1$ for $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ data.

41) An estimator $(T, \boldsymbol{C})$ of multivariate location and dispersion (MLD), needs to estimate $p(p+3)/2$ unknown parameters when there are p random variables. For $(\overline{\boldsymbol{x}}, \boldsymbol{S})$ or $(\overline{\boldsymbol{z}}, \boldsymbol{R})$, want $n > 10p$. Want $n > 20p$ for FCH, RFCH or RMVN.

42) Brand name robust MLD estimators from the Rousseeuw and Yohai paradigm take too long to compute: fake-brand name estimators that are not backed by breakdown or large sample theory are actually used. Fake-MCD, Fake-MVE, Fake-S, Fake-MM, Fake-τ, Fake-constrained-M and Fake-Stahel-Donoho are especially common.

Sections covered: Olive (2012) 1.1, 1.2, 1.4, ch. 2, ch. 3 (skim ∮ 3.4) skim ch.4 with emphasis on p. 62, DGK, MG, FCH, RFCH and RMVN estimators, DD plot. From ∮ 5.1, Def. 5.1, Applications 5.1 and 5.2.

Johnson and Wichern (1988): 1.3, 1.4, ch. 2 is a review of linear algebra p. 45, 46, and sections 2.4, 2.5, 2.6 are important. Ch. 3: p. 89, 100, 103-4, ∮ 3.5 are important. Ch. 4: ∮ 4.2, 4.3 (omit proofs), p. 144-145, 155-157