Exam 1 review. 5 sheets of notes and a calculator. Friday Sept. 18.

Types of problems.

1) Given a small data set, find $\overline{Y}$, S, MED(n) and MAD(n). See HW1 problem 2.10 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) Find the population median $M = \text{MED}(Y)$ by solving the equation $F(M) = 0.5$ for M where the cdf $F(y) = P(Y \leq y)$. If Y has a pdf $f(y)$ that is symmetric about μ, then $M = \mu$. If $W = a + bY$, then $\text{MED}(W) = a + b\text{MED}(Y)$. Often $a = \mu$ and $b = \sigma$. See HW2 11.2.

3) To find the population median absolute deviation $D = \text{MAD}(Y)$, first find $M = \text{MED}(Y)$ as in 2) above. a) Then solve $F(M + D) - F(M - D) = 0.5$ for D.
b) If Y has a pdf that is symmetric about μ, then let $U = y_{0.75}$ where $P(Y \leq y_\alpha) = \alpha$, and y_α is the 100αth percentile of Y for $0 < \alpha < 1$. Hence $M = y_{0.5}$ is the 50th percentile and U is the 75th percentile. Solve $F(U) = 0.75$ for U. Then $D = U - M$.
c)If $W = a + bY$, then $\text{MAD}(W) = |b|\text{MAD}(Y)$. See HW2 11.3.

MED(Y) and MAD(Y) need not be unique, but for "brand name" continuous random variables, they are unique.

4) A large sample 100 $(1 - \alpha)$% confidence interval (CI) for θ is

$$\hat{\theta} \pm t_{p,1-\frac{\alpha}{2}} SE(\hat{\theta})$$

where $P(t_p \leq t_{p,1-\frac{\alpha}{2}}) = 1 - \alpha/2$ if t_p is from a t distribution with p degrees of freedom. We will use 95% CIs so $\alpha = 0.05$ and $t_{p,1-\frac{\alpha}{2}} = t_{p,0.975} \approx 1.96$ for $p > 20$. Be able to find $\hat{\theta}$, p and $SE(\hat{\theta})$ for the following three estimators. See Q2, Q3, HW2 problem 2.14.

a) The **classical CI for the population mean** $\theta = \mu$ uses $\hat{\theta} = \overline{Y}$, $p = n - 1$ and $SE(\overline{Y}) = S/\sqrt{n}$.

Let $\lfloor x \rfloor$ denote the "greatest integer function". Then $\lfloor x \rfloor$ is the largest integer less than or equal to x (eg, $\lfloor 7.7 \rfloor = 7$). Let $\lceil x \rceil$ denote the smallest integer greater than or equal to x (eg, $\lceil 7.7 \rceil = 8$).

b) Let $U_n = n - L_n$ where $L_n = \lfloor n/2 \rfloor - \lceil \sqrt{n/4}\, \rceil$. Then the **CI for the population median** $\theta = \text{MED}(Y)$ uses $\hat{\theta} = \text{MED}(n)$, $p = U_n - L_n - 1$ and

$$SE(\text{MED}(n)) = 0.5(Y_{(U_n)} - Y_{(L_n+1)}).$$

c) The 25% trimmed mean

$$T_n = T_n(L_n, U_n) = \frac{1}{U_n - L_n} \sum_{i=L_n+1}^{U_n} Y_{(i)}$$

where $L_n = \lfloor n/4 \rfloor$ and $U_n = n - L_n$. That is, order the data, delete the L_n smallest cases and the L_n largest cases and take the sample mean of the remaining $U_n - L_n$ cases. The 25% trimmed mean is estimating the population truncated mean

$$\mu_T = \int_{y_{0.25}}^{y_{0.75}} 2y f_Y(y) dy.$$

To perform inference, find $d_1, ..., d_n$ where

$$d_i = \begin{cases} Y_{(L_n+1)} & i \le L_n \\ Y_{(i)}, & L_n + 1 \le i \le U_n \\ Y_{(U_n)}, & i \ge U_n + 1. \end{cases}$$

(The "half set" of retained cases is not changed, but replace the L_n smallest deleted cases by the smallest retained case $Y_{(L_n+1)}$ and replace the L_n largest deleted cases by the largest retained case $Y_{(U_n)}$.) Then the Winsorized variance is the sample variance $S_n^2(d_1, ..., d_n)$ of $d_1, ..., d_n$, and the scaled Winsorized variance

$$V_{SW}(L_n, U_n) = \frac{S_n^2(d_1, ..., d_n)}{([U_n - L_n]/n)^2}.$$

Then the **CI for the population truncated mean** $\theta = \mu_T$ uses $\hat{\theta} = T_n$, $p = U_n - L_n - 1$ and

$$SE(T_n) = \sqrt{V_{SW}(L_n, U_n)/n}.$$

5) Consider intervals that contain c cases $[Y_{(1)}, Y_{(c)}], [Y_{(2)}, Y_{(c+1)}], ..., [Y_{(n-c+1)}, Y_{(n)}]$. Compute $Y_{(c)} - Y_{(1)}, Y_{(c+1)} - Y_{(2)}, ..., Y_{(n)} - Y_{(n-c+1)}$. Then the estimator shorth(c) = $[Y_{(s)}, Y_{(s+c-1)}]$ is the interval with the shortest length. The shorth(c) interval is a large sample $100(1-\delta)$% PI if $c/n \to 1 - \delta$ as $n \to \infty$ that estimates the population shorth. Hence the shorth PI often asymptotically optimal.

6) A large sample $100(1-\delta)$% prediction interval (PI) $[\hat{L}_n, \hat{U}_n]$ is such that $P(Y_f \in [\hat{L}_n, \hat{U}_n])$ is eventually bounded below by $1-\delta$ as $n \to \infty$. A large sample $100(1-\delta)$% PI is *asymptotically optimal* if it has the shortest asymptotic length: the length of $[\hat{L}_n, \hat{U}_n]$

converges to $U_s - L_s$ as $n \to \infty$ where $[L_s, U_s]$ is the *population shorth*: the shortest interval covering at least $100(1-\delta)\%$ of the mass. So $F(U_s) - F(L_s-) \geq 1-\delta$, and if $F(b) - F(a-) \geq 1-\delta$, then $b - a \geq U_s - L_s$. The population shorth need not be unique, but the length of the population shorth is unique.

7) The interval $[\hat{L}_n, \hat{U}_n]$ is a large sample $100(1-\delta)\%$ *confidence interval* for θ if $P(\hat{L}_n \leq \theta \leq \hat{U}_n)$ is eventually bounded below by $1-\delta$ as $n \to \infty$.

8) The δ quantile or 100δth percentile $y_\delta = \pi_\delta = \xi_\delta$ satisfies $P(Y \leq y_\delta) = \delta$. The *sample δ quantile* or sample 100δth percentile $\hat{\xi}_{n,\rho} = Y_{(\lceil n\delta \rceil)}$. Software often uses a slightly different definition.

9) Given B samples drawn with replacement from the cases (nonparametric bootstrap), be able to compute simple statistics T_j^* from the jth sample such as the sample mean, the sample median, the max, the min, the range = max − min. See Example 2.10. The bagging estimator is $\overline{T}^* = \frac{1}{B}\sum_{j=1}^{B} T_j^*$.

10) The bootstrap sample is $T_1^*, ..., T_B^*$. Often B is a fixed number such as $B = 1000$, but using $B = \max(1000, \lceil n \log(n) \rceil)$ works better if you want the coverage of the bootstrap CI to converge to $1-\delta$ as $n \to \infty$.

11) Given a bootstrap sample $T_1^*, ..., T_B^*$, let the order statistics be $T_{(1)}^*, ..., T_{(B)}^*$. Applying certain PIs to the bootstrap sample results in CIs. The shorth(c) CI is found as in 5) with n replaced by B. The prediction region method CI is $[\overline{T}^* - a, \overline{T}^* + a]$, which is the interval centered at $\overline{T}^*$ just long enough to contain $U_B \approx \lceil B(1-\delta) \rceil$ of the T_j^*. The modified Bickel and Ren CI is $[T_n - b, T_n + b]$, which is the interval centered at T_n just long enough to contain U_B of the T_j^*. Let $k_1 = \lceil B\delta/2 \rceil$ and $k_2 = \lceil B(1-\delta/2) \rceil$. The percentile CI is $[T_{(k_1)}^*, T_{(k_2)}^*]$, which deletes the $k_1 - 1$ smallest and $B - k_2$ largest T_j^*.

12) For a large sample level δ test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, reject H_0 if θ_0 is not in the large sample $100(1-\delta)\%$ confidence interval (CI) for θ. A bootstrap test corresponds to a bootstrap CI.

13) Given MED$(Y) = g(\theta, \sigma)$ and MAD$(Y) = h(\theta, \sigma)$, Solve MED$(n) \stackrel{set}{=} g(\theta, \sigma)$ and MAD$(n) \stackrel{set}{=} h(\theta, \sigma)$ for (θ, σ), and call the solution $(\hat{\theta}, \hat{\sigma})$. This method of obtaining robust estimators is similar to the method of moments when there are two parameters. Also see Chapter 11. Often MED$(Y) = g(\theta)$ and MAD$(Y) = h(\sigma)$.

Ch. 3

14) The *population mean* of a random $p \times 1$ vector $\boldsymbol{x} = (x_1, ..., x_p)^T$ is $E(\boldsymbol{x}) = \boldsymbol{\mu} = (E(x_1), ..., E(x_p))^T$. The *$p \times p$ population covariance matrix* $\text{Cov}(\boldsymbol{x}) = E(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x} - E(\boldsymbol{x}))^T = (\sigma_{ij}) = \boldsymbol{\Sigma}_{\boldsymbol{x}}$. The *$p \times p$ population correlation matrix* $\text{Cor}(\boldsymbol{x}) = \boldsymbol{\rho}_{\boldsymbol{x}} = (\rho_{ij})$.

15) 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 $\text{Cov}(\boldsymbol{AY}) = \boldsymbol{A}\text{Cov}(\boldsymbol{Y})\boldsymbol{A}^T$.

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

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

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

18) **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 $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$.

19)

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

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 $\sigma_Y > 0$.

20) 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,

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

21) Notation:

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

22) Be able to compute the above quantities if X_1 and X_2 are scalars.

23) Given a table of data $\boldsymbol{W}$ for variables $X_1, ..., X_p$, be able to find the **coordinatewise median** $\text{MED}(\boldsymbol{W})$ 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 $\text{MED}(\boldsymbol{W}) = (\text{MED}_{X_1}(n), ..., \text{MED}_{X_p}(n))^T$ 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.