Exam 3, Wednesday, Dec. 4, 15 sheets of notes and a calculator,
Final, day, TBA, time TBA 30 sheets of notes and a calculator

Know 89) from Exam 2 review.

90) A *large sample* $100(1-\delta)\%$ *prediction region* is a set $\mathcal{A}_n$ such that $P(\boldsymbol{x}_f \in \mathcal{A}_n)$ is eventually bounded below by $1-\delta$ as $n \to \infty$. A prediction region is *asymptotically optimal* if its volume converges in probability to the volume of the minimum volume covering region or the highest density region of the distribution of $\boldsymbol{x}_f$. Let $\boldsymbol{W}$ be the matrix with ith row $\boldsymbol{x}_i^T$.

91) Let the $p \times 1$ column vector $T = T(\boldsymbol{W})$ be a multivariate location estimator, and let the $p \times p$ symmetric positive definite matrix $\boldsymbol{C} = \boldsymbol{C}(\boldsymbol{W})$ be a dispersion estimator. Let $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ be the data where $\boldsymbol{x}_i$ is a $p \times 1$ vector. The **sample mean**

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

where $\boldsymbol{1}$ is the $n \times 1$ vector of ones. The **sample covariance matrix**

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

That is, the ij entry of $\boldsymbol{S}$ is the sample covariance S_{ij}. The *classical estimator of multivariate location and dispersion* is $(T, \boldsymbol{C}) = (\overline{\boldsymbol{x}}, \boldsymbol{S})$.

92) The ith *Mahalanobis distance* $D_i = \sqrt{D_i^2}$ where the ith *squared Mahalanobis distance* is

$$D_i^2 = D_i^2(T, \boldsymbol{C}) = D^2_{\boldsymbol{x}_i}(T, \boldsymbol{C}) = (\boldsymbol{x}_i - T)^T \boldsymbol{C}^{-1} (\boldsymbol{x}_i - T)$$

for each point $\boldsymbol{x}_i$ where

$$D^2_{\boldsymbol{x}}(T, \boldsymbol{C}) = (\boldsymbol{x} - T)^T \boldsymbol{C}^{-1} (\boldsymbol{x} - T).$$

Notice that D_i^2 is a random variable (scalar valued).

Note: Let $q_n = \min(1-\delta+0.05, 1-\delta+p/n)$ for $\delta > 0.1$ and

$$q_n = \min(1-\delta/2, 1-\delta+10\delta p/n), \quad \text{otherwise.}$$

If $1-\delta < 0.999$ and $q_n < 1-\delta+0.001$, set $q_n = 1-\delta$. Let $D_{(U_n)}$ be the $100 q_n$th sample percentile of the D_i.

93) Assume that $\boldsymbol{x}_1, ..., \boldsymbol{x}_n, \boldsymbol{x}_f$ are iid from a distribution with mean $E(\boldsymbol{x}) = \boldsymbol{\mu}$ and nonsingular covariance matrix $\text{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}_{\boldsymbol{x}}$. The large sample $100(1-\delta)\%$ *nonparametric prediction region* for a future value $\boldsymbol{x}_f$ is

$$\{\boldsymbol{z} : D^2_{\boldsymbol{z}}(\overline{\boldsymbol{x}}, \boldsymbol{S}) \le D^2_{(U_n)}\}$$

if $D^2_{1-\delta}$ is a continuity point of the cdf $F_{D^2}(y)$.

94) Assume that $\boldsymbol{x}_1, ..., \boldsymbol{x}_n, \boldsymbol{x}_f$ are iid $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{x}})$. Then the large sample $100(1-\delta)\%$ *MVN prediction region* is

$$\{\boldsymbol{z} : D^2_{\boldsymbol{z}}(\overline{\boldsymbol{x}}, \boldsymbol{S}) \le \chi^2_{p,1-\delta}\}.$$

95) The interval $[L_n, U_n]$ is a large sample $100(1-\delta)$ % CI for θ if

$$P_\theta(L_n \le \theta \le U_n)$$

is eventually bounded below by $1-\delta$ for all $\theta \in \Theta$ as the sample size $n \to \infty$.

96) Let the data $Y_1, ..., Y_n$ have joint pdf or pmf $f(\boldsymbol{y}|\boldsymbol{\theta})$ with parameter space Θ and support $\mathcal{Y}$. The quantity $R(\boldsymbol{Y}|\boldsymbol{\theta})$ is a **pivot** or pivotal quantity if the distribution of $R(\boldsymbol{Y}|\boldsymbol{\theta})$ is independent $\boldsymbol{\theta}$. The quantity $R(\boldsymbol{Y}, \boldsymbol{\theta})$ is an **asymptotic pivot** or asymptotic pivotal quantity if the limiting distribution of $R(\boldsymbol{Y}, \boldsymbol{\theta})$ is independent of $\boldsymbol{\theta}$.

97) A *bootstrap data set*, such as $Y_1^*, ..., Y_n^*$ or $(Y_1^*, \boldsymbol{x}_1^*), ..., (Y_n^*, \boldsymbol{x}_n^*)$, is used to compute a bootstrap statistic T^*. Repeat B times to get a *bootstrap sample* $T_1^*, ..., T_B^*$.

98) A *case* consists of the measurements taken on a person, object, or thing. If the data is $(Y_1, \boldsymbol{x}_1^T)^T, ..., (Y_n, \boldsymbol{x}_n^T)^T$, then the ith case is $(Y_i, \boldsymbol{x}_i^T)^T$. The *nonparametric bootstrap* creates a bootstrap data set by drawing a sample of size n with replacement from the n cases. Then T^* is computed from the bootstrap data set.

If $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ is the data set and if $\boldsymbol{x}_1^*, ..., \boldsymbol{x}_n^*$ is the bootstrap data set, then the bootstrap distribution for the bootstrap data is the empirical distribution: if $\boldsymbol{w}$ is a random vector from the empirical distribution, then $P(\boldsymbol{w} = \boldsymbol{x}_i) = 1/n$ for $i = 1, ..., n$. This distribution is discrete with a pmf. $E(\boldsymbol{w}) = \overline{\boldsymbol{x}}_n$ and $\text{Cov}(\boldsymbol{w}) = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}} = (n-1)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}/n = (n-1)\boldsymbol{S}_n/n$

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

99) We do not have an iid sample of size B: $T_1, ..., T_B$, of the statistic T_n (with $T_i = T_{in}$), but $\sqrt{n}(T_1^* - T_n), ..., \sqrt{n}(T_B^* - T_n)$ can be regarded as pseudodata for $\sqrt{n}(T_1 - \boldsymbol{\theta}), ..., \sqrt{n}(T_B - \boldsymbol{\theta})$ if $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$ and $\sqrt{n}(T_i^* - T_n) \xrightarrow{D} \boldsymbol{u}$. Let the sample mean and sample covariance matrix of the bootstrap sample $T_1^*, ..., T_B^*$ be

$$\overline{T}^* = \frac{1}{B}\sum_{i=1}^B T_i^* \quad \text{and} \quad \boldsymbol{S}_{\mathrm{T}}^* = \frac{1}{\mathrm{B}-1}\sum_{\mathrm{i}=1}^{\mathrm{B}} (\mathrm{T}_{\mathrm{i}}^* - \overline{\mathrm{T}}^*)(\mathrm{T}_{\mathrm{i}}^* - \overline{\mathrm{T}}^*)^{\mathrm{T}}.$$

Then $\overline{T}^*$ is known as the *bagging estimator*.

100) The bootstrap *percentile method* large sample $100(1-\delta)$% confidence interval for θ is an interval $[T_{(k_L)}^*, T_{(K_U)}^*]$ containing $\approx \lceil B(1-\delta) \rceil$ of the T_i^*. Let $k_1 = \lceil B\delta/2 \rceil$ and $k_2 = \lceil B(1-\delta/2) \rceil$. A common choice is $[T_{(k_1)}^*, T_{(k_2)}^*]$.

101) The large sample $100(1-\delta)$% *shorth*(c) *CI*

$$[T_{(s)}^*, T_{(s+c-1)}^*] \quad \text{where} \quad \mathrm{c} = \min(\mathrm{B}, \lceil \mathrm{B}[1 - \delta + 1.12\sqrt{\delta/\mathrm{B}}\ ]\ \rceil).$$

102) The large sample $100(1-\delta)$% *standard bootstrap confidence region* for $\boldsymbol{\theta}$ is $\{\boldsymbol{w} : (\boldsymbol{w} - T_n)^T [\boldsymbol{S}_T^*]^{-1} (\boldsymbol{w} - T_n) \le D_{1-\delta}^2\} =$

$$\{\boldsymbol{w} : D_{\boldsymbol{w}}^2(T_n, \boldsymbol{S}_T^*) \le D_{1-\delta}^2\}$$

where $D_{1-\delta}^2 = \chi_{g,1-\delta}^2$ or $D_{1-\delta}^2 = d_n F_{g,d_n,1-\delta}$ where $d_n \to \infty$ as $n \to \infty$. This confidence region needs $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} N_g(\boldsymbol{0}, \boldsymbol{\Sigma}_A)$ and $n\boldsymbol{S}_T^* \xrightarrow{P} \boldsymbol{\Sigma}_A > 0$ as $n, B \to \infty$.

103) The large sample $100(1-\delta)\%$ *prediction region method confidence region* for $\boldsymbol{\theta}$ is $\{\boldsymbol{w} : (\boldsymbol{w} - \overline{T}^*)^T[\boldsymbol{S}_T^*]^{-1}(\boldsymbol{w} - \overline{T}^*) \leq D^2_{(U_B)}\} =$

$$\{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{T}^*, \boldsymbol{S}_T^*) \leq D^2_{(U_B)}\}$$

where $D^2_{(U_B)}$ is computed from $D_i^2 = (T_i^* - \overline{T}^*)^T[\boldsymbol{S}_T^*]^{-1}(T_i^* - \overline{T}^*)$ for $i = 1, ..., B$. Note that the corresponding test for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ rejects H_0 if $(\overline{T}^* - \boldsymbol{\theta}_0)^T[\boldsymbol{S}_T^*]^{-1}(\overline{T}^* - \boldsymbol{\theta}_0) > D^2_{(U_B)}$.

Let $q_B = \min(1 - \delta + 0.05, 1 - \delta + g/B)$ for $\delta > 0.1$ and

$$q_B = \min(1 - \delta/2, 1 - \delta + 10\delta g/B), \quad \text{otherwise.}$$

If $1 - \delta < 0.999$ and $q_B < 1 - \delta + 0.001$, set $q_B = 1 - \delta$. Then $D_{(U_B)}$ is the $100q_B$th sample quantile of the D_i.

104) CI 100) is the percentile PI 84) applied to the bootstrap sample, CI 101) is the shorth PI 86) applied to the bootstrap sample, confidence region 103) is the nonparametric prediction region 93) applied to the bootstrap sample. (Confidence region 102) is nearly the MVN prediction region 94) applied to the bootstrap sample.)

105) a) The large sample $100(1-\delta)\%$ *BR confidence region* is

$$\{\boldsymbol{w} : n(\boldsymbol{w} - T_n)^T \boldsymbol{C}_n^{-1}(\boldsymbol{w} - T_n) \leq D^2_{(U_B,T)}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(T_n, \boldsymbol{C}_n/n) \leq D^2_{(U_B,T)}\}$$

where the cutoff $D^2_{(U_B,T)}$ is the $100q_B$th sample quantile of the $D_i^2 = n(T_i^* - T_n)^T \boldsymbol{C}_n^{-1}(T_i^* - T_n)$.

b) The large sample $100(1-\delta)\%$ *PR confidence region* for $\boldsymbol{\theta}$ is

$$\{\boldsymbol{w} : n(\boldsymbol{w} - \overline{T}^*)^T \boldsymbol{C}_n^{-1}(\boldsymbol{w} - \overline{T}^*) \leq D^2_{(U_B)}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{T}^*, \boldsymbol{C}_n/n) \leq D^2_{(U_B)}\}$$

where $D^2_{(U_B)}$ is computed from $D_i^2 = n(T_i^* - \overline{T}^*)^T \boldsymbol{C}_n^{-1}(T_i^* - \overline{T}^*)$ for $i = 1, ..., B$.

106) Assume i) $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$, ii) $\sqrt{n}(T_i^* - T_n) \xrightarrow{D} \boldsymbol{u}$, iii) $\sqrt{n}(\overline{T}^* - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$, and iv) $\sqrt{n}(T_i^* - \overline{T}^*) \xrightarrow{D} \boldsymbol{u}$. Also assume $\boldsymbol{C}_n^{-1} \xrightarrow{P} \boldsymbol{C}^{-1}$. Then

$$D_1^2 = D^2_{T_i^*}(\overline{T}^*, \boldsymbol{C}_n/n) = \sqrt{n}(T_i^* - \overline{T}^*)^T \boldsymbol{C}_n^{-1} \sqrt{n}(T_i^* - \overline{T}^*),$$

$$D_2^2 = D^2_{\boldsymbol{\theta}}(T_n, \boldsymbol{C}_n/n) = \sqrt{n}(T_n - \boldsymbol{\theta})^T \boldsymbol{C}_n^{-1} \sqrt{n}(T_n - \boldsymbol{\theta}),$$

$$D_3^2 = D^2_{\boldsymbol{\theta}}(\overline{T}^*, \boldsymbol{C}_n/n) = \sqrt{n}(\overline{T}^* - \boldsymbol{\theta})^T \boldsymbol{C}_n^{-1} \sqrt{n}(\overline{T}^* - \boldsymbol{\theta}), \quad \text{and}$$

$$D_4^2 = D^2_{T_i^*}(T_n, \boldsymbol{C}_n/n) = \sqrt{n}(T_i^* - T_n)^T \boldsymbol{C}_n^{-1} \sqrt{n}(T_i^* - T_n)$$

satisfy $D_j^2 \xrightarrow{D} D^2 = \boldsymbol{u}^T \boldsymbol{C}^{-1} \boldsymbol{u}$, and 105) gives two large sample confidence regions. If $\boldsymbol{C}_n^{-1}$ is "not too ill conditioned" then $D_j^2 \approx \boldsymbol{u}^T \boldsymbol{C}_n^{-1} \boldsymbol{u}$ for large n, and the confidence regions in 105) will have coverage near $1 - \delta$.

107) Theorem: Bootstrap Proof Technique (BPT): Suppose $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} N_g(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}_n \xrightarrow{P} \boldsymbol{\Sigma}$ as $n \to \infty$, and for fixed n, $\sqrt{m}(T^*_{n,m} - T_n) \xrightarrow{D} N_g(\boldsymbol{0}, \boldsymbol{\Sigma}_n)$ as $m \to \infty$. Then a) $\sqrt{m}(T^*_{n,m} - T_n) \xrightarrow{D} N_g(\boldsymbol{0}, \boldsymbol{\Sigma})$ as $m, n \to \infty$. Also b) $\sqrt{n}(T^*_n - T_n) \xrightarrow{D} N_g(\boldsymbol{0}, \boldsymbol{\Sigma})$ as $n \to \infty$ where $T^*_n = T^*_{n,n}$ has $m = n$.

108) CI technique when $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2(\theta))$.

$$\left[T_n - z_{1-\delta/2} \frac{\sigma(\hat{\theta})}{\sqrt{n}}, T_n + z_{1-\delta/2} \frac{\sigma(\hat{\theta})}{\sqrt{n}} \right]$$

is a large sample $100(1-\delta)\%$ CI for θ.

Make sure you use $\hat{\theta}$. If $\sigma^2(\beta) = 1/I_1(\beta)$, then $\sigma(\hat{\beta}) = 1/\sqrt{I_1(\hat{\beta})}$.

Linear Model Olive Ch 5:

109) The *multiple linear regression model with iid errors* is $Y_i = Y_i|\boldsymbol{x}_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ for $i = 1, ..., n$ where the e_i are iid with $E(e_i) = 0$ and $V(e_i) = \sigma^2$. Unless told otherwise, $\boldsymbol{x}_i = (1, x_{i2}, ..., x_{ip})^T$ so a constant is in the model. In matrix form, $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ where $E(\boldsymbol{e}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{Y}) = \text{Cov}(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_n$. Treat $\boldsymbol{X\beta}$ as a constant (if necessary condition on $\boldsymbol{X}$) for this model, which is fit with ordinary least squares (OLS). The ith row of $\boldsymbol{X}$ is $\boldsymbol{x}_i^T$. Assume $\boldsymbol{X}$ is an $n \times p$ matrix with full rank p.

a) $\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$.

b) $E(\hat{\boldsymbol{\beta}}_{OLS}) = \boldsymbol{\beta}$. Hence $\hat{\boldsymbol{\beta}}_{OLS}$ is an unbiased estimator of $\boldsymbol{\beta}$.

c) $\text{Cov}(\hat{\boldsymbol{\beta}}_{OLS}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

d)

$$\hat{\boldsymbol{\beta}}_{OLS} \sim AN_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}) \ \text{ and } \ \hat{\boldsymbol{\beta}}_{\text{OLS}} \sim \text{AN}_\text{p}(\boldsymbol{\beta}, \text{MSE}(\boldsymbol{X}^\text{T}\boldsymbol{X})^{-1})$$

where $MSE = \dfrac{1}{n-p} \displaystyle\sum_{i=1}^n r_i^2$ where the ith fitted value $\hat{Y}_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{OLS}$ and the ith residual $r_i = Y_i - \hat{Y}_i$. MSE is a consistent estimator of σ^2 under mild conditions.

e) $\boldsymbol{H} = \boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$. Let $h_i = \boldsymbol{H}_{ii}$ be the ith diagonal element of $\boldsymbol{H}$. Note that $\boldsymbol{H} = \boldsymbol{H}^T = \boldsymbol{H}^2$.

f) The vector of *predicted* or *fitted values* $\widehat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{HY}$.

g) The vector of residuals is $\boldsymbol{r} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$.

h) $\boldsymbol{HX} = \boldsymbol{X}$, so $\boldsymbol{X}^T\boldsymbol{H} = \boldsymbol{X}^T$.

i) If $\boldsymbol{e} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$, then $\boldsymbol{Y} \sim N_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I}_n)$, and

$$\hat{\boldsymbol{\beta}}_{OLS} \sim N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}).$$

Then if $\boldsymbol{A}$ is a $k \times p$ matrix, $\boldsymbol{AY} \sim N_k(\boldsymbol{AX\beta}, \sigma^2\boldsymbol{AA}^T)$, $\hat{\boldsymbol{Y}} = \boldsymbol{HY} \sim N_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{H})$, and $\boldsymbol{r} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} \sim N_n(\boldsymbol{0}, \sigma^2(\boldsymbol{I} - \boldsymbol{H}))$.

110) (**OLS CLT**:) Suppose that MLR model in 109) holds. Assume $\max(h_i) \to 0$ and

$$\frac{\boldsymbol{X}^T\boldsymbol{X}}{n} \to \boldsymbol{W}^{-1}$$

as $n \to \infty$. Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2 \ \boldsymbol{W}).$$

111) **Know:** If $\boldsymbol{Z}_n \xrightarrow{D} N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{AZ}_n + \boldsymbol{b} \xrightarrow{D} N_m(\boldsymbol{A\mu} + \boldsymbol{b}, \boldsymbol{A\Sigma A}^T)$ where $\boldsymbol{A}$ is an $m \times k$ constant matrix and $\boldsymbol{b}$ is an $m \times 1$ constant vector.

112) The model $Y = \boldsymbol{x}^T \boldsymbol{\beta} + e$ that uses all of the predictors is called the *full model.* A model $Y = \boldsymbol{x}_I^T \boldsymbol{\beta}_I + e$ that only uses a subset $\boldsymbol{x}_I$ of the predictors is called a *submodel.* The **full model is always a submodel**.

113) Let $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$. A *model for variable selection* is $\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{x}_S^T \boldsymbol{\beta}_S + \boldsymbol{x}_E^T \boldsymbol{\beta}_E = \boldsymbol{x}_S^T \boldsymbol{\beta}_S$ where $\boldsymbol{x} = (\boldsymbol{x}_S^T, \boldsymbol{x}_E^T)^T$, $\boldsymbol{x}_S$ is an $a_S \times 1$ vector, and $\boldsymbol{x}_E$ is a $(p - a_S) \times 1$ vector. Let $\boldsymbol{x}_I$ be the vector of a terms from a candidate subset indexed by I, and let $\boldsymbol{x}_O$ be the vector of the remaining predictors (out of the candidate submodel). If $S \subseteq I$, then $\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{x}_S^T\boldsymbol{\beta}_S = \boldsymbol{x}_S^T\boldsymbol{\beta}_S + \boldsymbol{x}_{I/S}^T \boldsymbol{\beta}_{(I/S)} + \boldsymbol{x}_O^T \boldsymbol{0} = \boldsymbol{x}_I^T \boldsymbol{\beta}_I$ where $\boldsymbol{x}_{I/S}$ denotes the predictors in I that are not in S. Since this is true regardless of the values of the predictors, $\boldsymbol{\beta}_O = \boldsymbol{0}$ if $S \subseteq I$. Note that $\boldsymbol{\beta}_E = \boldsymbol{0}$. Let $k_S = a_S - 1 =$ the number of population active nontrivial predictors. Then $k = a - 1$ is the number of active predictors in the candidate submodel I. If $S \subseteq I$, then $\boldsymbol{Y} = \boldsymbol{X}_I \boldsymbol{\beta}_I + \boldsymbol{e}$. Then $\boldsymbol{H}\boldsymbol{H}_I = \boldsymbol{H}_I \boldsymbol{H} = \boldsymbol{H}_I$.

114) The **parametric bootstrap** for MLR has $\boldsymbol{Y}^* \sim N_n(\boldsymbol{X}\hat{\boldsymbol{\beta}}, \hat{\sigma}_n^2 \boldsymbol{I}) \sim N_n(\boldsymbol{H}\boldsymbol{Y}, \hat{\sigma}_n^2 \boldsymbol{I})$ where **we are not assuming** that the $e_i \sim N(0, \sigma^2)$, and $\hat{\sigma}_n^2 = MSE$ where the residuals are from the full OLS model. Here $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{OLS}$. Thus $\hat{\boldsymbol{\beta}}_I^* = (\boldsymbol{X}_I^T \boldsymbol{X}_I)^{-1} \boldsymbol{X}_I^T \boldsymbol{Y}^* \sim N_{a_I}(\hat{\boldsymbol{\beta}}_I, \hat{\sigma}_n^2 (\boldsymbol{X}_I^T \boldsymbol{X}_I)^{-1})$ since $E(\hat{\boldsymbol{\beta}}_I^*) = (\boldsymbol{X}_I^T \boldsymbol{X}_I)^{-1} \boldsymbol{X}_I^T \boldsymbol{H}\boldsymbol{Y} = \hat{\boldsymbol{\beta}}_I$ because $\boldsymbol{H}\boldsymbol{X}_I = \boldsymbol{X}_I$, and $\text{Cov}(\hat{\boldsymbol{\beta}}_I^*) = \hat{\sigma}_n^2 (\boldsymbol{X}_I^T \boldsymbol{X}_I)^{-1}$. Hence

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_I^* - \hat{\boldsymbol{\beta}}_I) \sim N_{a_I}(\boldsymbol{0}, n\hat{\sigma}_n^2 (\boldsymbol{X}_I^T \boldsymbol{X}_I)^{-1}) \xrightarrow{D} N_{a_I}(\boldsymbol{0}, \sigma^2 \boldsymbol{W}_I)$$

as $n, B \to \infty$ if $S \subseteq I$.

115) Second way to compute $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{OLS}$: Assume a constant β_1 is in the model so the first column of $\boldsymbol{X}$ is $\boldsymbol{x}_1 = \boldsymbol{1}$, an $n \times 1$ vector of ones. Let $\boldsymbol{x} = (1 \;\; \boldsymbol{u}^T)^T$.

a) If $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}}^{-1}$ exists, then $\hat{\beta}_1 = \overline{Y} - \hat{\boldsymbol{\beta}}_2^T \overline{\boldsymbol{u}}$ and

$$\hat{\boldsymbol{\beta}}_2 = \frac{n}{n-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}}^{-1} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{u}}^{-1} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}}^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y}.$$

b) Suppose that $(Y_i, \boldsymbol{u}_i^T)^T$ are iid random vectors such that σ_Y^2, $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}$, and $\boldsymbol{\Sigma}_{\boldsymbol{u}Y}$ exist. Then $\hat{\beta}_1 \xrightarrow{P} \beta_1$ and

$$\hat{\boldsymbol{\beta}}_2 \xrightarrow{P} \boldsymbol{\beta}_2 \quad \text{as} \quad \text{n} \to \infty.$$

c) Alternatively, take $Y = \alpha + \boldsymbol{\beta}^T \boldsymbol{x} + e$. Hence $\alpha = \beta_1$, $\boldsymbol{\beta} = \boldsymbol{\beta}_2$, and $\boldsymbol{x} = \boldsymbol{u}$ in the above model. Then $\hat{\alpha}_{OLS} = \overline{Y} - \hat{\boldsymbol{\beta}}^T \overline{\boldsymbol{x}}$ and

$$\hat{\boldsymbol{\beta}}_{OLS} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}.$$

Then $\hat{\alpha} \xrightarrow{P} \alpha = E(Y) - \boldsymbol{\beta}^T E(\boldsymbol{x})$ while $\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$.

116) Let $Y = \alpha + \boldsymbol{\beta}^T \boldsymbol{x} + e$. Assume the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid. Assume $Y \perp\!\!\!\perp \boldsymbol{x} | \boldsymbol{\beta}^T \boldsymbol{x}$ and that the e_i are iid with $E(e_i) = 0$ and $V(e_i) = \sigma^2$. Note that $\text{Cov}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}$ $\text{Cov}(\boldsymbol{x})\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{A}^T$, and $\text{Cov}(\boldsymbol{A}\boldsymbol{x}, Y) = \boldsymbol{A}$ $\text{Cov}(\boldsymbol{x}, Y) = \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{x},Y}$. Similarly, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{A}\boldsymbol{x}} = \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\boldsymbol{A}^T$, $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{A}\boldsymbol{x}} = \boldsymbol{A}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\boldsymbol{A}^T$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{A}\boldsymbol{x},Y} = \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y}$, and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{A}\boldsymbol{x},Y} = \boldsymbol{A}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y}$. Let $\boldsymbol{\beta}_q$ be a $q \times 1$ vector and $\boldsymbol{\Lambda}_q$ be a $q \times 1$ unknown constant compressor matrix. Let $\boldsymbol{z} = \boldsymbol{z}_q = \boldsymbol{\Lambda}_q \boldsymbol{x}$ be a $q \times 1$ unknown vector of latent variables. The population PLS (partial least squares) models are $Y = \alpha + \boldsymbol{\beta}_Q^T \boldsymbol{\Lambda}_q^T \boldsymbol{x} + e$ for $q = 1, ..., p$ where $\boldsymbol{\Lambda}_p = \boldsymbol{I}$ and $\boldsymbol{\beta}_p = \boldsymbol{\beta}$. Then

$\boldsymbol{\beta}^T = \boldsymbol{\beta}_q^T \boldsymbol{\Lambda}_q^T = [\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}]^T$ with $\boldsymbol{\beta} = \boldsymbol{\Lambda}_q \boldsymbol{\beta}_q = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. We also have that $\boldsymbol{\beta}_q = \boldsymbol{\Sigma}_{\boldsymbol{z}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{z}Y} = (\boldsymbol{\Lambda}_q^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\Lambda}_q)^T \boldsymbol{\Lambda}_q^T \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Then $\boldsymbol{\beta}_{qPLS} = \boldsymbol{\Lambda}_q \boldsymbol{\beta}_q = \boldsymbol{\beta}_{OLS} = \boldsymbol{\beta} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. In particular, the one component pop. PLS model has

$$\boldsymbol{\beta}_{1PLS} = \boldsymbol{\beta}_{OPLS} = \frac{\boldsymbol{\Lambda}_1 \boldsymbol{\Lambda}_1^T}{\boldsymbol{\Lambda}_1^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\Lambda}_1} \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \lambda \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$$

where

$$\lambda = \frac{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T \boldsymbol{\Sigma}_{\boldsymbol{x}Y}}{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}}.$$

117) Let $\hat{\boldsymbol{\beta}}_{OPLS} = \hat{\lambda} \hat{\boldsymbol{\eta}}$ where $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$. Let $\boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \boldsymbol{\Sigma}_{\boldsymbol{w}}$ where $\boldsymbol{w}_i = (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})(Y_i - \boldsymbol{\mu}_Y)$. Let $\boldsymbol{z}_i = \boldsymbol{x}_i(Y_i - \overline{Y}_n)$. If $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})^T$, assume the $E(x_{ij}^k Y_i^m)$ exist for $j = 1, ..., p$ and $k, m = 0, 1, 2$. An estimator of $\lambda^2 \boldsymbol{A} \boldsymbol{\Sigma}_{\boldsymbol{\eta}} \boldsymbol{A}^T$ is $\hat{\lambda}^2 \boldsymbol{A} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}} \boldsymbol{A}^T$.

OPLS Theorem: Assume

$$\sqrt{n} \left(\begin{pmatrix} \hat{\lambda} \\ \hat{\boldsymbol{\eta}} \end{pmatrix} - \begin{pmatrix} \lambda \\ \boldsymbol{\eta} \end{pmatrix} \right) \xrightarrow{D} N_{p+1} \left(\begin{pmatrix} 0 \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \Sigma_\lambda & \boldsymbol{\Sigma}_{\lambda \boldsymbol{\eta}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\eta} \lambda} & \boldsymbol{\Sigma}_{\boldsymbol{\eta}} \end{pmatrix} \right) \sim N_{p+1}(\boldsymbol{0}, \boldsymbol{\Sigma}).$$

a) $\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}})$.

b)

$$\sqrt{n}(\hat{\lambda} \hat{\boldsymbol{\eta}} - \lambda \boldsymbol{\eta}) = \sqrt{n}(\hat{\boldsymbol{\beta}}_{OPLS} - \boldsymbol{\beta}) \xrightarrow{D} N_p \left(\boldsymbol{0}, \boldsymbol{D} \boldsymbol{\Sigma} \boldsymbol{D}^T \right)$$

with $\boldsymbol{D} = [\boldsymbol{\eta} \;\; \lambda \boldsymbol{I}_p]$ where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix.

c) Let $\boldsymbol{A}$ be a $k \times p$ full rank constant matrix with $k \leq p$ and $\boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{0} = \boldsymbol{A}\boldsymbol{\eta}$. Then

$$\sqrt{n}(\boldsymbol{A} \hat{\boldsymbol{\beta}}_{OPLS} - \boldsymbol{0}) \xrightarrow{D} N_k \left(\boldsymbol{0}, \lambda^2 \boldsymbol{A} \boldsymbol{\Sigma}_{\boldsymbol{\eta}} \boldsymbol{A}^T \right).$$

118) A random vector $\boldsymbol{u}$ has a *mixture distribution* if $\boldsymbol{u}$ equals a random vector $\boldsymbol{u}_j$ with probability π_j for $j = 1, ..., J$.

The distribution of a $g \times 1$ random vector $\boldsymbol{u}$ is a mixture distribution if the cumulative distribution function (cdf) of $\boldsymbol{u}$ is

$$F_{\boldsymbol{u}}(\boldsymbol{t}) = \sum_{j=1}^J \pi_j F_{\boldsymbol{u}_j}(\boldsymbol{t})$$

where the probabilities π_j satisfy $0 \leq \pi_j \leq 1$ and $\sum_{j=1}^J \pi_j = 1$, $J \geq 2$, and $F_{\boldsymbol{u}_j}(\boldsymbol{t})$ is the cdf of a $g \times 1$ random vector $\boldsymbol{u}_j$. Then $\boldsymbol{u}$ has a mixture distribution of the $\boldsymbol{u}_j$ with probabilities π_j.

119) **Theorem.** Suppose $E(h(\boldsymbol{u}))$ and the $E(h(\boldsymbol{u}_j))$ exist. Then

$$E[h(\boldsymbol{u})] = \sum_{j=1}^J \pi_j E[h(\boldsymbol{u}_j)].$$

Hence

$$E(\boldsymbol{u}) = \sum_{j=1}^J \pi_j E[\boldsymbol{u}_j],$$

and $Cov(\boldsymbol{u}) = E(\boldsymbol{u}\boldsymbol{u}^T) - E(\boldsymbol{u})E(\boldsymbol{u}^T) = E(\boldsymbol{u}\boldsymbol{u}^T) - E(\boldsymbol{u})[E(\boldsymbol{u})]^T =$
$\sum_{j=1}^J \pi_j E[\boldsymbol{u}_j \boldsymbol{u}_j^T] - E(\boldsymbol{u})[E(\boldsymbol{u})]^T =$

$$\sum_{j=1}^J \pi_j Cov(\boldsymbol{u}_j) + \sum_{j=1}^J \pi_j E(\boldsymbol{u}_j)[E(\boldsymbol{u}_j)]^T - E(\boldsymbol{u})[E(\boldsymbol{u})]^T.$$

If $E(\boldsymbol{u}_j) = \boldsymbol{\theta}$ for $j = 1, ..., J$, then $E(\boldsymbol{u}) = \boldsymbol{\theta}$ and

$$Cov(\boldsymbol{u}) = \sum_{j=1}^J \pi_j Cov(\boldsymbol{u}_j).$$

120) Suppose $Y \perp\!\!\!\perp \boldsymbol{x}|\boldsymbol{\beta}^T\boldsymbol{x}$ (the regression model is not necessarily MLR). Let I_{min} correspond to the set of predictors selected by a variable selection method such as forward selection or lasso variable selection. If $\hat{\boldsymbol{\beta}}_I$ is $a \times 1$, use zero padding to form the $p \times 1$ vector $\hat{\boldsymbol{\beta}}_{I,0}$ from $\hat{\boldsymbol{\beta}}_I$ by adding 0s corresponding to the omitted variables. Then the observed variable selection estimator $\hat{\boldsymbol{\beta}}_{VS} = \hat{\boldsymbol{\beta}}_{I_{min},0}$. As a statistic, $\hat{\boldsymbol{\beta}}_{VS} = \hat{\boldsymbol{\beta}}_{I_k,0}$ with probabilities $\pi_{kn} = P(I_{min} = I_k)$ for $k = 1, ..., J$ where there are J subsets, e.g. $J = 2^p - 1$.

121) Let $\hat{\boldsymbol{\beta}}_{MIX}$ be a random vector with a mixture distribution of the $\hat{\boldsymbol{\beta}}_{I_k,0}$ with probabilities equal to π_{kn}. Hence $\hat{\boldsymbol{\beta}}_{MIX} = \hat{\boldsymbol{\beta}}_{I_k,0}$ with the same probabilities π_{kn} of the variable selection estimator $\hat{\boldsymbol{\beta}}_{VS}$, but the I_k are randomly selected.

122) **Theorem.** Assume $P(S \subseteq I_{min}) \to 1$ as $n \to \infty$, and let $\hat{\boldsymbol{\beta}}_{MIX} = \hat{\boldsymbol{\beta}}_{I_k,0}$ with probabilities π_{kn} where $\pi_{kn} \to \pi_k$ as $n \to \infty$. Denote the positive π_k by π_j. Assume $\boldsymbol{u}_{jn} = \sqrt{n}(\hat{\boldsymbol{\beta}}_{I_j,0} - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{u}_j \sim N_p(\boldsymbol{0}, \boldsymbol{V}_{j,0})$. a) Then

$$\boldsymbol{u}_n = \sqrt{n}(\hat{\boldsymbol{\beta}}_{MIX} - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{u} \tag{1}$$

where the cdf of $\boldsymbol{u}$ is $F_{\boldsymbol{u}}(\boldsymbol{t}) = \sum_j \pi_j F_{\boldsymbol{u}_j}(\boldsymbol{t})$.

b) Let $\boldsymbol{A}$ be a $g \times p$ full rank matrix with $1 \le g \le p$. Then

$$\boldsymbol{v}_n = \boldsymbol{A}\boldsymbol{u}_n = \sqrt{n}(\boldsymbol{A}\hat{\boldsymbol{\beta}}_{MIX} - \boldsymbol{A}\boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{A}\boldsymbol{u} = \boldsymbol{v} \tag{2}$$

where $\boldsymbol{v}$ has a mixture distribution of the $\boldsymbol{v}_j = \boldsymbol{A}\boldsymbol{u}_j \sim N_g(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{V}_{j,0}\boldsymbol{A}^T)$ with probabilities π_j.

c) The estimator $\hat{\boldsymbol{\beta}}_{VS}$ is a $\sqrt{n}$ consistent estimator of $\boldsymbol{\beta}$: $\sqrt{n}(\hat{\boldsymbol{\beta}}_{VS} - \boldsymbol{\beta}) = O_P(1)$.

d) If $\pi_d = 1$, then $\sqrt{n}(\hat{\boldsymbol{\beta}}_{SEL} - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{u} \sim N_p(\boldsymbol{0}, \boldsymbol{V}_{d,0})$ where SEL is VS or MIX.

123) The following subscript notation is useful. Subscripts before the MIX are used for subsets of $\hat{\boldsymbol{\beta}}_{MIX} = (\hat{\beta}_1, ..., \hat{\beta}_p)^T$. Let $\hat{\boldsymbol{\beta}}_{i,MIX} = \hat{\beta}_i$. Similarly, if $I = \{i_1, ..., i_a\}$, then $\hat{\boldsymbol{\beta}}_{I,MIX} = (\hat{\beta}_{i_1}, ..., \hat{\beta}_{i_a})^T$. Subscripts after MIX denote the ith vector from a sample $\hat{\boldsymbol{\beta}}_{MIX,1}, ..., \hat{\boldsymbol{\beta}}_{MIX,B}$. Similar notation is used for other estimators such as $\hat{\boldsymbol{\beta}}_{VS}$. The subscript 0 is still used for zero padding. We may use $FULL$ to denote the full model $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{FULL}$.

124) Use the conditional distribution of $\hat{\boldsymbol{\beta}}_{VS}|(\hat{\boldsymbol{\beta}}_{VS} = \hat{\boldsymbol{\beta}}_{I_k,0})$ to find the distribution of $\boldsymbol{w}_n = \sqrt{n}(\hat{\boldsymbol{\beta}}_{VS} - \boldsymbol{\beta})$. Let $\hat{\boldsymbol{\beta}}^C_{I_k,0}$ be a random vector from the conditional distribution

$\hat{\boldsymbol{\beta}}_{I_k,0}|(\hat{\boldsymbol{\beta}}_{VS} = \hat{\boldsymbol{\beta}}_{I_k,0})$. Let $\boldsymbol{w}_{kn} = \sqrt{n}(\hat{\boldsymbol{\beta}}_{I_k,0} - \boldsymbol{\beta})|(\hat{\boldsymbol{\beta}}_{VS} = \hat{\boldsymbol{\beta}}_{I_k,0}) \sim \sqrt{n}(\hat{\boldsymbol{\beta}}^C_{I_k,0} - \boldsymbol{\beta})$. Denote $F_{\boldsymbol{z}}(\boldsymbol{t}) = P(z_1 \le t_1, ..., z_p \le t_p)$ by $P(\boldsymbol{z} \le \boldsymbol{t})$. Then

$$F_{\boldsymbol{w}_n}(\boldsymbol{t}) = P[n^{1/2}(\hat{\boldsymbol{\beta}}_{VS} - \boldsymbol{\beta}) \le \boldsymbol{t}] = \sum_{k=1}^{J} F_{\boldsymbol{w}_{kn}}(\boldsymbol{t})\pi_{kn}.$$

Hence $\hat{\boldsymbol{\beta}}_{VS}$ has a mixture distribution of the $\hat{\boldsymbol{\beta}}^C_{I_k,0}$ with probabilities π_{kn}, and $\boldsymbol{w}_n$ has a mixture distribution of the $\boldsymbol{w}_{kn}$ with probabilities π_{kn}.

125) **Theorem.** Assume $P(S \subseteq I_{min}) \to 1$ as $n \to \infty$, and let $\hat{\boldsymbol{\beta}}_{VS} = \hat{\boldsymbol{\beta}}_{I_k,0}$ with probabilities π_{kn} where $\pi_{kn} \to \pi_k$ as $n \to \infty$. Denote the positive π_k by π_j. Assume $\boldsymbol{w}_{jn} = \sqrt{n}(\hat{\boldsymbol{\beta}}^C_{I_j,0} - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{w}_j$. Then

$$\boldsymbol{w}_n = \sqrt{n}(\hat{\boldsymbol{\beta}}_{VS} - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{w}$$

where the cdf of $\boldsymbol{w}$ is $F_{\boldsymbol{w}}(\boldsymbol{t}) = \sum_j \pi_j F_{\boldsymbol{w}_j}(\boldsymbol{t})$.

126) The linear model is $\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{\eta} + \boldsymbol{e}$.

Theorem: Assume that the sample correlation matrix $\boldsymbol{R}_{\boldsymbol{u}} = \dfrac{\boldsymbol{W}^T\boldsymbol{W}}{n} \xrightarrow{P} \boldsymbol{V}^{-1}$. Let $\boldsymbol{H} = \boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{W}^T = (h_{ij})$, and assume that $\max_{i=1,...,n} h_{ii} \xrightarrow{P} 0$ as $n \to \infty$. Let $\hat{\boldsymbol{\eta}}_A$ be $\hat{\boldsymbol{\eta}}_{EN}$, $\hat{\boldsymbol{\eta}}_L$, or $\hat{\boldsymbol{\eta}}_R$. Let p be fixed.

i) OLS CLT: $\sqrt{n}(\hat{\boldsymbol{\eta}}_{OLS} - \boldsymbol{\eta}) \xrightarrow{D} N_{p-1}(\boldsymbol{0}, \sigma^2\boldsymbol{V})$.

ii) *If* $\hat{\lambda}_{1,n}/\sqrt{n} \xrightarrow{P} 0$, *then*

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_A - \boldsymbol{\eta}) \xrightarrow{D} N_{p-1}(\boldsymbol{0}, \sigma^2\boldsymbol{V}).$$

iii) *If* $\hat{\lambda}_{1,n}/\sqrt{n} \xrightarrow{P} \tau \ge 0$, $\hat{\alpha} \xrightarrow{P} \psi \in [0,1]$, *and* $\boldsymbol{s}_n \xrightarrow{P} \boldsymbol{s} = \boldsymbol{s}_{\boldsymbol{\eta}}$, *then*

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_{EN} - \boldsymbol{\eta}) \xrightarrow{D} N_{p-1}\left(-\boldsymbol{V}[(1-\psi)\tau\boldsymbol{\eta} + \psi\tau\boldsymbol{s}], \sigma^2\boldsymbol{V}\right).$$

iv) *If* $\hat{\lambda}_{1,n}/\sqrt{n} \xrightarrow{P} \tau \ge 0$, *then*

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_R - \boldsymbol{\eta}) \xrightarrow{D} N_{p-1}(-\tau\boldsymbol{V}\boldsymbol{\eta}, \sigma^2\boldsymbol{V}).$$

v) *If* $\hat{\lambda}_{1,n}/\sqrt{n} \xrightarrow{P} \tau \ge 0$ *and* $\boldsymbol{s}_n \xrightarrow{P} \boldsymbol{s} = \boldsymbol{s}_{\boldsymbol{\eta}}$, *then*

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_L - \boldsymbol{\eta}) \xrightarrow{D} N_{p-1}\left(\frac{-\tau}{2}\boldsymbol{V}\boldsymbol{s}, \sigma^2\boldsymbol{V}\right).$$

ii) and v) are the Lasso CLT, ii) and iv) are the RR CLT, and ii) and iii) are the EN CLT.

127) Let $\boldsymbol{\Sigma}_i$ be the nonsingular population covariance matrix of the ith population. Assume $n_i/n \xrightarrow{P} \pi_i$ where $0 < \pi_i < 1$ and $\sum_{i=1}^p \pi_i = 1$. Let $\overline{\boldsymbol{x}}_i$ be the sample mean of the cases from the ith population with $\boldsymbol{\mu}_i$ the ith population mean. Then $\sqrt{n_i}(\overline{\boldsymbol{x}}_i - \boldsymbol{\mu}_i) \xrightarrow{D} N_m(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$. Then $\sqrt{n}(\overline{\boldsymbol{x}}_i - \boldsymbol{\mu}_i) \xrightarrow{D} N_m\left(\boldsymbol{0}, \dfrac{\boldsymbol{\Sigma}_i}{\pi_i}\right)$ for $i = 1, .., p$.