Assuming nothing, what is $Var(\hat{\beta})$?

$$\mathrm{var}(\hat{\beta} \mid X) = \mathrm{var}\left\{(X^T X)^{-1} X^T \mathbf{Y}\right\} =$$
$$(X^T X)^{-1} X^T \mathrm{var}\left\{\mathbf{Y} \mid X\right\} X (X^T X)^{-1} =$$
$$(X^T X)^{-1} X^T V X (X^T X)^{-1}$$

- So, if one knows V, then the variance of the estimated coefficients is known.
- However, one never knows V so you have to estimate V.
- Estimating V will require some assumption about how the Y_{ij}' s are correlated (the form of the V_i' s).

$Var(\hat{\beta})$ from variance assumptions of OLS (Independence and constant variance)

$$
\underset{NxN}{V(\vec{Y})} = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \ldots & 0 \\ 0 & \sigma^2 & 0 & 0 \ldots & 0 \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ 0 & 0 & 0 & 0 \ldots & \sigma^2 \end{bmatrix}
$$

- Inference returned by standard OLS assumes, the observations are assumed to be independent and *homoskedastic* (constant variance).

Estimating $Var(\vec{Y})$ to get estimate of $Var(\hat{\beta})$ under assumptions of OLS

- To estimate σ^2, which is now the variance of the Y_{ij}' s. Note that:

$$\mathrm{var}(\mathrm{Y_{ij}} \mid \mathbf{X}_{ij}) = \mathrm{var}\left\{e_{ij} \mid \mathbf{X}_{ij}\right\} = \sigma^2$$

- We don' t know the e_{ij}' s, but we can estimate them from the residuals:

$$r_{ij} = Y_{ij} - \hat{Y}_{ij}(\vec{X}_{ij}) = Y_{ij} - \left\{\hat{\beta}_0 + \hat{\beta}_1 X_{ij1} + \ldots + \hat{\beta}_p X_{ijp}\right\}$$

OLS estimate of $Var(\hat{\beta})$

- To get a estimate of σ^2, get an average (sort of – adjusted denominator, *N-p*, reduces small sample bias) of the squared residuals:

$$\hat{\sigma}^2 = \frac{1}{N-p} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (r_{ij})^2$$

- Then, this is plugged back in V_i to get:

$\hat{V}_i = \hat{\sigma}^2 I_{n_i x n_i}$ or $\hat{V} = \hat{\sigma}^2 I_{NxN}$,

where I represents identify matrix

$$I_{4x4} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

OLS estimate of $Var(\hat{\beta})$

- Finally, plug this back into to get the estimated variance of the coefficients

$$\hat{\text{var}}(\hat{\beta} \mid X) = (X^T X)^{-1} X^T \hat{V} X (X^T X)^{-1}$$
$$= (X^T X)^{-1} X^T \hat{\sigma}^2 I X (X^T X)^{-1}$$
$$= \hat{\sigma}^2 (X^T X)^{-1} X^T X (X^T X)^{-1}$$
$$= \hat{\sigma}^2 (X^T X)^{-1}$$

$$\text{var}(\hat{\beta})_{(p+1)x(p+1)} = \begin{bmatrix} \text{var}(\hat{\beta}_0) & \text{cov}(\hat{\beta}_0, \hat{\beta}_1) & \ldots & \text{cov}(\hat{\beta}_0, \hat{\beta}_p) \\ \text{cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \ldots \\ \ldots & \ldots & \ldots & \\ \text{cov}(\hat{\beta}_0, \hat{\beta}_p) & \text{cov}(\hat{\beta}_1, \hat{\beta}_p) & \ldots & \text{var}(\hat{\beta}_p) \end{bmatrix}$$

- This is a matrix where the diagonal elements are the estimated variances of the coefficient estimates. The SE($\hat{\beta}$) are the square-roots of these estimated variances, e.g., $SE(\hat{\beta}_0) \equiv \sqrt{\hat{\text{var}}(\hat{\beta}_0)}$

What if OLS model for is wrong ?

- Ignoring the correlation, as in the standard errors returned by OLS, can give biased estimates of $\text{var}(\hat{\beta})$ (and thus biased $SE(\hat{\beta})$).

- This can lead to erroneous confidence intervals and tests.

- However, we can still use OLS if we can just repair the estimates of $\text{var}(\hat{\beta})$.

- That means getting a better (maybe even) unbiased estimate of $Var(\vec{Y})$, *so choose bigger variance-covariance model*.

Estimating *V* under bigger model

- First, we will take the case that measurements are time-structured and the same for each person ($n_i = n$) - the dental example.

- For this type of data, one can assume the most general structure about V (unstructured).

- The goal is to estimate the individual components of V_i (the v_{ijk}) using the residuals of OLS.

Form of the each unit's Variance-Covariance matrix, V_i.

- In order to estimate the *V*, we will assume that the individual covariance matrices (the V_i) are all the same: $V_1 = V_2 = \ldots = V_m$.

- That is equivalent to saying the $v_{ijk} = v_{jk}$, or the variance and covariances of equivalent observations are the same for every individual.

- That way, we will only have to estimate one V_i of in order to build *V*.

- Thus, we can call each $V_1, V_2, \ldots, V_m = V_0$.

The Variance-Covariance of Y

$$Var(\vec{Y}) = \begin{bmatrix} V_0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & V_0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & V_0 & 0 & \ldots & 0 \\ . & . & & & \ldots & \\ . & . & . & & & . \\ . & . & . & & & . \\ 0 & 0 & 0 & 0 & \ldots & V_0 \end{bmatrix}$$

Each one of these represents a matrix, e.g.

$$V_0(\vec{Y}_1)_{nxn} = \begin{bmatrix} \operatorname{var}(Y_{11}) & \operatorname{cov}(Y_{11},Y_{12}) & \ldots & \operatorname{cov}(Y_{11},Y_{1n}) \\ \operatorname{cov}(Y_{11},Y_{12}) & \operatorname{var}(Y_{12}) & \operatorname{cov}(Y_{12},Y_{13}) & \ldots \\ \ldots & \ldots & \ldots & 0 \\ \ldots & \ldots & \ldots & \operatorname{var}(Y_{1n}) \end{bmatrix}$$

Estimating the components of V_0.

- If we observed the errors, then it would be easy to estimate the v_{jk}, because, given the X_{ij}'s are treated as fixed, $v_{jk}=cov(Y_{ij}, Y_{ik})=$ $cov(e_{ij},e_{ik})=E(e_{ij}*e_{ik})-E(e_{ij})E(e_{ik})=E(e_{ij}*e_{ik})$.
- So estimate cov with average:

$$\hat{v}_{jk} = \frac{1}{m}\sum_{i=1}^{m} e_{ij}e_{ik} \ .$$

- We don't observe the errors, but we can still estimate the covariance of any two measurements on the same person using the residuals.

$$\hat{v}_{jk} = \frac{1}{m}\sum_{i=1}^{m} r_{ij}r_{ik}, r_{ij} = Y_{ij} - \hat{Y}_{ij}(\vec{X}_{ij}), \text{ where } \hat{Y}_{ij}(\vec{X}_{ij}) \text{ is the predicted value of outcome}$$

Plugging Into the Big $V(\vec{Y})$

- Now that one has derived the estimates of each unique variance and covariance among the set of measurements on each unit, then just plug these into to get:

$$\hat{V}_0(\vec{Y}_i)_{nxn} = \begin{bmatrix} a\tilde{v}e(r_{i1}^2) & a\tilde{v}e(r_{i1} * r_{i2}) & \ldots & a\tilde{v}e(r_{i1} * r_{in}) \\ a\tilde{v}e(r_{i1} * r_{i2}) & a\tilde{v}e(r_{i2}^2) & \ldots & a\tilde{v}e(r_{i2} * r_{in}) \\ \ldots & \ldots & \ldots & \ldots \\ a\tilde{v}e(r_{i1} * r_{in}) & a\tilde{v}e(r_{i2} * r_{in}) & \ldots & a\tilde{v}e(r_{in}^2) \end{bmatrix}, \text{ where } a\tilde{v}e \text{ represents adusted for small sample}$$

$$\hat{V}(\vec{Y})_{NxN} = \begin{bmatrix} \hat{V}_{0,nxn} & 0_{nxn} & \ldots & 0_{nxn} \\ 0_{nxn} & \hat{V}_{0,nxn} & \ldots & \ldots \\ \ldots & 0_{nxn} & \ldots & \ldots \\ \ldots & \ldots & \ldots & \hat{V}_{0,nxn} \end{bmatrix}, \text{lumped together to form on big variance-cov matrix}$$

Good start, but…

- The LS solution is still (theoretically) unbiased even if there is residual correlation among repeated measurements on a subject (and design is balanced).

- However, the inference is not unbiased, so must account for correlation when calculating $SE(\hat{\beta})$

- One solution for regularly measured data (e.g., all subjects have same measurements at same times) is to use residuals and estimate the variance-covariance matrix of the data assuming each subject has the same matrix.

- However, can even get more flexible (robust SE's).

Robust (no Model) SE's

- Instead of assuming that very subject has the same V_i, one can estimate each subjects V-C matrix separately.

- Thus, the estimate of the variance of the first measurement on subject i is:

$$\hat{v}_{i11} = \hat{\mathrm{var}}(Y_{i1}) = r_{i1}^2 = (Y_{i1} - \hat{Y}_{i1})^2$$

- The correlation between the 1st and 2nd measurement on the ith person is estimated as: $$\hat{v}_{i12} = \hat{\mathrm{cov}}(Y_{i1}, Y_{i2}) = r_{i1} r_{i2} = (Y_{i1} - \hat{Y}_{i1})(Y_{i2} - \hat{Y}_{i2})$$

Robust SE's, cont.

- This procedure is repeated until all the variances and covariances are estimated for each subject (again, we always assume no between subject correlation).
- These are terrible estimators of the variances and covariances.
- However, do we care? No we don't!!!
- Because, they still result, under certain assumptions, in a consistent estimator of $SE(\hat{\beta})$
- The estimate, $\hat{\mathrm{var}}(\hat{\beta})$, sums over the, $\hat{V}$ which means we gain precision by averaging over the individual estimates of the $\hat{v}_{ijk}$

Get robust inference via Influence Curve (IC)

- Estimators of coefficients in these regression models are so-called <u>asymptotically linear</u>,

$$\left(\beta-\hat{\beta}\right)\approx\frac{1}{m}\sum_{i=1}^{m}IC(Y_i,X_i;\beta)\text{ so}$$

$$\operatorname{var}(\hat{\beta})=\frac{\operatorname{var}(IC)}{m}$$

- In case of OLS, simple model $E(Y_{ij}|X_{ij})=\alpha+\beta X_{ij}$

$$IC(Y_i,X_i;\beta)=\frac{1}{E\sum_j X_{ij}^2}\sum_j X_{ij}\{Y_{ij}-(\alpha+\beta X_{ij})\}$$

- Also, *clustered* bootstrap