# BIOS 7345: Advanced Regression Analysis I

**Andrew J. Spieker, Ph.D.**

Assistant Professor of Biostatistics
Vanderbilt University Medical Center

Set 14: Sandwich and bootstrap methods for GLMs

Version: 09/17/2023

# Table of Contents

**Recall**:

- A GLM involves specifying a parametric form for $Y$ (given $\mathbf{X}$) that can be factored into a nice exponential form.
- The score equations take the form:

$$\mathbf{D}^T\mathbf{V}^{-1}(\mathbf{y}-\boldsymbol{\mu})/\phi = \mathbf{0}.$$

- From the prior set of notes, we saw that for a GLM,

$$\mathcal{I}(\boldsymbol{\beta},\phi) = \begin{bmatrix} \mathbf{D}^T\mathbf{V}^{-1}\mathbf{D}/\phi & \mathbf{0}^T \\ \mathbf{0} & \cdots \end{bmatrix}$$

- The block-diagonal structure of the information tells us we need not propagate uncertainty in estimation of $\phi$ in estimating variance of $\widehat{\boldsymbol{\beta}}$.

# Variance based on likelihood theory

**Asymptotic distribution**:

- Likelihood theory (suitable regularity conditions):

$$\widehat{\boldsymbol{\beta}} \dot{\sim} \mathcal{N}\Big(\boldsymbol{\beta}, \phi(\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}\Big)\,.$$

- To estimate $\text{Cov}[\widehat{\boldsymbol{\beta}}]$, we could be in one of two cases:
  1. $\phi = 1$, in which case $\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = (\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}$.
  2. Otherwise, we require a consistent estimate of $\phi$. This one will do:

$$\widehat{\phi} = \frac{1}{N-K}\sum_{i=1}^{N}\frac{(y_i - \widehat{\mu}_i)^2}{V(\widehat{\mu}_i)},$$

and we have that $\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = \widehat{\phi}(\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}$

# Table of Contents

# Variance based on theory of estimating equations

**Ideas**:

- We don't need to solve for $\phi$ to solve the score equations for $\boldsymbol{\beta}$. Namely, we can simply solve the following estimating equations:

$$\mathbf{D}^T\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}.$$

- If you believe the mean model is correct, these are referred to as *unbiased* estimating equations.
    - $\mathrm{E}[\mathbb{G}_N(\boldsymbol{\beta}; \mathbf{X}, \mathbf{y})] = \mathbf{0}$, where the expectation is either over $\mathbf{y}|\mathbf{X}$ or $(\mathbf{X}, \mathbf{y})$.
- What if the likelihood is not correctly specified?
    - For instance, what if the mean model is not correct?
    - For instance, what if the mean-variance relationship is not correct?
    - For instance, what if the third (or higher) moment is not correct?
- Can we derive an expression for $\mathrm{Cov}[\widehat{\boldsymbol{\beta}}]$ based on the theory of estimating equations rather than likelihood theory?
    - We did something similar for OLS; we again must assume $\mathbf{X}$ is random.

**Notation**:

- $\widehat{\boldsymbol{\beta}}_N$: solution to estimating equations.
- $\boldsymbol{\beta}_0$: the true, unknown value to be estimated.
  - If the mean model is not correctly specified, then $\boldsymbol{\beta}_0$ can be understood as "the quantity for which $\widehat{\boldsymbol{\beta}}_N$ is consistent."
- $\mathbb{G}_N(\boldsymbol{\beta}; \mathbf{X}, \mathbf{y}) = \mathbf{D}^T \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \sum_{i=1}^{N} \mathbf{G}(\boldsymbol{\beta}; \mathbf{x}_i, Y_i)$.
- $\mathbf{A}(\boldsymbol{\beta}) = \mathrm{E}\left[-\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{G}(\boldsymbol{\theta}; \mathbf{x}, Y)\Big|_{\boldsymbol{\theta}=\boldsymbol{\beta}}\right]$
- $\mathbf{B}(\boldsymbol{\beta}) = \mathrm{E}[\mathbf{G}(\boldsymbol{\beta}; \mathbf{x}, Y)\mathbf{G}(\boldsymbol{\beta}; \mathbf{x}, Y)^T]$

# Variance based on theory of estimating equations

**Taylor expansion**:

- Because $\widehat{\boldsymbol{\beta}}_N$ solves the estimating equations, it follows that:

$$\mathbf{0} = \frac{1}{N}\mathbb{G}_N(\widehat{\boldsymbol{\beta}}_N; \mathbf{X}, \mathbf{y})$$

- If $\mathbf{G}$ is analytic (has a Taylor series), we can expand about $\boldsymbol{\beta}_0$:

$$\begin{aligned} \mathbf{0} &\approx \frac{1}{N}\mathbb{G}_N(\boldsymbol{\beta}_0; \mathbf{X}, \mathbf{y}) + \frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{N}\mathbb{G}_N(\boldsymbol{\beta}; \mathbf{X}, \mathbf{y})\right]\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) \\ &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{G}(\boldsymbol{\beta}_0; \mathbf{x}_i, Y_i) + \left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{G}(\boldsymbol{\beta}; \mathbf{x}_i, Y_i)\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right](\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) \end{aligned}$$

**Rearrangement**:

- Assume that $\sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{G}(\boldsymbol{\beta}_0; \mathbf{x}_i, Y_i)$ is invertible.
- Rearranging the equation on the prior slide (and leaving the details surrounding the regularity conditions on the remainder term of the Taylor expansion to a more theory-oriented course):

$$(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) \approx \left[ -\frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{G}(\boldsymbol{\beta}; \mathbf{x}_i, Y_i) \bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right]^{-1} \left[ \frac{1}{N} \sum_{i=1}^{N} \mathbf{G}(\boldsymbol{\beta}_0; \mathbf{x}_i, Y_i) \right]$$

# Variance based on theory of estimating equations

**Invoking asymptotics**:

- Multiplying both sides by $\sqrt{N}$, we then have:

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) \approx \left[ -\frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{G}(\boldsymbol{\beta}; \mathbf{x}_i, Y_i) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right]^{-1} \left[ \frac{\sqrt{N}}{N} \sum_{i=1}^{N} \mathbf{G}(\boldsymbol{\beta}_0; \mathbf{x}_i, Y_i) \right]$$

- By the weak law of large numbers,

$$\left[ -\frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{G}(\boldsymbol{\beta}; \mathbf{x}_i, Y_i) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right] \longrightarrow_p \mathbf{A}(\boldsymbol{\beta}_0)$$

- By the central limit theorem*,

$$\frac{\sqrt{N}}{N} \sum_{i=1}^{N} \mathbf{G}(\boldsymbol{\beta}_0; \mathbf{x}_i, Y_i) = \left[ \sqrt{N} \left( \frac{1}{N} \sum_{i=1}^{N} \mathbf{G}(\boldsymbol{\beta}_0; \mathbf{x}_i, Y_i) - \mathbf{0} \right) \right] \longrightarrow_d \mathcal{N}(\mathbf{0}, \mathbf{B}(\boldsymbol{\beta}_0)).$$

**Point of nuance**:

- There is a nuanced point here that's easy to miss.
- In the previous slide, we assumed $\mathrm{E}[\mathbf{G}(\boldsymbol{\beta}_0; \mathbf{x}, \mathbf{y})] = \mathbf{0}$ in this step:

$$\left[\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{G}(\boldsymbol{\beta}_0; \mathbf{x}_i, Y_i) - \mathbf{0}\right)\right] \longrightarrow_d \mathcal{N}(\mathbf{0}, \mathbf{B}(\boldsymbol{\beta}_0)).$$

- We're in the middle of deriving the asymptotic distribution for $\widehat{\boldsymbol{\beta}}_N$ in the setting that the mean model may not be correct. Is this fair?
- When we refer to $\boldsymbol{\beta}_0$ as the "true value of the parameter," it is more accurate to think of it as the value for which the implicit solution to the estimating equations is consistent.
- Since we set the estimating equations to zero to solve for $\widehat{\boldsymbol{\beta}}_N$, the expectation will be zero at $\boldsymbol{\beta}_0$ even if the mean model is not correct (that is, if the estimating equations are not unbiased).

# Variance based on theory of estimating equations

**More asymptotics**:

- Returning to the derivation, it follows from Slutsky's theorem that

$$\widehat{\boldsymbol{\beta}}_N \dot{\sim} \mathcal{N}\left(\boldsymbol{\beta}_0, \frac{1}{N}[\mathbf{A}(\boldsymbol{\beta}_0)]^{-1}\mathbf{B}(\beta_0)[\mathbf{A}(\boldsymbol{\beta}_0)]^{-1}\right)$$

- To estimate $\text{Cov}[\widehat{\boldsymbol{\beta}}]$, we can plug in estimators of $\mathbf{A}(\boldsymbol{\beta}_0)$ and $\mathbf{B}(\boldsymbol{\beta}_0)$ (such estimators are known as *sandwich* estimators).

**Plug-in estimators**:

- How do we estimate $\mathbf{A}(\boldsymbol{\beta}_0)$? Recall that:

$$\mathbb{A}_N^{\text{obs}}(\boldsymbol{\beta}) \quad = \quad -\frac{\partial}{\partial \boldsymbol{\beta}} \mathbb{G}_N(\boldsymbol{\beta}; \mathbf{X}, \mathbf{y}).$$

- By the weak law of large numbers,

$$\frac{1}{N} \mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}}) \quad = \quad -\frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{G}(\widehat{\boldsymbol{\beta}}, \mathbf{x}_i, Y_i) \longrightarrow_p \mathbf{A}(\boldsymbol{\beta}_0).$$

- This statement is valid even if the mean model is not correct.

**Plug-in estimators**:

- How do we estimate $\mathbf{B}(\boldsymbol{\beta}_0)$? Let

$$\begin{aligned}\mathbb{B}_N^{\text{obs}}(\boldsymbol{\beta}) &= \sum_{i=1}^{N} \mathbf{G}(\boldsymbol{\beta}; \mathbf{x}_i, y_i)\mathbf{G}(\boldsymbol{\beta}; \mathbf{x}_i, y_i)^T \\ &= \mathbf{D}^T(\boldsymbol{\beta})\mathbf{V}^{-1}(\boldsymbol{\beta})\text{diag}(y_i - \mu_i(\boldsymbol{\beta}))^2\mathbf{V}^{-1}(\boldsymbol{\beta})\mathbf{D}(\boldsymbol{\beta}).\end{aligned}$$

- By the weak law of large numbers,

$$\frac{1}{N}\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{G}(\widehat{\boldsymbol{\beta}}; \mathbf{x}_i, y_i)\mathbf{G}(\widehat{\boldsymbol{\beta}}; \mathbf{x}_i, y_i)^T \longrightarrow_p \mathbf{B}(\boldsymbol{\beta}_0).$$

- This statement is valid even if the mean-variance relationship is not correctly specified.

# Variance based on theory of estimating equations

**Plug-in estimators**:

- We can therefore estimate $\text{Cov}[\widehat{\boldsymbol{\beta}}]$ as follows:

$$\begin{aligned}\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] &= \frac{1}{N}\left(\frac{1}{N}\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})\right)^{-1}\left(\frac{1}{N}\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})\right)\left(\frac{1}{N}\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})\right)^{-1} \\ &= \left(\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})\right)^{-1}\left(\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})\right)\left(\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})\right)^{-1},\end{aligned}$$

  where $\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})$ and $\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})$ are presented on the prior slides.
- This is *one* version of the sandwich estimator, and it is asymptotically valid even if neither aspect of the GLM (meaning the mean model and the mean-variance relationship) is correctly specified.

**Correct mean model**:

- If the mean model is correct, it is straightforward to show:

$$\begin{aligned}\mathbf{A}(\boldsymbol{\beta}_0) &= \mathrm{E}_{\mathbf{X}}[\mathbf{D}^T(\boldsymbol{\beta}_0)\mathbf{V}^{-1}(\boldsymbol{\beta}_0)\mathbf{D}^T(\boldsymbol{\beta}_0)] \\ &= \mathrm{E}_{\mathbf{X}}\Big[\mathbf{X}^T\mathbf{W}(\boldsymbol{\beta}_0)\mathbf{X}\Big].\end{aligned}$$

- If we believe the mean model to hold, it therefore seems sensible to estimate $\mathbf{A}(\boldsymbol{\beta}_0)$ based on $\mathbb{A}_N(\widehat{\boldsymbol{\beta}})$ rather than $\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})$. By the weak law of large numbers (if the mean model is correct):

$$\frac{1}{N}\mathbb{A}_N(\widehat{\boldsymbol{\beta}}) = \frac{1}{N}\mathbf{X}^T\mathbf{W}(\widehat{\boldsymbol{\beta}})\mathbf{X} \longrightarrow_p \mathbf{A}(\boldsymbol{\beta}_0).$$

- Keep in mind: sometimes $\mathbb{A}_N(\boldsymbol{\beta})$ and $\mathbb{A}_N^{\text{obs}}(\boldsymbol{\beta})$ are the same. When?

# VARIANCE BASED ON THEORY OF ESTIMATING EQUATIONS

**Plug-in estimators**: When the mean model is correct

- We can estimate $\text{Cov}[\widehat{\boldsymbol{\beta}}]$ as follows:

$$\begin{aligned}\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] &= \frac{1}{N}\left(\frac{1}{N}\mathbb{A}_N(\widehat{\boldsymbol{\beta}})\right)^{-1}\left(\frac{1}{N}\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})\right)\left(\frac{1}{N}\mathbb{A}_N(\widehat{\boldsymbol{\beta}})\right)^{-1} \\ &= \left(\mathbb{A}_N(\widehat{\boldsymbol{\beta}})\right)^{-1}\left(\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})\right)\left(\mathbb{A}_N(\widehat{\boldsymbol{\beta}})\right)^{-1}\end{aligned}$$

- This is another version of the sandwich estimator; it is asymptotically valid if the mean-variance relationship of the GLM is misspecified.
- Under the canonical link, this will be equivalent to the previous sandwich and validity will not depend upon mean model being correct.
- Under a non-canonical link, validity depends upon the mean model being correctly specified.

# Variance based on theory of estimating equations

**Correct mean model and mean-variance relationship**:

- If both the mean model and mean-variance relationship are correct, it is straightforward to show:

$$\mathbf{B}(\boldsymbol{\beta}_0) \;\; = \;\; \phi\mathbf{A}(\boldsymbol{\beta}_0)$$

- If we believe the mean model to be correct and the mean-variance relationship to hold, it therefore seems sensible to estimate $\mathbf{B}(\boldsymbol{\beta}_0)$ based on $\mathbb{B}_N(\widehat{\boldsymbol{\beta}})$ rather than $\mathbb{B}_N^{\text{obs}}(\boldsymbol{\beta})$:

$$\frac{1}{N}\mathbb{B}_N(\widehat{\boldsymbol{\beta}}) = \frac{1}{N}\widehat{\phi}\mathbb{A}_n(\widehat{\boldsymbol{\beta}}) \longrightarrow_p \mathbf{B}(\boldsymbol{\beta}_0)$$

**Model-based estimation**:

- Take note that if we believe the mean model and the mean-variance relationship are correct, there is cancellation: $\mathbb{B}_N(\widehat{\boldsymbol{\beta}}) = \phi \mathbb{A}_N(\widehat{\boldsymbol{\beta}})$.
- The variance estimator based on the assumption of both a correct mean model and mean-variance relationship collapses as follows:

$$\begin{aligned}\widehat{\mathrm{Cov}}[\widehat{\boldsymbol{\beta}}] &= (\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}\mathbb{B}_N(\widehat{\boldsymbol{\beta}})(\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1} \\ &= \widehat{\phi}(\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}.\end{aligned}$$

which is exactly the formula based on the Fisher information!

**Model-based estimation**:

- We previously saw that the solution to the GLM is determined by a user-specified mean model and mean-variance relationship. We have just argued the same for an asymptotically valid variance estimator.
- We are working within the theory of estimating equations—*not* likelihood theory—so what we have effectively just argued is that the likelihood-based approach will be valid so long as the mean model and mean-variance relationship are correctly specified.
- Though we often use likelihood language to describe a GLM, we don't rely on the third (and higher) moments implied by that likelihood.
- This is not your first exposure to the concept described on the previous slide; we already have seen this in OLS (which, as we also know, is a specific example of a GLM).

**Example**: Derivation of variance for OLS

- Consider OLS linear regression, which be can be thought of:
  - Parametrically: A normal GLM with the canonical link.
  - Semi-parametrically: A GLM based on the mean model $\text{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$ and a working mean-variance relationship $\mathbf{V} = \mathbf{I}$.
- Let's use our findings to derive the sandwich variance estimator:
  - $\mathbb{A}_N^{\text{obs}}(\boldsymbol{\beta}) = \mathbb{A}_N(\boldsymbol{\beta}) = \mathbf{X}^T\mathbf{X}$.
  - $\mathbb{B}_N^{\text{obs}}(\boldsymbol{\beta}) = \mathbf{X}^T\text{diag}(y_i - \mu_i(\boldsymbol{\beta}))^2\mathbf{X}$.
  - Sandwich: $\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\text{diag}(y_i - \mathbf{x}_i^T\widehat{\boldsymbol{\beta}})^2\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$
- If we believe the mean model and the mean-variance relationship, we would replace the meat (or cheese, or veggies) of the sandwich with $\widehat{\phi}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}) = \widehat{\sigma}^2(\mathbf{X}^T\mathbf{I}\mathbf{X})$:

$$\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = \widehat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}.$$

- We recognize this formula!

**Example**: Derivation of variance for WLS

- Consider WLS linear regression based on $\text{Var}[Y|\mathbf{X}] \propto \mathbf{V}(\mathbf{X})$, which be can be thought of:
  - Parametrically: A normal GLM with the canonical link, and a dispersion parameter $\phi_i$ that depends upon $\mathbf{X}_i$.
  - Semi-parametrically: A GLM based on the mean model $\text{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$ and a working mean-variance relationship, $\mathbf{V} = \mathbf{V}(\mathbf{X})$.
- I leave it to you to verify the following sandwich variance estimator:

$$\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\text{diag}(y_i - \mathbf{x}_i^T\widehat{\boldsymbol{\beta}})^2\mathbf{W}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1},$$

  where $\mathbf{W} = \mathbf{V}^{-1}(\mathbf{X})$.
- If we believe the mean model and the mean-variance relationship, we replace the meat/cheese/veggies with $\widehat{\phi}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}) = \widehat{\phi}(\mathbf{X}^T\mathbf{W}\mathbf{X})$:

$$\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = \left(\frac{1}{N-K}\sum_{i=1}^{N} w_i(y_i - \mathbf{x}_i^T\widehat{\boldsymbol{\beta}})^2\right)(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}.$$

# Variance based on theory of estimating equations

**Variance formulas so far**: Canonical link

- $\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = \widehat{\phi}(\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}$.
  - Relies on correct mean model.
  - Relies on correct mean-variance relationship.
  - Estimation of $\widehat{\phi}$ not necessary if there is no nuisance parameter.
- $\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = (\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})(\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}$.
  - $\mathbb{A}_N(\widehat{\boldsymbol{\beta}}) = \mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})$.
  - Does not rely on correct mean model.
  - Does not rely on correct mean-variance relationship.

**Variance formulas so far**: Non-canonical link

- $\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = \widehat{\phi}(\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}$.
  - Relies on correct mean model.
  - Relies on correct mean-variance relationship.
  - Estimation of $\widehat{\phi}$ not necessary if there is no nuisance parameter.
- $\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = (\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})(\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}$.
  - $\mathbb{A}_N(\widehat{\boldsymbol{\beta}}) \neq \mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})$.
  - Relies on correct mean model.
  - Does not rely on correct mean-variance relationship.
- $\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = (\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}}))^{-1}\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})(\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}}))^{-1}$.
  - $\mathbb{A}_N(\widehat{\boldsymbol{\beta}}) \neq \mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})$.
  - Does not rely on correct mean model.
  - Does not rely on correct mean-variance relationship.

**Other considerations**:

- There are other versions of the sandwich variance; most are simply modifications to the ones we've already discussed.
    - In the language of the `sandwich()` package in R, the ones we've discussed fall under the category of `HC0`.
- Some re-scale by a factor of $N/(N-K)$ to add a correction for degrees of freedom.
- Some studentize the residuals of $\mathbb{B}_N(\widehat{\boldsymbol{\beta}})$.
- In large samples, discrepancies across these versions are comparatively minor.

# Variance based on theory of estimating equations

**Example**: Normal distribution (log link)

- Revisiting a prior example, suppose our GLM is based on:
  - $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2)$.
  - $g(\cdot)$ given by the log link (i.e., $g(\boldsymbol{\mu}) = \log(\boldsymbol{\mu})$).
- We derived the following estimating equations for $\boldsymbol{\beta}$:

$$\mathbf{X}^T \text{diag}(\exp(\mathbf{x}_i^T \boldsymbol{\beta}))(\mathbf{y} - \exp(\mathbf{X}\boldsymbol{\beta})) = \mathbf{0}.$$

- We determined the following:

$$\begin{aligned} \mathbb{A}_N(\widehat{\boldsymbol{\beta}}) &= \mathbf{X}^T \text{diag}(\exp(\mathbf{x}_i^T \widehat{\boldsymbol{\beta}}))^2 \mathbf{X}. \\ \mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}}) &= \mathbf{X}^T \text{diag}(\exp(\mathbf{x}_i^T \widehat{\boldsymbol{\beta}}))^2 \mathbf{X} - \mathbf{X}^T \text{diag}(\exp(\mathbf{x}_i^T \widehat{\boldsymbol{\beta}})(y_i - \exp(\mathbf{x}_i^T \widehat{\boldsymbol{\beta}})))\mathbf{X}. \end{aligned}$$

- I leave it for you to show that:

$$\mathbb{B}_N(\boldsymbol{\beta}) = \mathbf{X}^T \text{diag}(\exp(\mathbf{x}_i^T \widehat{\boldsymbol{\beta}})(y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}))^2 \mathbf{X}.$$

# Variance based on theory of estimating equations

**Example**: Normal distribution (log link)

- We continue from our previous fit of the (simulated) data from this GLM.
- Consider the following variance estimators:
  1. A "model-based" estimator.
  2. An estimator that allows a misspecified mean-variance relationship.
  3. An estimator that allows a misspecified mean model and mean-variance relationship.

**Example**: Normal distribution (log link)

```
## Store final iteration
betahat <- betaj

## Linear predictor
etahat <- c(X %*% betahat)

## Estimating function
Gn <- t(X) %*% diag(exp(etahat)) %*% (y - exp(etahat))

## An
An <- t(X) %*% diag(exp(etahat)^2) %*% X

## W
W <- diag(exp(etahat))

## AnObs
AnObs <- An - t(X) %*% W %*% diag(y - exp(etahat)) %*% X

## Bn
Bn <- t(X) %*% W %*% diag((y - exp(etahat)))^2 %*% W %*% X
```

**Example**: Normal distribution (log link)

```
V1 <- phi * solve(An)
V2 <- solve(An) %*% Bn %*% solve(An)
V3 <- solve(AnObs) %*% Bn %*% solve(AnObs)
V2star <- V2 * (n)/(n - 2)
V3star <- V3 * (n)/(n - 2)

 >     sqrt(diag(V1))
[1] 0.1804808 0.1373220

>     sqrt(diag(V2))
[1] 0.1649462 0.1077723

>     sqrt(diag(V3))
[1] 0.1650517 0.1079106

>     sqrt(diag(V2star))
[1] 0.1683475 0.1099946

>     sqrt(diag(V3star))
[1] 0.1684552 0.1101358
```

# Variance based on theory of estimating equations

**Example**: Normal distribution (log link)

```
zz <- glm(y ~ X[,2], start = c(1,1), family = gaussian(link = "log"))
V4 <- sandwich(zz)

>     sqrt(diag(V4))
(Intercept)      X[, 2]
0.1649462   0.1077723
```

- And just like that, we've taken some of the magic away from the `sandwich()` function in R! What have we learned about which version of the sandwich is being used?

**Design matrix**: Fixed vs. random

- The sandwich variance estimator(s) were derived based on the theory of estimating equations under the premise that sampling is from the joint distribution $(\mathbf{X}, \mathbf{y})$.
- If the mean model is correct, the validity of the sandwich still holds even if $\mathbf{X}$ is fixed.
  - The argument for this lies in showing that when the mean model is correct, $\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})$, $\mathbb{A}_N(\widehat{\boldsymbol{\beta}})$, and $\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})$ are all consistent for the same (respective) quantities for which they are consistent when $\mathbf{X}$ is random.
  - This will *not* be the case when the mean model is misspecified.
- We didn't have to care so much about this when we were thinking of GLMs through the likelihood framework, in which we were always assuming the model to be correctly specified.

**Summarizing what we know so far**:

| **X** | Link | $g^{-1}(\cdot)$ | **V** | $\widehat{\phi}\mathbb{A}_N^{-1}$ | $(\mathbb{A}_N^{-1})\mathbb{B}_N^{\text{obs}}(\mathbb{A}_N)^{-1}$ | $(\mathbb{A}_N^{\text{obs}})^{-1}\mathbb{B}_N^{\text{obs}}(\mathbb{A}_N^{\text{obs}})^{-1}$ |
|---|---|---|---|---|---|---|
| Fixed | C | ✓ | ✓ | ✓ | ✓ | ✓ |
| Fixed | C | ✓ | ✗ | ✗ | ✓ | ✓ |
| Fixed | C | ✗ | ✓ | ✗ | ✗ | ✗ |
| Fixed | C | ✗ | ✗ | ✗ | ✗ | ✗ |
| Fixed | NC | ✓ | ✓ | ✓ | ✓ | ✓ |
| Fixed | NC | ✓ | ✗ | ✗ | ✓ | ✓ |
| Fixed | NC | ✗ | ✓ | ✗ | ✗ | ✗ |
| Fixed | NC | ✗ | ✗ | ✗ | ✗ | ✗ |
| Random | C | ✓ | ✓ | ✓ | ✓ | ✓ |
| Random | C | ✓ | ✗ | ✗ | ✓ | ✓ |
| Random | C | ✗ | ✓ | ✗ | ✓ | ✓ |
| Random | C | ✗ | ✗ | ✗ | ✓ | ✓ |
| Random | NC | ✓ | ✓ | ✓ | ✓ | ✓ |
| Random | NC | ✓ | ✗ | ✗ | ✓ | ✓ |
| Random | NC | ✗ | ✓ | ✗ | ✗ | ✓ |
| Random | NC | ✗ | ✗ | ✗ | ✗ | ✓ |

C: Canonical; NC: Non-canonical
$g^{-1}(\cdot)$ mean model correctly specified?
**V** mean-variance correctly specified?

# Table of Contents

**Main ideas**:

- Let $F$ denote CDF for $(\mathbf{X}, Y)$ or $(Y|\mathbf{X})$, depending on context; let $\mathbb{F}_N$ denote empirical CDF based on $N$ observations.
  - $\boldsymbol{\beta} = T(F)$, and hence $\widehat{\boldsymbol{\beta}} = T(\mathbb{F}_N)$.
  - Absent parametric form, $\mathbb{F}_N$ is our best estimate of $F$.
- Repeat-sampling from $\mathbb{F}_N$ with replacement gives information on distribution of $\widehat{\boldsymbol{\beta}}^* = T(\mathbb{F}_N^*)$; asterisk denotes fixed $\mathbb{F}_N$.
- Let $\{\widehat{\boldsymbol{\beta}}_b^*\}_{b=1}^B$ denote the (bootstrap) samples.
- Note two layers of variation:
  - How well $\mathbb{F}_N$ approximates $F$.
    - ⋆ Glivencko-Cantelli: $\sup_{t\in[0,1]} |F(t) - \mathbb{F}_N(t)| \longrightarrow_{\text{a.s.}} 0$ as $N \nearrow \infty$.
  - How well $\{\widehat{\boldsymbol{\beta}}_b^*\}_{b=1}^B$ approximates $T(\mathbb{F}_N^*)$
    - ⋆ Better as $B \nearrow \infty$.
- Which source of variation can we control once given the data?

**Estimator-attributed bias**:

- Let $\widehat{\boldsymbol{\beta}}_b^* = T(\mathbb{F}_{N:b}^*)$ denote estimate based on $b^{\text{th}}$ bootstrap sample. We may estimate bias as follows:

$$\begin{aligned}\widehat{\text{Bias}} &= \frac{1}{B}\sum_{b=1}^{B}(T(\mathbb{F}_{N:b}^*) - T(\mathbb{F}_N)) \\ &= \frac{1}{B}\sum_{b=1}^{B}\widehat{\boldsymbol{\beta}}_b^* - \widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}} \approx \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta},\end{aligned}$$

  where $\widehat{\boldsymbol{\beta}}^* = \frac{1}{B}\sum_{b=1}^{B}\widehat{\boldsymbol{\beta}}_b^*$.
- Correction won't catch external sources of bias; be warned.

**Covariance**:

- We may estimate the covariance as well:

$$\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\boldsymbol{\beta}}_b^* - \widehat{\boldsymbol{\beta}}^*)(\widehat{\boldsymbol{\beta}}_b^* - \widehat{\boldsymbol{\beta}}^*)^T$$

- For the $k^{\text{th}}$ coefficient, we have:

$$\widehat{v}_k \quad = \quad \widehat{\text{Var}}[\widehat{\beta}_k] = \frac{1}{B-1}\sum_{b=1}^{B}([\widehat{\boldsymbol{\beta}}_b^*]_k - \widehat{\beta}_k^*)^2$$

**Confidence intervals**: Normal approximation (bias-correction)

- Symmetric $(1-\alpha)$ CI:

$$(\widehat{\beta}_k - \widehat{\text{Bias}}_k) \pm \sqrt{\widehat{v}_k} z_{1-\alpha/2}.$$

- Assumptions:
  - $\widehat{\beta}_k - \beta_k \dot{\sim} \mathcal{N}(\text{Bias}_k, \sigma^2)$, which is symmetric and pivotal.
  - $\widehat{\text{Bias}}_k$ and $\widehat{v}_k$ are good estimates of $\text{Bias}_k$ and $\sigma^2$.
- Good for cases where $N$ is large enough that normal approximation holds, but no known theoretical formula for asymptotic variance.
- Can use QQ-plots to evaluate departures from normality.

**Confidence intervals**: Pivot based

- Let $\widehat{\beta}^*_{k(p)}$ denote $p^{\text{th}}$ quantile of $k^{\text{th}}$ coefficient of $\{\widehat{\beta}^*_b\}_{b=1}^B$.
- Behavior of $\beta_k - \widehat{\beta}_k$ approximately that of $\widehat{\beta}_k - \widehat{\beta}^*_k$:

$$\begin{aligned} 0.95 &\approx \mathrm{P}\left(\widehat{\beta}^*_{k(\alpha/2)} \leq \widehat{\beta}^*_k \leq \widehat{\beta}^*_{k(1-\alpha/2)}\right) \\ &= \mathrm{P}\left(\widehat{\beta}_k - \widehat{\beta}^*_{k(1-\alpha/2)} \leq \widehat{\beta}_k - \widehat{\beta}^*_k \leq \widehat{\beta}_k - \widehat{\beta}^*_{k(\alpha/2)}\right) \\ &\approx \mathrm{P}\left(\widehat{\beta}_k - \widehat{\beta}^*_{k(1-\alpha/2)} \leq \beta_k - \widehat{\beta}_k \leq \widehat{\beta}_k - \widehat{\beta}^*_{k(\alpha/2)}\right) \\ &= \mathrm{P}\left(2\widehat{\beta}_k - \widehat{\beta}^*_{k(1-\alpha/2)} \leq \beta_k \leq 2\widehat{\beta}_k - \widehat{\beta}^*_{k(\alpha/2)}\right) \end{aligned}$$

- Assumptions:
  - $\widehat{\beta}_k - \beta_k$ asymptotically pivotal (not necessarily symmetric).

**Confidence intervals**: Quantile-based

- Let $\widehat{\beta}^*_{k(p)}$ denote $p^{\text{th}}$ quantile of $k^{\text{th}}$ coefficient of $\{\widehat{\beta}^*_b\}_{b=1}^B$.
- One can form a $100(1-\alpha)\%$ CI as:

$$[\widehat{\beta}^*_{k(\alpha/2)}, \widehat{\beta}^*_{k(1-\alpha/2)}].$$

- Assumptions:
  - There is a monotone $h(\cdot)$ for which the distribution of $h(\widehat{\beta}^*_k)$ is symmetric, and that $h(\widehat{\beta}^*_k)$ is pivotal.
  - $h(\widehat{\beta}_k)$ is unbiased.

**Linear regression**: Bootstrap procedures

- The following three slides outline reasonable bootstrap procedures for linear regression; all but one will generalize to GLMs.

**Linear regression**: Random design

- Re-sample pairs $(\mathbf{x}_i^*, y_i^*)$ from existing observations $\{\mathbf{x}_i, y_i\}_{i=1}^N$ with replacement.
- Estimate $\widehat{\boldsymbol{\beta}}_b^*$ for $b = 1, \ldots, B$; form estimates/confidence intervals of your choosing from prior methods.
- Design changes with each sample.
- Consistent with an observational study with random sampling irrespective of exposure/outcome.
- Consistent with fully/purely randomized experiment (like a coin toss).

**Linear regression**: Fixed design, correct mean model, homoscedasticity

- Fit model $E[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$ and extract residuals $\{\widehat{\epsilon_i}\}_{i=1}^{N}$.
- Stratify unconditional bootstrap procedure by subgroups defined by $\mathbf{X}$.
- Estimate $\widehat{\boldsymbol{\beta}}_b^*$ for $b = 1, \ldots, B$; form estimates/confidence intervals of your choosing from prior methods.
- Allows heteroscedasticity; allows mean model misspecification.
- Example: designed experiment with a small number of large groups.

**Linear regression**: Fixed design, correct mean model, homoscedasticity

- Fit model $\text{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$ and extract residuals $\{\widehat{\epsilon}_i\}_{i=1}^N$.
- Re-sample $N$ residuals $\widehat{\epsilon}_i^*$ with replacement.
- Keep $\mathbf{x}_i$ intact; form new outcomes $y_i^* = \mathbf{x}_i^T \widehat{\boldsymbol{\beta}} + \widehat{\epsilon}_i^*$ for $i = 1, \ldots, N$.
- Estimate $\widehat{\boldsymbol{\beta}}_b^*$ for $b = 1, \ldots, B$; form estimates/confidence intervals of your choosing from prior methods.
- Assumptions:
  - Homoscedasticity of errors.
  - Correct mean-model.
- Example: designed experiment with many discrete categories of $\mathbf{X}$ that each have relatively small samples.
- Does not generalize to GLMs (can't necessarily form $y_i^*$ based on $\widehat{\epsilon}_i^*$).

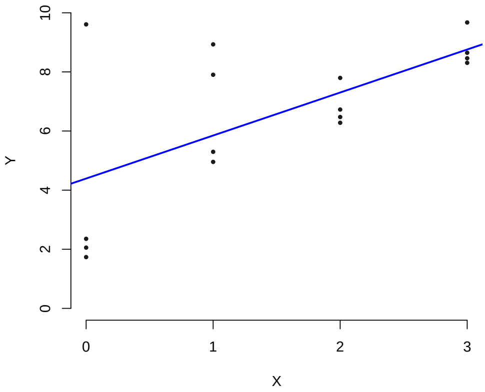
**More generally**: Fixed vs. random design

- If the mean model is correct, either version of the bootstrap should perform well regardless of whether $\mathbf{X}$ is fixed or random.
- If $\mathbf{X}$ is fixed, mean-model misspecification will tend to result in an overstated variance if you treat $\mathbf{X}$ as random.
- If $\mathbf{X}$ is random, mean-model misspecification will tend to result in an understated variance if you treat $\mathbf{X}$ as fixed.

**Example**: Bootstrap under high error skewness

```
## Set seed for reproducibility
set.seed(7345)

## Set sample size
n <- 16

## Generate predictor
x <- c(rep(0,n/4),rep(1,n/4),rep(2,n/4),rep(3,n/4))

## Generate outcome (linearity correct)
y <- 1 + 3*(x == 1) + 5*(x == 2) + 7*(x == 3) + rexp(n, 1/2)

## Create data frame
dat <- data.frame(cbind(x, y))

## Analysis on original data
zz <- lm(y ~ x, data = dat)

## Estimated slope
bhat <- as.numeric(zz$coef[2])

## Set bootstrap replicates
B <- 5000
```

**Example**: Bootstrap under high error skewness



**Example**: Bootstrap under high error skewness

```
## Model-based and sandwich-based standard errors
se.model <- as.numeric(sqrt(diag(vcov(zz)))[2])
se.sandwich <- as.numeric(sqrt(diag(sandwich(zz)))[2])

## Model-based CI
> c(EST = bhat,
+   CILO = bhat - qnorm(0.975)*se.model,
+   CIHI = bhat + qnorm(0.975)*se.model)
      EST      CILO      CIHI
1.4550308 0.5314966 2.3785650

## Sandwich-based CI
> c(EST = bhat,
+   CILO = bhat - qnorm(0.975)*se.sandwich,
+   CIHI = bhat + qnorm(0.975)*se.sandwich)
      EST      CILO      CIHI
1.4550308 0.4480281 2.4620335
```

**Example**: Bootstrap under high error skewness

```
## BOOTSTRAP METHOD 1: FULL-RESAMPLING

## Create a place to store results
b.results <- matrix(0, nrow = B, ncol = 1)

## Conduct bootstrap samples
for (j in 1:B)
{
  ## Random sample with replacement with original sample size in mind
  samp <- sample(1:n, replace = TRUE)
  bdat <- dat[samp,]

  ## Run model on bootstrap sample
  bzz <- lm(y ~ x, data = bdat)

  ## Extract results
  b.results[j,1] <- coef(bzz)[2]
}
```

**Example**: Bootstrap under high error skewness

```
## Bootstrap standard error
> sd(b.results)
[1] 0.5314059

## Symmetric large-sample-justified CI
> c(CILO = mean(b.results) - qnorm(0.975)*sd(b.results),
+   CIHI = mean(b.results) + qnorm(0.975)*sd(b.results))
     CILO      CIHI
0.4049048 2.4879778

## Asymmetric Pivot-based CI
qlo <- quantile(b.results, 0.025)
qhi <- quantile(b.results, 0.975)
> c(CILOW = as.numeric(2*mean(b.results) - qhi),
+   CIHI = as.numeric(2*mean(b.results) - qlo))
    CILOW      CIHI
0.6392062 2.6397876

## Quantile-based CI
> c(CILOW = as.numeric(quantile(b.results, c(0.025))),
+   CIHI = as.numeric(quantile(b.results, c(0.975))))
   CILOW     CIHI
0.253095 2.253676
```

**Example**: Bootstrap under high error skewness

```
## BOOTSTRAP METHOD 2: CONDITIONAL (FIXED X)

## Set bootstrap replicates
B <- 5000

## Create a place to store results
b.results <- matrix(0, nrow = B, ncol = 1)

## Extract residuals from fitted model
rsdls <- zz$residuals

## Keep a "fixed" version of the exposure
x.fixed <- dat$x

## Extract estimate of beta
bhat <- as.numeric(zz$coef)
```

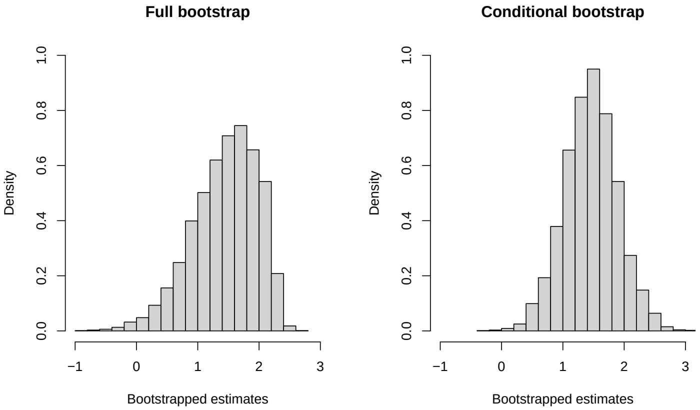
**Example**: Bootstrap under high error skewness

```
for (j in 1:B)
{
  ## Random sample of residuals with replacement
  samp <- sample(1:n, replace = TRUE)
  brsdls <- rsdls[samp]

  ## Append residuals to create a bootstrap FEV
  by <- bhat[1] + bhat[2]*x.fixed + brsdls

  ## Create bootstrap data set
  bdat <- data.frame(cbind(x.fixed, by))

  ## Run model on bootstrap sample
  bzz <- lm(by ~ x.fixed, data = bdat)

  ## Extract results
  b.results[j,1] <- coef(bzz)[2]
}
```

**Example**: Bootstrap under high error skewness

```
## Bootstrap standard error
> as.numeric(sd(b.results))
[1] 0.4352239

## Symmetric large-sample-justified CI
> c(CILO = mean(b.results) - qnorm(0.975)*sd(b.results),
+   CIHI = mean(b.results) + qnorm(0.975)*sd(b.results))
    CILO     CIHI
0.603114 2.309160

## Asymmetric Pivot-based CI
qlo <- quantile(b.results, 0.025)
qhi <- quantile(b.results, 0.975)
> c(CILOW = as.numeric(2*mean(b.results) - qhi),
+   CIHI = as.numeric(2*mean(b.results) - qlo))
    CILOW      CIHI
0.5893383 2.3304830

## Quantile-based CI
> c(CILOW = as.numeric(quantile(b.results, c(0.025))),
+   CIHI = as.numeric(quantile(b.results, c(0.975))))
    CILOW      CIHI
0.5817912 2.3229358
```

**Example**: Bootstrap under high error skewness



# The nonparametric bootstrap

**Example**: Bootstrap under high error skewness

- Why does the distribution of the bootstrapped estimates look so different between the two approaches?
- One point is clearly highly influential!
- Probability of inclusion in a single unconditional bootstrap replicate:

$$
\begin{aligned}
\text{P}(\geq 1 \text{ Inclusion}) &= 1 - (15/16)^{16} \approx 0.64. \\
\text{P}(\geq 2 \text{ Inclusions}) &= 1 - (15/16)^{16} - 15(15/16)^{15}(1/16) \approx 0.26. \\
\text{P}(\geq 3 \text{ Inclusions}) &\approx 0.074. \\
\text{P}(\geq 4 \text{ Inclusions}) &\approx 0.015. \\
\text{P}(\geq 5 \text{ Inclusions}) &\approx 0.0023.
\end{aligned}
$$

- Probability of inclusion in a single conditional bootstrap replicate:

$$
\begin{aligned}
\text{P}(\geq 1 \text{ Inclusion}) &= 1 - (3/4)^{4} \approx 0.68. \\
\text{P}(\geq 2 \text{ Inclusions}) &= 1 - (3/4)^{4} - 4(3/4)^{3}(1/4) \approx 0.26. \\
\text{P}(\geq 3 \text{ Inclusions}) &\approx 0.051. \\
\text{P}(4 \text{ Inclusions}) &\approx 0.0039. \\
\text{P}(\geq 5 \text{ Inclusions}) &= 0
\end{aligned}
$$

**Example**: Clever uses of the bootstrap

- We can use the bootstrap to answer questions that would otherwise be difficult or impossible to analytically answer.
- As an example, consider the following two models (unadjusted and adjusted) using OLS linear regression:

$$\begin{aligned} \mathrm{E}[Y|X=x] &= \alpha_0 + \alpha_1 x \\ \mathrm{E}[Y|X=x, Z=z] &= \beta_0 + \beta_1 x + \beta_2 z. \end{aligned}$$

- What is $\mathrm{Cov}[\widehat{\alpha}_1, \widehat{\beta}_1]$?

**Example**: Set up simulation

```
## Set seed for reproducibility
set.seed(7345)

## Set number of simulations
nsim <- 500

## Set sample size
n <- 500

## Set number of bootstrap replicates
B <- 100

## Store results
res <- matrix(0, nrow = nsim, ncol = 3)
```

**Example**: Simulation (part 1 - original data)

```
## Conduct simulation
for (j in 1:nsim)
{
  ## Generate predictors and outcomes
  X <- runif(n, 0, 5)
  Z <- runif(n, 0, 5)
  Y <- 1 + X + Z + rnorm(n, 0, 5)

  ## Fit adjusted and unadjusted models
  XU <- cbind(1, X)
  XA <- cbind(1, X, Z)
  res[j,1] <- (solve(t(XU) %*% XU) %*% (t(XU)) %*% Y)[2]
  res[j,2] <- (solve(t(XA) %*% XA) %*% (t(XA)) %*% Y)[2]
```

**Example**: Simulation (part 2 - bootstrap)

```
## Store bootstrapped results
bres <- matrix(0, nrow = B, ncol = 2)

for (b in 1:B)
{
  ## Full-size sample with replacement
  samp <- sample(1:n, size = n, replace = TRUE)
  bX <- X[samp]
  bZ <- Z[samp]
  bY <- Y[samp]

  ## Fit adjusted and unadjusted models on bootstrapped data
  bXU <- cbind(1, bX)
  bXA <- cbind(1, bX, bZ)
  bzz1 <- (solve(t(bXU) %*% bXU) %*% t(bXU) %*% bY)[2]
  bzz2 <- (solve(t(bXA) %*% bXA) %*% t(bXA) %*% bY)[2]

  ## Extracted data
  bres[b,1] <- bzz1
  bres[b,2] <- bzz2
}
```

**Example**: Report results

```
  ## Store estimated covariance
  res[j,3] <- cov(bres[,1],bres[,2])

  ## Track progress
  if (round(j/50) == (j/50)) {print(paste(j, "sims complete!"))}
}

## Average estimated covariance by bootstrap
> colMeans(res)[3]
[1] 0.02374372

## Actual covariance by simulation
> cov(res[,1],res[,2])
[1] 0.0229229
```

# Summary

**So far**:

- Sandwich and bootstrap estimation.

# SUMMARY

**Up next**:

- Overdispersion and quasilikelihood.