How do you characterize uncertainty for an estimator?

What are the range of possible values for monthly German tank production?

Do we really have evidence of low numbers of female births in China?

What is the probability that our estimate equals the population parameter value?

How would you characterize uncertainty for an estimator?

we must infer the variance of the estimator
German Tank Problem

- $S_i$ - serial number of salvaged tank $i$, total of $n$ salvaged tanks

$$S_i = \beta_0 + U_i$$

- unknown $\beta_0$
- formula $B_0 = \bar{S}_n$ estimator (sample mean)
- known $\bar{S}_n$ estimate of sample mean

- formula $\text{Var}(B_0) \equiv \sigma^2_{B_0}$
- known $\hat{\sigma}^2_{B_0}$ estimate of variance of $B_0$

- today we focus on constructing $\hat{\sigma}^2_{B_0}$
- "estimating the variance of the estimator"
Recall: Variance of the Estimator

- What are the covariates?
  
  \[ S_i = \beta_0 \cdot 1 + U_i \]

- implicit covariate \( X_i = 1 \)

- What is the variance of \( B_0 \)?
  
  \[
  \frac{\sum_{i=1}^{n} X_i^2 \mathbb{E}(U_i^2 | X) + \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j \mathbb{E}(U_i U_j | X)}{(\sum_{i=1}^{n} X_i^2)^2}
  \]

- \( X = (1, \ldots, 1) \) implies

  \[
  \frac{\sum_{i=1}^{n} \mathbb{E}(U_i^2 | X) + \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}(U_i U_j | X)}{n^2}
  \]
Is it reasonable to assume the variance of $U_i$ is constant?

- Consider: $J = 149 \Rightarrow \beta_0 = \mathbb{E} (S_i) = 75$
  - $U_i \in \{-74, \ldots, 0, \ldots, 74\}$ equally likely to take any value
  - true for all $i$ (each observation is equally noisy)
  - variance of $U_i$ is constant across $i$

- We assume $\mathbb{E} (U_i^2 | X) = \omega^2$

\[
\text{Var} (B_0) = \frac{n \cdot \omega^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E} (U_i U_j | X)}{n^2}
\]
Is it reasonable to assume that $U_i$ and $U_j$ are uncorrelated?

$U$ - discrete, $K$ possible values

- Each salvage tank number is drawn from a different production run
  - value that $U_i$ takes has no information regarding value $U_j$ takes
  - values of $U_i$ and $U_j$ are uncorrelated

- We assume $\mathbb{E}(U_i U_j | X) = 0$

$$\text{Var}(B_0) = \frac{n \cdot \omega^2 + 0}{n^2} = \frac{\omega^2}{n}$$
**Initial Question Answered**

Homoskedastic, Uncorrelated Variance Estimator

\[
\text{Var} \left( B_0 \right) = \frac{1}{n^2} \cdot \mathbb{E} \left( U_j^2 \right)
\]

- Method of Moments
  - Population Moment: \( \mathbb{E} \left( U_j^2 \right) = \omega^2 \)
  - Sample Moment: \( \frac{1}{n} \sum_{i=1}^{n} U_i^2 \)

- \( U_i \) is unobserved
- we observe \( \hat{S}_i = B_0 \) and

\[
S_i - \hat{S}_i = \hat{U}_i
\]

- \( \hat{U}_i \) - the estimated residual

- How should we estimate \( \mathbb{E} \left( U_j^2 \right) \)?
  - \( \frac{1}{n} \sum_{i=1}^{n} \hat{U}_i^2 = \hat{\omega}^2 \)

- What is the estimator of \( \text{Var} \left( B_0 \right) \)?
  - \( \hat{\omega}^2 \)
  - \( \frac{\hat{\omega}^2}{n} \)
- \( M_i \) - fraction of male births in country \( i \), total of \( n \) countries

\[
M_i = \beta_0 + \beta_1 C_i + U_i
\]

- Deviation-from-means form

\[
M_i^* = \beta_1 C_i^* + U_i^*
\]

- \( C_i^* = C_i - \bar{C} \)

- Review: What does \( \sum_{i=1}^{n} C_i^* \) equal?
  - \( 0 \)
  - \( \sum_{i=1}^{n} C_i^* = \sum_{i=1}^{n} (C_i - \bar{C}) = \sum_{i=1}^{n} C_i - n\bar{C} \)
Is it reasonable to assume that $U_i$ and $U_j$ are uncorrelated?

- $U_i$ - all factors that cause the male birth fraction in country $i$ to differ from the overall average
  - if variation in fraction of male births is random
  - values of $U_i$ and $U_j$ are uncorrelated

- We assume $\mathbb{E}(U_iU_j|C^*) = 0$

$$Var(B_1) = \frac{\sum_{i=1}^{n} (C_i^*)^2 \mathbb{E}(U_i^2|C^*)}{\left(\sum_{i=1}^{n} (C_i^*)^2\right)^2}$$
Is it reasonable to assume the variance of $U_i$ is constant?

- variation in fraction of male births is larger in countries with smaller populations
  - observations are not equally noisy (large population countries are less noisy)
  - variance of $U_i$ varies $i$

- We assume $\mathbb{E} (U_i^2 | X) = \omega_i^2$

\[
\text{Var} (B_1) = \frac{\sum_{i=1}^{n} (C_i^*)^2 \omega_i^2}{\left(\sum_{i=1}^{n} (C_i^*)^2\right)^2}
\]
Heteroskedasticity Robust Variance Estimator

Heteroskedastic, Uncorrelated Variance Estimator

\[
\text{Var}(B_1) = \frac{\sum_{i=1}^{n} (C_i^*)^2 \omega_i^2}{\left(\sum_{i=1}^{n} (C_i^*)^2\right)^2}
\]

- estimated residual

\[
\hat{U}_i = M_i^* - B_1 C_i^*
\]

- How should we estimate \( \mathbb{E}(U_i^2) = \omega_i^2 \)?
  - \( \hat{U}_i^2 = \hat{\omega}_i^2 \)
  - only one observation!
  - we don’t need to learn \( \omega_i^2 \), but rather the weighted average
    \[
    \frac{1}{n} \sum_{i=1}^{n} (C_i^*)^2 \omega_i^2
    \]

- What is the estimator of \( \text{Var}(B_1) \)?
  
  \[
  \frac{\sum_{i=1}^{n} (C_i^*)^2 \hat{U}_i^2}{\left(\sum_{i=1}^{n} (C_i^*)^2\right)^2}
  \]