1. $Y \sim X$, n, X_i fixed

$\hat{\beta_0}, \hat{\beta_1}$ LS estimates.

$\hat{\epsilon_i} = Y_i - (\hat{\beta_0} + \hat{\beta_1} X_i)$

ϵ_i do not depend on X

$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

$E[\epsilon_i] = 0 \ \forall i$, $Var[\epsilon_i] = \sigma^2$, ϵ's iid $\Big\}$ could assume

a) For a particular $i \in 1 \ldots n$,

LS estimates unbiased

$E[\hat{\epsilon_i}] = E[Y_i - \hat{\beta_0} - \hat{\beta_1} X_i] = E[Y_i] - \beta_0 - \beta_1 X_i$

$= \beta_0 + \beta_1 X_i + 0 - \hat{\beta_0} - \hat{\beta_1 X_i}$ — assuming moment assumptions

$= 0$ if we assume $E[\epsilon_i] = 0$ and w/ LS estimates.

(ii)

b) For some i, j, $i \neq j$,

$Cov(\hat{\epsilon_i}, \hat{\epsilon_j})$

$= E[(\hat{\epsilon_i} - E[\hat{\epsilon_i}])(\hat{\epsilon_j} - E[\hat{\epsilon_j}])]$

From the previous part a):

$= E[\hat{\epsilon_i} \cdot \hat{\epsilon_j}]$ under moment assumptions

Since ϵ_i, ϵ_j also independent under moment assumptions:

$\hat{\epsilon_i} = Y_i - \hat{Y_i} = \beta_0 + \beta_1 X_i + \epsilon_i - \hat{\beta_0} - \hat{\beta_1} X_i = \overbrace{(\beta_0 - \hat{\beta_0}) + (\beta_1 - \hat{\beta_1}) X_i}^{\text{fixed}} + \epsilon_i$

$\hat{\epsilon_j} = Y_j - \hat{Y_j} = \beta_0 + \beta_1 X_j + \epsilon_j - \hat{\beta_0} - \hat{\beta_1} X_j = \overbrace{(\beta_0 - \hat{\beta_0}) + (\beta_1 - \hat{\beta_1}) X_j}^{\text{fixed}} + \epsilon_j$

Then the functions of independent r.v.'s are also independent.

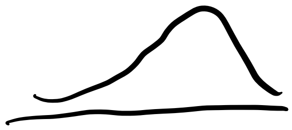
so $\hat{\epsilon_i}, \hat{\epsilon_j}$ are also indep. under moment assumptions.

so $Cov(\hat{\epsilon_i}, \hat{\epsilon_j}) = 0$ if

$\Rightarrow$ (ii)

c) 90% PI

ϵ_i NOT normal.



PI assumes normal distribution so coverage should be way worse than "usual" since $\epsilon_i \not\sim$ Normal.

overall, (ii)

d) Bootstrapping the residual will NOT reproduce the issues. Specifically, this method by definition chooses ϵ_i assuming a normal distribution, so it will ultimately "cover up" the lack of normality of Y.

(ii)

2. $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ SLR, X_i's fixed

a) Reduce data set into $\{(X_i, Y_i) : x_{0.05} \le X_i \le x_{0.95}\} = M$

aka middle 90% of the data (X values)

i) $\hat{\beta}_{1\,new} = \dfrac{\sum_{i \in M} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i \in M} (X_i - \bar{X})^2}$ vs. $\hat{\beta}_{1\,old} = \dfrac{\sum_{i \in whole} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i \in whole} (X_i - \bar{X})^2}$

Note that New dataset M is of size $0.9n$ ← n = original dataset size

$E[\hat{\beta}_{1\,old}] = \beta_1$ is unbiased, we know.

We expect that $\hat{\beta}_{1\,new}$ will still be an unbiased estimator for β_1 because the new dataset only excludes high leverage points based on large X values, since $\beta_1 = \frac{S_{xy}}{S_{xx}}$, the magnitude of change will be similar for both the numerator and denominator, so the ratio $S_{xy} : S_{xx}$ should stay the same, so $\hat{\beta}_{1\,new}$ will be unbiased.

ii) $Var[\hat{\beta}_{1\,old}] = \dfrac{1}{\sum_{i=1-n} (X_i - \bar{X})^2}$, $Var[\hat{\beta}_{new}] = \dfrac{1}{\sum_{i=0.05n}^{0.95n} (X_i - \bar{X})^2}$

Note that $\sum_{i=0.05n}^{0.95n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 - \underbrace{\sum_{i=1}^{0.05n} (X_i - \bar{X})^2}_{large} - \underbrace{\sum_{i=0.95n}^{n} (X_i - \bar{X})^2}_{large}$

Clearly, by creating the new dataset, we are removing those points that were unusually large or tiny (aka unusually far from $\bar{X}$). So,

$$\sum_{i=0.05n}^{0.95n} (X_i - \bar{X})^2 < \sum_{i=1}^{n} (X_i - \bar{X})^2$$ by definition. Thus, we expect

$Var[\hat{\beta}_{1\,new}]$ to be larger than $Var[\hat{\beta}_{1\,old}]$ from LS, since the denominator is much smaller than before. Also, a smaller multiple of n in the denom contributes to variance being larger.

b) Now, keep all data points. Instead, truncate the values:

$$\tilde{X}_i = \begin{cases} X_{0.05} & , X_i < 0.05 \\ X_i & , X_{0.05} < X_i < X_{0.95} \\ X_{0.95} & , X_i > 0.95 \end{cases}$$

but not assigning zeros

Note that this is similar to using Least Trimmed Squares instead of Least Squares. This method will for sure be more robust to the impact of outliers.

i) Now, we assign the same x, y values of $X_{0.05} \to y_{0.05}$ and $X_{0.95} \to y_{0.95}$ to all points outside the middle 90% of points and refit the model with all n of these points

In $\hat{\beta}_{1\,new} = \frac{Sxy_{new}}{Sxx_{new}}$ vs. $\hat{\beta}_{1\,old} = \frac{Sxy_{old}}{Sxx_{old}}$,

The new model of $\hat{y}$ is now: $\tilde{Y}_i = \begin{cases} \beta_0 + \beta_1 \tilde{X}_i + \epsilon_i & \text{if } \tilde{X}_i = X_i \\ y_{0.05} & \text{if } \tilde{X}_i = x_{0.05} \\ y_{0.95} & \text{if } \tilde{X}_i = x_{0.95} \end{cases}$

So, $\hat{\beta}_{1\,new}$ will now be biased

ii) $Var[\hat{\beta}_{1\,new}]$ should be lower than $\hat{\beta}_1$ now

c) Now remove pts by Y values, not X.

i) $\hat{\beta}_{1\,new}$ will be biased now since $E\left[\hat{\beta}_{1\,new}\right] = E\left[\frac{S_{xy}}{S_{xx}}\right]$ will be affected by change in $\sum_{0.05n}^{0.95n} (y_i - \bar{y})^2$

3. weight ~ temp + hum + fert , $n=37$
↳ all centered + scaled.

a) Residual Standard error $\hat{\sigma} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2 \leftarrow RSS_{full}}{df = 33}}$

$= \sqrt{\frac{RSS}{33}}$ ← $RSS = y^T(I-H) \cdot y$

$H = X(X^TX)^{-1}X^T$

$F = \frac{RSS_{partial} / \cancel{3}}{RSS_{full} / \cancel{33}\ 11} = 9.311 = \frac{11 \cdot RSS_{part}}{RSS_{full}} =$

b) $\text{Adj } R^2 = 1 - (1-R^2) \frac{n-1}{n-p-1}$

$= 1 - (1 - 0.4584) \cdot \frac{36}{32}$

$= 1 - 0.6093 = 0.3907$

c) The standard errors for hum and temp being quite high while for fert and intercept being low is likely due to collinearity between hum and temp. Since these covariates are very dependent on each other, they in turn reduce each others' significance in this full model, (both have high p-values) while fert is relatively independent of them so its variance estimate is definitely lower.

d). $\frac{\hat{\sigma}^2}{\hat{\sigma}'^2} \sim \frac{\chi^2_{33}}{\chi^2_{16}} \sim \boxed{F_{33,16}}$

$n=37$

$n=20$

e)