$g(\hat\beta_1, \hat\beta_2)$ $\quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{p\times 1}$ $\quad p = k+1$ $\quad s.e.(\hat\beta_2)$ $\quad \boxed{\sigma^2 (X'X)^{-1}_{33}}$ 3rd diag. element $\quad s.e.(\hat\beta_1)$

If X is fixed $\Rightarrow V(\hat\beta) = \sigma^2 (X'X)^{-1}$ $\qquad \hat\beta = (X'X)^{-1}X'y$

Bootstrapping Regression Models

- Consider the multiple linear regression model $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i$, $i = 1, \ldots, n$, where each $\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{ik})'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ is a $p \times 1$ parameter vector, and $p = k + 1$. It follows that

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

$\rightarrow$ $\epsilon \sim N(0, \sigma^2 I_n)$

where

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$\sigma^2 \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} = \begin{bmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{bmatrix}$

- The error terms $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed random variables with $E(\epsilon_i) = 0$ and $V(\epsilon_i) = \sigma^2$ for all i.

- Oftentimes, we also assume normality for the error terms.

Example (~~Givens and Hoeting (2013)) Copper-Nickel~~ Alloy GPA (y_i)

The director of the office of admissions wants to know if a student's grade point average (y_i) can be predicted from the entrance test score (x_i). The simple linear regression model is given by

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$

- Suppose that the 20 students were randomly sampled from a population of students. X & Y are random
- Find the bootstrap estimate of the standard error of $\hat\beta_1$.
- Obtain the bootstrap 95% percentile CI for β_1
- The dataset is given in the admissions.csv file.

$\frac{\beta_1}{\beta_0}$

Example (Givens and Hoeting (2013)) Copper-Nickel Alloy

The table below gives 13 measurements of corrosion loss (y_i) in copper-nickel alloys, each with a specific iron content (x_i). Of interest is the corrosion loss in the alloys as the iron content increases, relative to the corrosion loss when there is no iron. Thus, consider the estimation of $\theta = \beta_1/\beta_0$ in a simple linear regression:

X is fixed
Y is random

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$

$E(y) = \beta_0 + \beta_1 x$

- Suppose that the iron content measurements are fixed. Letting $\boldsymbol{z}_i = (x_i, y_i)'$ for $i = 1, \ldots, 13$, the data are $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_{13}\}$.
- The point estimate of θ based on the observed data is $\hat\theta = \hat\beta_1/\hat\beta_0 = -0.185$. $\quad \theta = \frac{\beta_1}{\beta_0}$
- We also want ① an estimate of $s.e.\left(\hat\theta\right)$ and ② a CI for θ based on the bootstrap.
- ~~How to obtain the bootstrap samples? We obtain 10,000 bootstrap samples of the form~~ $\{\boldsymbol{z}_1^*, \boldsymbol{z}_2^*, \ldots, \boldsymbol{z}_{13}^*\}$ by resampling from $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_{13}\}$.

Table. Copper-nickel alloy data

x_i	0.01	0.48	0.71	0.95	1.19	0.01	0.48
y_i	127.6	124.0	110.8	103.9	101.5	130.1	122.0
x_i	1.44	0.71	1.96	0.01	1.44	1.96	
y_i	92.3	113.1	83.7	128.0	91.4	86.2	

similar

R EXAMPLE

$\hat\theta$ (sampling dist.)

P (unknown) $\rightarrow (x_1, \ldots, x_n)$ 1 sample $\rightarrow \hat P \rightarrow x^* \Rightarrow \hat\theta^*$

Real world $\qquad$ Bootstrap dist

s.e.$\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix}$ bootstrap! $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$, $\varepsilon_1, \dots, \varepsilon_n \overset{iid}{\sim} (0, \sigma^2)$

$P = (\beta, F)$

How to obtain the bootstrap samples?

- Note that the probability model has two components $P = (\boldsymbol{\beta}, F)$ where F represents the distribution of the error term (or some other distributional assumption on the data)
- Thus $\hat{P} = \left(\hat{\boldsymbol{\beta}}, \hat{F}\right)$, $\hat{\boldsymbol{\beta}}$ is the least squares estimator of $\boldsymbol{\beta}$. Not so clear how to estimate F.
- The following procedure is a naive bootstrapping mistake:
  - resample from the collection of response values $\{y_1, \dots, y_n\}$ a new pseudo-response, say Y_i^*, for each observed $\boldsymbol{x}_i$, thereby generating a new dataset.

$(y_1^*, \underline{x}_1), \dots, (y_n^*, \underline{x}_n)$

$(y_1, \dots, y_n)$

$\rightarrow (y_1^*, \dots, y_n^*)$

$$\boldsymbol{y}^* = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix}$$

$\hat{\beta}^* = (X'X)^{-1}X'\underline{y}^*$

  - Then a bootstrap parameter vector estimate $\hat{\boldsymbol{\beta}}^*$ would be calculated from these pseudo-data
  - After repeating this process many times, the bootstrap distribution of $\hat{\boldsymbol{\beta}}^*$ would be used for inference about $\boldsymbol{\beta}$

  $y_i \sim N(\underline{x}_i'\beta, \sigma^2)$

  The mistake: $Y_i|\boldsymbol{x}_i$ are not i.i.d, they have different conditional means.

  - Thus it's not appropriate to generate bootstrap regression datasets this way.
- *We must ask what variables are iid* in order to determine a correct bootstrapping approach.

$\underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ $\quad X_{n \times p} = \begin{bmatrix} \underline{x}_1' \\ \vdots \\ \underline{x}_n' \end{bmatrix}$

	No dist. assumption on error (nonpara. Case 1a)	ε_i are normal (parametric)
Case 1 (X's are fixed)	1	2
Case 2 (X's are random)	3	4

Case 1a

$y_i = x_i'\beta + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} F(\epsilon_i)$, $V(\epsilon_i)=\sigma^2$, x_i's are fixed

$X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}$ kind of iid

Case 1: The covariate values $\boldsymbol{x}_i$, $i = 1, \dots, n$ are fixed. *Bootstrap the residuals.*

Case 1a. No distributional assumption is made on $\epsilon_i s$. In this case, we compute $\hat{\boldsymbol{\beta}}$, obtain the residuals, and then sample from the empirical distribution of the residuals. $V(\epsilon_i) = \sigma_i^2, i=1,\dots,n$
Procedure to obtain the bootstrap samples:

1. Under the model $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $V(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_p$, estimate $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$
2. Get the residuals $e_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$, $i = 1, \dots n$ $\quad y_i - \hat{y}_i \to x_i'\hat\beta$
3. Repeat B times:
   - Sample $e_1^*, e_2^*, \dots, e_n^*$ using SRSWR from $\{e_1, e_2, \dots, e_n\}$ $\quad \{r_1, \dots, r_n\}$ nonparametric bootstrap
   - Compute $y_i^* = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} + e_i^*$, $i = 1, \dots, n$, and let $\boldsymbol{y}^*$ be $\boldsymbol{y}^* = (y_1^*, \dots, y_n^*)'$
4. The bootstrap data sets are $\{\boldsymbol{y}^*, \mathbf{X}\}$ (a total of B sets of these). We can also get $\hat{\boldsymbol{\beta}}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}^*$.
   $\{y_b^*, X\}, b=1,\dots,B$

This process can be repeated many times to build an emprical distribution for $\hat{\boldsymbol{\beta}}^*$ that can be used for inference. Note that the e_i^* are actually not independent, though they are usually roughly so. The strategy of bootstrapping residuals is at the core of simple bootstrapping methods for other models such as autoregressive models, nonparametric regression, and generalized linear models.

Improvement (from Davison & Hinkley)

- Note that the vector of residuals is $\boldsymbol{e} = (\boldsymbol{I}_n - \mathbf{H})\boldsymbol{y}$. Thus $V(\boldsymbol{e}) = \sigma^2(\boldsymbol{I}_n - \mathbf{H})$
- In Step 3, instead of using the residuals e_i, use the modified residuals $r_i = \frac{e_i}{\sqrt{1-h_{ii}}}$, $i = 1, \dots, n$, where h_{ii} is the ith diagonal element of $\mathbf{H}$. This $e_1^*, e_2^*, \dots, e_n^*$ are sampled from $r_1, \dots, r_n$ or their centered counterparts $r_1 - \bar{r}, \dots, r_n - \bar{r}$.

$f(e_i) = \begin{cases} e_i\left[\frac{1+\sqrt{5}}{2}\right], & v_i = 0 \\ e_i\left(\frac{1-\sqrt{5}}{2}\right), & v_i = 1 \end{cases}$, $v_i \sim$

Under Nonconstant variance

In the heteroskedastic case, an alternative to the residual bootstrap is the *wild bootstrap*, where the y_is are obtained as follows

$$y_i^* = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} + f(e_i),$$

$e_i^* = f(e_i)$

where $f(e_i) = e_i\left[\frac{1+(-1)^{v_i}\sqrt{5}}{2}\right]$ and $v_i \sim$ Bernoulli$\left(\frac{5+\sqrt{5}}{10}\right)$. (see Davison and Hinkley)

Case 1b. The $\epsilon_i s$ are assumed to be $N(0, \sigma^2)$. In this case, we compute $\hat{\boldsymbol{\beta}}$, obtain the residuals, and then sample from the plug-in dustribution

1. Under the model $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N_p(\mathbf{0}, \sigma^2\boldsymbol{I}_p)$, estimate $\quad \hat\beta = (X'X)^{-1}X'y$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}\right)^2$$

(MLE) $\quad \hat\sigma^2 = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - x_i'\hat\beta)^2$ (Least squares)

2. Repeat B times
   - Obtained the bootstrapped residuals as $e_i^* \sim N(0, \hat{\sigma}^2)$, $i = 1, \dots, n$ $\quad (e_1^*, \dots, e_n^*)$
   - Compute $y_i^* = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} + e_i^*$, $i = 1, \dots, n$, and let $\boldsymbol{y}^*$ be $\boldsymbol{y}^* = (y_1^*, \dots, y_n^*)'$
3. The bootstrap data sets are $\{\boldsymbol{y}^*, \mathbf{X}\}$ (a total of B sets of these).

parametric bootstrap

$V(\hat{\beta}_i)$, $V(\sqrt{\hat{\beta}_i})$ $\quad$ $\text{s.e.}(\bar{x}) = \frac{s}{\sqrt{n}}$, $i = 1, \ldots, p$

Remarks about Case 1

- We don't need to do Monte Carlo simulations to figure out the bootstrap standard error for the components of $\hat{\boldsymbol{\beta}}^*$, since we have a closed-form expression

$$
\begin{aligned}
V_{\hat{F}}\left(\hat{\boldsymbol{\beta}}^*\right) &= V\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{y}^*\right) \\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'V_{\hat{F}}\left(\boldsymbol{y}^*\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\
&= \hat{\sigma}^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} \qquad \hat{\sigma}^2 (X'X)^{-1} \\
&= \hat{V}\left(\hat{\boldsymbol{\beta}}\right)
\end{aligned}
$$

  Note that $V_{\hat{F}}\left(\boldsymbol{y}^*\right)$ is taken with respect to the distribution of $\boldsymbol{e}^*$ so $\hat{\boldsymbol{\beta}}$ is fixed.

- Thus the bootstrap estimate of the SE of $\hat{\beta}_j$ is the same as usual estimate of $\widehat{\text{s.e.}}\left(\hat{\beta}_j\right)$.

- We can go on to apply the bootstrap to more general regression models that have no mathematical solution: where the regression function is nonlinear in the parameters.

The case for random covariates $\quad$ experiment (randomization)

- Suppose that the data arose from an observational study, where both response and predictors are measured from a collection of individuals selected at random.

- In this case, the random variables $\boldsymbol{w}_i = (\boldsymbol{x}_i, y_i)'$ are iid from a joint response-predictor (multivariate distribution) distribution F.

- To bootstrap, sample $\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_n^*$ completely at random with replacement from $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$

- Apply the regression model to the resulting pseudo-dataset to obtain a bootstrap parameter estimate $\hat{\boldsymbol{\beta}}^*$.

- Repeat these steps many times, then proceed to inference as in the first approach.

- This approach of bootstrapping the cases is sometimes called the *paired bootstrap*.

$$
\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \underset{\sim}{x}_1' \\ \vdots \\ \underset{\sim}{x}_n' \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad \text{sampling pairs!}
$$

$$
\begin{bmatrix} y_i \\ \underset{\sim}{x}_i \end{bmatrix}_{p\times 1} \overset{iid}{\sim} \text{Dist}(\quad), \quad i = 1, \ldots, n
$$

$$
\left\{ \begin{bmatrix} y_1 \\ \underset{\sim}{x}_1 \end{bmatrix}, \ldots, \begin{bmatrix} y_n \\ \underset{\sim}{x}_n \end{bmatrix} \right\}
$$

sample from this set

Case 2: The covariate values $\boldsymbol{x}_i$, $i = 1, \ldots, n$ are random. *Bootstrap the pairs.*

Case 2a. No distributional assumption is made. In this case, we compute sample from the empirical distribution of the pairs.

Under the model $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i$, $i = 1, \ldots, n$ where $\boldsymbol{w}_i = (\boldsymbol{x}_i, y_i)'$, $i = 1, \ldots, n$ are iid from some unknown distribution $F(\cdot)$,

1. Repeat the following B times
   - Resample $\boldsymbol{w}_1^*, \boldsymbol{w}_2^*, \ldots, \boldsymbol{w}_n^*$ using SRSWR from $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$
   - Compute $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\boldsymbol{y}^*$ where $\mathbf{X}^* = [\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*]'$.
2. The bootstrap datasets are the B sets of $\{\boldsymbol{w}_1^*, \boldsymbol{w}_2^*, \ldots, \boldsymbol{w}_n^*\}$.

Case 2b. We make a distributional assumption.

Consider the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, \ldots, n$, where $\begin{bmatrix} x_i \\ \epsilon_i \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_e^2 \end{bmatrix}\right)$, $i = 1, \ldots, n$. Equivalently the joint distribution of $\begin{bmatrix} x_i \\ y_i \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \beta_0 + \beta_1\mu_x \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \beta_1\sigma_x^2 \\ \beta_1\sigma_x^2 & \beta_1^2\sigma_x^2 + \sigma_e^2 \end{bmatrix}\right)$.

1. Compute $\hat{\mu}_x = \bar{X}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ as least squares estimators, $\hat{\sigma}_x^2 =$ sample variance of $x_1, \ldots, x_n$, and $\hat{\sigma}_e^2 = \frac{1}{n}\sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$
2. Repeat the following B times
   - Resample $\begin{bmatrix} x_i^* \\ y_i^* \end{bmatrix} \sim N\left(\begin{bmatrix} \hat{\mu}_x \\ \hat{\beta}_0 + \hat{\beta}_1\hat{\mu}_x \end{bmatrix}, \begin{bmatrix} \hat{\sigma}_x^2 & \hat{\beta}_1\hat{\sigma}_x^2 \\ \hat{\beta}_1\hat{\sigma}_x^2 & \hat{\beta}_1^2\hat{\sigma}_x^2 + \hat{\sigma}_e^2 \end{bmatrix}\right)$, $i = 1, \ldots, n$.
   - Compute $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$ similarly as $\hat{\beta}_0$ and $\hat{\beta}_1$.
3. The bootstrap datasets are the B sets of $\left\{\begin{bmatrix} x_1^* \\ y_1^* \end{bmatrix}, \begin{bmatrix} x_2^* \\ y_2^* \end{bmatrix}, \ldots, \begin{bmatrix} x_n^* \\ y_n^* \end{bmatrix}\right\}$.

Notes about the bootstrap for random covariates.

- If you have doubts about the adequacy of the regression model, the constancy of the residual variance, or other regression assumptions, the paired bootstrap will be less sensitive to violations in the assumptions than will bootstrapping the residuals.
- The paired bootstrap sampling more directly mirrors the original data generation mechanism in cases where the predictors are not considered fixed.

Do the following for the Copper-Nickel example.

- Verify the estimate of $\hat{\theta}$
- Obtain a histogram of the bootstrap estimates of $\hat{\theta}$ (i.e., the $\hat{\theta}^*$s) obtained from regressions of the bootstrap datasets. The histogram summarizes the sampling variability of $\hat{\theta}$ as an estimator of θ.
- Obtain a bootstrap estimate of the standard error of $\hat{\theta}$
- Obtain a 95% confidence interval for θ based on the percentile method.