**Example (Efron & Tibshirani) The Patch Data**

Eight subjects wore medical patches designed to infuse a certain naturally-occurring hormone into the blood stream. Each subjects had his blood levels of the hormone measured after wearing three different patches: a placebo patch containing no hormone, and "old patch" manufactured at an older plant, and a "new" patch manufactured at a newly-opened plant. The first three columns of the table below show the three blood-level measurements for each subject.

The purpose of the patch experiment was to show bioequivalence. Patches manufactured at the old plant had already been approved for sale by the Food and Drug Administration (FDA). Patches from the new facility did not require a full new FDA investigation. They would be approved for sale if it can be shown that they were bioequivalent to those from the old facility. The FDA criterion for bioequivalence is that the expected value of the new patches match that of the old patches in the sense that

$$\frac{|E\,(\text{new}) - E\,(\text{old})|}{E\,(\text{old}) - E\,(\text{placebo})} \leq 0.20.$$

In other words, the FDA wants the new facility to match the old facility within 20% of the amount of hormone the old drug adds to placebo blood levels. Let $\theta$ be the parameter

$$\theta = \frac{E\,(\text{new}) - E\,(\text{old})}{E\,(\text{old}) - E\,(\text{placebo})}.$$

For now, we only consider the bias and standard error of the plug-in estimate $\hat{\theta}$. Later we will do a CI for $\theta$.

| subject | placebo | old patch | new patch | $z =$ old $-$ placebo | $y =$ new $-$ old |
|---|---|---|---|---|---|
| 1 | 9,243 | 17,649 | 16,449 | 8,406 | -1,200 |
| 2 | 9,671 | 12,013 | 14,614 | 2,342 | 2,601 |
| 3 | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 4 | | | | | |
| 5 | | | | | |
| 6 | | | | | |
| 7 | | | | | |
| 8 | | | | | |
| mean | | | | 6,342 | -452.3 |

- In the patch data example, we are interested in two statistics $z_i$ and $y_i$ obtained for each of the eight subjects

$$z_i = \text{old patch measurement} - \text{placebo measurement}$$
$$y_i = \text{new patch measurement} - \text{old patch measurement}$$

- Assume that the pairs $x_i = (y_i, z_i)$, $i = 1, \ldots, n$ are obtained by random sampling from an unknown bivariate distribution $F$, $F \longrightarrow \boldsymbol{x} = (x_1, \ldots, x_8)'$, then the parameter $\theta$ is

$$\theta = \frac{E_F\,(y)}{E_F\,(z)}$$

- The plug-in estimator of $\theta$ is given by

$$\hat{\theta} = \frac{\bar{y}}{\bar{z}} = \frac{-452.3}{6,342} = -0.0713,$$

which is less than 0.20 in absolute value.

**Bootstrap procedure to estimate the bias in Efron & Tibshirani's patch data**

1. Sample $\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \ldots, \boldsymbol{X}_8^*$ from $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_8\}$ with replacement
2. Calculate $\hat{\theta}^* = \frac{\bar{y}^*}{\bar{z}^*}$ using $\{\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \ldots, \boldsymbol{X}_8^*\}$
3. Repeat (1) and (2) $B$ times
4. The bootstrap estimate of the bias is given by $\bar{\hat{\theta}}^* - \hat{\theta}$

- Implementing the above procedure, we get the following estimate

$$\widehat{\text{bias}}_\theta\left(\hat{\theta}\right) = \bar{\hat{\theta}}^* - \hat{\theta} = -0.0670 - (-0.0713) = 0.0043$$

- General rule of thumb: A bias of less than 0.25 standard errors can be ignored. Bias is "small" since

$$\frac{\widehat{\text{bias}}_\theta\left(\hat{\theta}\right)}{\widehat{\text{s.e.}}_\theta\left(\hat{\theta}\right)} = \frac{0.0043}{0.1050} = 0.041 \leq 0.25,$$

**Improved estimation of the bias**

- There is an improved estimator of the bias which applies when $\hat{\theta}$ is the plug-in estimator of $\theta$
- Let $\boldsymbol{X} = (X_1, \ldots, X_n)'$, and $\boldsymbol{X}^* = (X_1^*, \ldots, X_n^*)'$. Define also the following

$$\begin{aligned}
P_1^* &= \frac{\#\text{ of } X_i\text{s equal to } X_1}{n} = \text{proportion of } X_i\text{s equal to } X_1 \\
P_2^* &= \text{defined similarly} \\
&\vdots \\
P_n^* &= \\
\boldsymbol{P}^* &= (P_1^*, \ldots, P_n^*)', \quad \text{where } \sum_{i=1}^n P_i^* = 1.
\end{aligned}$$

- The vector $\boldsymbol{P}^*$ is called the *resampling vector*. To illustrate, if $\boldsymbol{X}^* = (X_1, X_6, X_6, X_5, X_7, X_1, X_3, X_8)'$ then $\boldsymbol{P}^* = \left(\frac{2}{8}, \frac{0}{8}, \frac{1}{8}, \frac{0}{8}, \frac{1}{8}, \frac{2}{8}, \frac{1}{8}, \frac{1}{8}\right)'$.
- Note that

$$\hat{\theta} = \frac{\bar{y}}{\bar{z}}, \quad \hat{\theta}^* = \frac{\bar{y}^*}{\bar{z}^*} = \frac{\sum_{j=1}^8 P_j^* y_j}{\sum_{j=1}^8 P_j^* z_j}$$

- Write $\hat{\theta}^* = T\left(\boldsymbol{P}^*\right)$. This makes sense as $\hat{\theta}^*$ depends on the $P_j^*$s chosen.
- Also if $\boldsymbol{P}_0 = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)'$ then

$$T\left(\boldsymbol{P}_0\right) = \frac{\sum_{j=1}^8 \frac{1}{n} y_j}{\sum_{j=1}^8 \frac{1}{n} z_j} = \hat{\theta} = t\left(\hat{F}\right).$$

- There are $B$ $\boldsymbol{P}^*$s: $\boldsymbol{P}^{*1}, \boldsymbol{P}^{*2}, \ldots, \boldsymbol{P}^{*B}$. Now let $\bar{\boldsymbol{P}}^* = \frac{1}{B}\sum_{b=1}^B \boldsymbol{P}^{*b}$.
- The improved bootstrap estimator of the bias is

$$\widetilde{\text{bias}}_\theta\left(\hat{\theta}\right) = \bar{\hat{\theta}}^* - T\left(\bar{\boldsymbol{P}}^*\right).$$