INFERENCES

- STANDARD ERROR (Jackknife)
- SAMPLING DIST'N (Bootstrap)

\( \hat{\theta}(x) = \hat{T} : \sum \psi \left( \frac{x_i - \hat{T}}{c_s} \right) = 0 \)

\( \hat{s} = s(\hat{T}) \)
THE SAMPLE MEAN

\[(x_1, \ldots, x_n) \rightarrow \bar{x}_n = \frac{1}{n} \sum x_i\]

\[n \hat{V}(\bar{x}_n) = \frac{1}{n-p} \sum (x_i - \bar{x}_n)^2\]

\[\text{"The usual"} \quad \text{MSE MLE UBE REML}\]

\[= \frac{(n-1)}{n} \cdot (-1, 0, 1)\]

"STUDENT" derived this and with \(p = 1\) got ME to distin
for: \(\ln x / \sigma\)

linearity helped him a lot!
ESTIMATING THE VARIANCE OF $\bar{X} = \bar{X}_n$

$T = \bar{X}_n = \frac{1}{n} \sum X_i$

$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 = \hat{\sigma}^2$

$\hat{V}(\bar{X}_n) = \left( \frac{S}{\sqrt{n}} \right)^2 = \left( \frac{\hat{\sigma}}{\sqrt{n}} \right)^2 \uparrow \text{SE.}$

ANOTHER WAY

- Delete $X_j$
- Compute $\bar{X}_{-j} = \frac{n \bar{X} - X_j}{n-1}$

$\hat{S}_C^*(X_j) = n(\bar{X}_n - \bar{X}_{-j})$

$\epsilon = \frac{1}{n}$

Sensitivity Curve

a "derivative" (difference)
\[ \overline{S}(x_j) = \frac{n}{n-1} (x_j - \overline{x}) \]

\[ = n (\overline{x}_n - \overline{x}_j) \]

\[ \frac{1}{n-1} \sum (x_j - \overline{x})^2 = S^2 = \left( \frac{n-1}{n} \right) \frac{1}{n} \sum_{j=1}^{n} \overline{S}(x_j) \]

"Sample Variance of \( X_j \)"

"Adjust to sample of size \( n \)"

\[ \text{What if we do this for a general } T(x)? \]
General \( SC \) (no *)

\[
SC(x) = \underbrace{\text{Book Defn}}_n \left[ T_n(x, \ldots, x_{n-1}; x) - T_{n-1}(x, \ldots, x_{n-1}) \right]
\]

- Measures "influence" of adding \( x \) to \((x, \ldots, x_{n-1})\)

- Can be used to produce

\[
\tilde{V}(T_n)
\]

\[
SC^*(x_j) = \left( \text{select } x_j \right)
\]

\[
= n \left[ T_n(x, \ldots, x_0) - T_{n-1}(x, \ldots, x_{n-1}) \right]
\]

\[
= n \left[ T_n - T_{n-1} \right]
\]
**SENSITIVITY CURVE**

\[ Sc^*(x_j) = n \left[ T_n(x_1, \ldots, x_n) 
- T_{n-1}(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \right] \]

\[ Sc^*(x_j) = n \left[ T_n - T_j \right] \]

How much does \( T \) depend on \( x_j \)?

**Variance**

\[ n \hat{V}(\hat{T}_n) = \left( \frac{n-1}{n} \right) \frac{1}{n} \sum_{j=1}^{n} Sc^2(x_j) \]

\[ SE(\hat{T}_n) = \sqrt{\hat{V}(\hat{T}_n)} \] **Correct for** \( n \rightarrow n-1 \)

**OK for** Smooth \( \{ \text{differentiable} \} \)

**Not for** \{ median, \( \hat{y} \) \}
What does $SC^*(x_j)$ measure?

$n \times \left[ \text{change from adding } x_j \text{ to } x_{-j} \right]$
The $SC^*(x)$ estimates (approximates) $\Psi$ for an M-estimate.

Data Point

\[
\begin{align*}
T &= 55.7 \\
\text{EST } T &= 55.7 \\
C &= 2 \\
\text{(small)}
\end{align*}
\]
Since

\[ n \left( \frac{1}{n} \sum_{i=0}^{n-1} \sqrt{c^*(x_i)^2} \right) \]

\[ n \hat{V}(\hat{T}) = \left( \frac{n-1}{n} \right) \frac{1}{n} \sum c^*(x_i)^2 \]

when \( T \) is resistant (\( \Psi \) a good shape)

outliers are down-weighted.

we get robust/resistant CI's and tests.

But only 1st + 2nd moment information
Newcomb's Speed of Light Data

\[ \text{Value} \times 10^{-3} = 24.8 \]

\[ h = 20 \]

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (n = 710)</td>
<td>21.75</td>
</tr>
<tr>
<td>( T (.05) )</td>
<td>24.39</td>
</tr>
<tr>
<td>( T (.20) )</td>
<td>25.67</td>
</tr>
<tr>
<td>( T (.40) )</td>
<td>25.50</td>
</tr>
<tr>
<td>Median</td>
<td>25.5</td>
</tr>
<tr>
<td>Trimean</td>
<td>25.50</td>
</tr>
<tr>
<td>Estimator</td>
<td>Estimate</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
</tr>
<tr>
<td>orig data mean</td>
<td>21.75</td>
</tr>
<tr>
<td>5% Trim</td>
<td>24.39</td>
</tr>
<tr>
<td>drop mean</td>
<td>25.21</td>
</tr>
<tr>
<td>5% Trim</td>
<td>25.44</td>
</tr>
</tbody>
</table>

\[ 100 \ N(0,1) \]

\[ \hat{X} \sim N_5 \]

\[ SD \ via \ SC^* \]

\[ -0.208 \]

\[ -0.211 \]

\[ .879/10 \]

\[ .923/10 \]
Computing for M-estimates:

1. Do WLS to get $T_n$
2. Delete "j"'s value by setting $w_j = 0$.
3. Repeat WLS, start at $T_n$ → $T_{-j}$; 0 or 2 iterations is enough.
4. $s_c^*(x_j) = n \{ T_n - T_{-j} \}$
5. Compute variance estimate
function(y, cc, coef = biw$est) {
    n <- length(y)
    curve <- NULL
    for(index in 1:n) {
        temp <- y[-index]
        estdel <- biweight(temp, cc, init = coef)$est
        curve <- c(curve, estdel)
    }
    curve <- n * (coef - curve)
    estvar <- (t(curve) ** curve) / n
    estvar <- (((n - 1)/n) * estvar) / n
    estse <- sqrt(estvar)
    list(data = y, curve = curve, est = coef, estvar = estvar, estse = estse)
}

biw <- biweight(y, c, iter)

# Consider increasing to be more accurate, but not by much.
An Analogue to Sample Mean

**Sample Mean**

\[ \bar{X}_n = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X}_n) \]

**General**

\[ T_n = \frac{1}{n} \sum_{j=1}^{n} (X_j - \bar{X}_n) = S C^* (X_j) \]

\[ S^2 = \text{sample variance} \]

\[ \hat{V}(\bar{X}) = \frac{S^2}{n} \]

\[ \hat{S}_E (\bar{X}) = \frac{S}{\sqrt{n}} \]

\[ \left( \frac{n-1}{n} \right) \frac{1}{n} \sum_{j=1}^{n} S C^*(X_j)^2 \quad \text{divided by} \quad n \]

Square root of the above
Influence Curve

$\text{SC}_n^*(x) \rightarrow \text{IC}(x, F, T)$

(new mass true distn estimator)

$(t - x_n) \rightarrow (t - \mu)$

0 T an M-estimator

$\text{IC}(x, F, T) = \text{const}(F, \psi) \psi(\frac{x - T}{S})$

$T = T(F, \psi; c)$

$S = S(F, \psi; c)$

Basic Result

$n \nu(t) \rightarrow E[\text{IC}(X, F, T)^2]$ for differentiable functionals

$X \sim F$ of $F \{ T(\cdot) \}$.
FUNCTIONALS OF DISTN'S T(F)

EXAMPLES

Mean \[ \int x dF(x) = \mu \]

Var \[ \int x^2 dF(x) - \left[ \int x dF(x) \right]^2 \]

Median \( s \) \[ F^{-1}(\frac{1}{2}) \]

\( \alpha \)-Trim \[ \frac{1}{1-2\alpha} \int F''(1-\alpha) x dF(x) \]

\( M \) \[ M-est \]

The sample variance (with \( n-1 \))

is not a functional \( \left( \sum_{i=1}^{n} \frac{(\tilde{x}_i - \tilde{x})^2}{n} \right) = \sigma_x^2 \)
\[ \lim_{\varepsilon \to 0} \frac{T((1-\varepsilon)F + \varepsilon H) - T(F)}{\varepsilon} \]

\[ \rightarrow IC(H,F,T) \quad \text{H a dist} \]

\[ = \int IC(T,F,T) \left( dH(x) - dF(x) \right) \]

let \( H = S_x \) a point mass at \( x \).

So,

\[ T(H) = T(F) + IC + \text{Remainder} \]
**Examples IC(\text{IC})**

\[
\text{mean } \mathbb{E}(X - \mu) \quad \mathbb{E}(X - \mu)^2 = \sigma^2
\]

\[
\text{var} \rightarrow \text{HW} \leftarrow \text{x-Time}
\]

\[
\text{median } \frac{\text{sgn}(X - \mu)}{2 f(\mu)} \quad \mu = \text{median}
\]

\[
\text{M-est} \quad \frac{\psi(X, t(f))}{- \int \frac{d}{dt} \psi(x, t) f(x) \, dx} = \frac{\psi}{\text{const}_\psi, \psi}
\]

\[
\text{Location} \quad \frac{\psi(X - t(f))}{\int \psi'(X - t(f)) f(x) \, dx}
\]
Influence Curve Sketch

Motivation

\[ x_1, \ldots, x_n \rightarrow \hat{\mu} \quad (m \in \mathbb{R}) \]

\[ \hat{g}(m) = g(\hat{\mu}) \]

\[ \text{1 dim (functional)} \quad \frac{1}{\text{dim}} \quad = g(m) + \{g'(m)\}^T (\hat{\mu} - m) \]

\[ \mathbb{E}[\hat{g}(\hat{\mu}) - g(m)]^2 \]

\[ = g'(m)^T \mathbb{E}[(m - \hat{\mu})(m - \hat{\mu})^T] g'(m) \]

\[ = g'(m)^T \text{diag}(g'(m)) g'(m) \]

\[ = \hat{\nabla}[g(\hat{\mu})] \]
VARIANCE

\[ n E \left[ \left( T(\hat{f}) - T(f) \right)^2 \right] = \sum_{x \in \Omega} \left( \hat{g}(x) - g(x) \right)^2 \]

\[ E = \left[ \int IC(x) d \left\{ \hat{v}_n \left( \hat{\gamma}, \hat{\nu} \right) \right\} \right] \times \]

\[ \circ \times \left[ \int IC(y) d \left\{ \hat{v}_n \left( \hat{\gamma}, \hat{\nu} \right) \right\} \right] \]

\[ \hat{v}_n (\hat{f} - f) \approx B^o (F(x)) \]

\[ \rightarrow \int \left[ IC^2(x) f(x) \right] dx \]

\[ = E \left[ IC^2(x) \right] \]
NON IID SAMPLES

Regression/ANOVA

Can think of \((X, Y)\) as \((iid)\)

Can condition on \(X\), and...

See Wu in The Annals 1990 (?)

NESTED MODELS

\[ \theta, \ldots, \theta_k \text{ iid } G(\cdot) \]

\[ Y_k | \theta_k \sim \mathcal{N}(\theta_k, \sigma^2_k) \]

The computer will run—be careful that it answers a (the) question you want.

\[ Y_k \sim \mathcal{N}(\mu, \tau^2 + \sigma^2_k) \]

If \(G\) is \(\mathcal{N}(\mu, \tau^2)\)
SUMMARY (JACKKNIFE)

- Reuses the sample
- First moment — bias / biascorrect
- Second moment — variance

- Susceptible to outliers in $\bar{X}$
- Transforms make a difference
- Don't get full sampling distn.

IS THERE AN ALTERNATIVE?

Variance stability symmetry
"match"

YES, THE BOOTSTRAP (US ONE).