Increase sample size; StdErr(\bar{x}) increases!

Actual problem: Alex B was measuring the effect on gum tissue of a particular method of doing a crown preparation. This was a pilot dataset.

- Upper right first molar of volunteers:
  - make a cast of tooth and gum before crown prep
  - do crown prep; wait a little while
  - make a second cast of tooth and gum
  - make digital 3-D images of “before” and “after” casts,
  - align digital images using fixed surface of tooth
  - compute change in gum height, (after) - (before)

- This measurement is, for practical purposes, without error

- Alex B considered a 46.5 mm length of gum between two landmarks.

Design Question: At how many locations in this 46.5 mm length should Alex measure?
**Design Question:** At how many locations in this 46.5 mm length should Alex measure?

**Other facts:** Comparing crown prep vs. no prep

$\Rightarrow$ comparison is between persons

- Although measurements are, in effect, without error, they are costly in time.
- Alex provided a dataset with 11 teeth (??) measured before and after at (??) 11 locations.
- I fit a bunch of spatial models to this dataset and ended up with this model:

$$\begin{align*}
Y_{si} &= \mu_i + \varepsilon_{si} \\
\bar{s} &= \frac{\sum (i-1)}{n-1} \\
&= 46.25 \\
i &= 1, \ldots, n
\end{align*}$$

$$\text{Cov}(\varepsilon_{si}) = \sigma^2 \left( \exp(-d_{ij}/22.5) \right)$$

$\sigma^2 = 95^2 \mu m$

$d_{ij} =$ distance (mm) between two measurements $= \left[ 46.25(i-j)/(n-1) \right] \text{mm}$.

**Very important:** As you take more measurements (as n increases), adjacent measurements are closer together.
Because I like to do dumb simple things for power calculations, I chose to use \( \bar{Y}_{i,n} \) as the summary of the \( n \) measures from subject \( i \).

- Easy to show: \( \text{Var}(\bar{Y}_{i,n}) = \frac{\sigma^2}{n} \left( n + 2 \sum_{i \neq j} \text{corr}(\frac{X_i - \bar{X}}{2}, \frac{X_j - \bar{X}}{2}) \right) \)

<table>
<thead>
<tr>
<th>Design</th>
<th># measures</th>
<th>( \text{Var}(\bar{Y}_{i,n}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Diagram" /></td>
<td>2</td>
<td>0.560 ( \sigma^2 )</td>
</tr>
<tr>
<td><img src="image2.png" alt="Diagram" /></td>
<td>3</td>
<td>0.5208 ( \sigma^2 )</td>
</tr>
<tr>
<td><img src="image3.png" alt="Diagram" /></td>
<td>4</td>
<td>0.5185 ( \sigma^2 )</td>
</tr>
<tr>
<td><img src="image4.png" alt="Diagram" /></td>
<td>5</td>
<td>0.5218 ( \sigma^2 )</td>
</tr>
<tr>
<td><img src="image5.png" alt="Diagram" /></td>
<td>6</td>
<td>0.5345 ( \sigma^2 )</td>
</tr>
</tbody>
</table>

Note: The diagram shows the correlation between measures.
This result does not depend on the choice of constants (22.5, 46.25).

It is not specific to this correlation function $e^{-d/\theta}$

- I got the same qualitative result using:
  
  \[ \text{corr}(i, j) = e^{-d_{ij}/\theta^2} \]
  
  \[ \text{corr}(i, j) = 1 - d_{ij}/46.25 \]

Unable to find any errors in my work, I asked several colleagues and was directed to


which proves this result for the covariance function

\[ \text{cov}(y_s, y_s') = \sigma^2 e^{15 - s} \]

They also show that if you leave out $\sigma^2$, this result no longer holds, i.e. $\text{Var}(\hat{\rho}_n)$ decreases monotonically in $n$, though as $n$ becomes large $\text{Var}(\hat{\rho}_n)$ becomes effectively flat (Hocq PG (1961), Ann. Math. Stat. 32: 1042-1047).

Jargon (e.g. N. Cressie’s spatial book): “Infill Asymptotics”
How can this possibly be true?

Morris and Ebey offer no intuition except to note that “as ... n increases, the serial correlation between adjacent observations increases.”

That is (using my correlation function):

$$\text{correlation (i, i+1) } = \exp\left(-\frac{46.25}{22.5} \frac{1}{n-1}\right)$$

which increases as n increases.

I conjecture that a similar result could be shown for all spatial problems.