When Validation Fails: Analysis of Data from an Imperfect Test Chamber

Kendal H. Ferguson
April, 2019
Aerosol Test Chamber

Chamber Test Locations

Height

Width
We have seven SUT locations, but only one referee location.

We validate chamber if we show homogeneity across all locations.

To validate, we conduct several trials with a filter sampler at each location.

Each filter sampler collects fluorescein during each trial. An instrument counts particles from each filter. Counts are subject to measurement error.

Compare the counts at the locations to assess homogeneity.
Model of Spatial Concentration Variability

True Concentration:

\[ \log(\text{Conc}_{lt}) = \log(\text{MedConc}_t) + \sum_{l=1}^{L} \beta_l X_l + \varepsilon_{lt} \]

Measured Concentration:

\[ \log(\text{MeasConc}_{lt}) = \log(\text{MedConc}_t) + \sum_{l=1}^{L} \beta_l X_l + \varepsilon_{lt} + \delta_{lt} \]

\( \varepsilon = \) True deviation from expected logged concentration

\( \delta = \) Measurement error.

Problem: Error terms are confounded!
\[ \log(\text{MedConc}_t) \sim N(\log(100), \text{stdev} = 0.1) \]

\[ \beta_I \sim N(0, \text{stdev} = 0.5) \]

Prior distributions on original scale (conc = 100):
Priors 2

\[ \varepsilon_{lt} \sim N(0, \sigma^2_{\varepsilon}) \]
\[ \varepsilon_{lt} + \delta_{lt} \sim N(0, \sigma^2_{\varepsilon+\delta}) \]
\[ \delta_{lt} \sim N(0, \sigma^2_{\delta}) \]

\[ \sigma^2_{\varepsilon+\delta} \sim \text{SI} \chi^2(\nu = 5, \sigma = 0.085) \]
\[ \sigma^2_{\delta} \sim \text{SI} \chi^2(\nu = 12, \sigma = 0.05) \]
\[ \sigma^2_{\varepsilon} \sim \text{SI} \chi^2(\nu = 4.57, \sigma = 0.071) \]
Error Variance Priors
Error Variance Joint Distribution
Data Generation

\[ \log(\text{MedConc}_t) = \log(100) \]

\[ \sigma^2 = 0.004 \]

\[ \sigma^2 = 0.003 \]

<table>
<thead>
<tr>
<th>Location</th>
<th>( \beta )</th>
<th>Percent of Chamber Median</th>
<th>Percent of Location 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.03</td>
<td>103.0</td>
<td>100.0</td>
</tr>
<tr>
<td>2</td>
<td>-0.05</td>
<td>95.1</td>
<td>92.3</td>
</tr>
<tr>
<td>3</td>
<td>-0.03</td>
<td>97.0</td>
<td>94.2</td>
</tr>
<tr>
<td>4</td>
<td>-0.01</td>
<td>99.0</td>
<td>96.1</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>104.1</td>
<td>101.0</td>
</tr>
<tr>
<td>6</td>
<td>0.07</td>
<td>107.3</td>
<td>104.1</td>
</tr>
<tr>
<td>7</td>
<td>0.09</td>
<td>109.4</td>
<td>106.2</td>
</tr>
<tr>
<td>8</td>
<td>0.12</td>
<td>112.7</td>
<td>109.4</td>
</tr>
</tbody>
</table>

Maximum deviation during 69% of trials will be within 0.2 logs (\( \sim \pm20\% \)) of Location 1. We want 80% of trials within 0.2 logs, therefore the chamber should NOT be validated.
For each sample from the posterior

betaDiff: Vector of expected differences:
\[
\begin{bmatrix}
\beta_2 - \beta_1 \\
\vdots \\
\beta_8 - \beta_1
\end{bmatrix}
\]

covMatrix: Covariance matrix for differences:
\[
\begin{bmatrix}
2\sigma_{true}^2 & \cdots & \sigma_{true}^2 \\
\vdots & \ddots & \vdots \\
\sigma_{true}^2 & \cdots & 2\sigma_{true}^2
\end{bmatrix}
\]

Library(mvtnorm)

pmvnorm(lower = -x, upper = x, mean = betaDiff, sigma = covMatrix)
For 85.9% of observations, both differences are within 0.2 logs
Probability of Sufficient Homogeneity: \( x = 2.74\% \)

Validation Failed!
Aerosol Test Chamber

Chamber Test Locations

Height

Width
Representing the posterior

- Use maximum likelihood to approximate the density of the 7 location deviations as multivariate normal.
- Use scaled inverse chi square to represent the true error variance.
- These will be used as “priors” for our logistic regressions. But we won’t update these priors. They just express our best knowledge.
Testing the SUTs

Chamber Validated:

\[ P(\text{Detect}_{lt}) = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1 \cdot \text{Concl}_{lt})}} \]

Chamber Not Validated:

\[ P(\text{Detect}_{lt}) = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1 \cdot \log(\text{Concl}_{lt} + \sum_{l=1}^{L} (\beta_l - \beta_1) X_l + \varepsilon_{lt}))}} \]
Implications of Simulated Data 1
(Back-Up Slide)

We want to know the implications of our location parameters ($\beta$) and our true variance of logged concentration ($\sigma^2_{true}$) on the probability of the concentration at all 7 SUT locations being within 0.2 logs of the concentration at the referee location.

The difference in logged true concentration between the first SUT location (Location 2) and the referee location (Location 1) can be expressed as:

$$diff1 = (\mu + \beta_2 + \epsilon_2) - (\mu + \beta_1 + \epsilon_1)$$

Because both $\epsilon_1$ and $\epsilon_1$ are normally distributed with the same variance, $diff1$ (and all the other seven differences between the SUT locations and the referee location) is distributed as:

$$diff1 \sim N(\beta_2 - \beta_1, variance = 2\sigma^2_{true})$$

The covariance between any two such differences, is obtained as follows (where concentration $x_1 = u + B_1 + e_1$, and other concentrations are notated similarly).

$$Cov(diff1, diff2) = E[(x_2 - x_1)(x_3 - x_1)] - E(x_2 - x_1)E(x_3 - x_1)$$
$$= E[x_2x_3 - x_1x_3 - x_1x_2 + x_1^2] - (\beta_2 - \beta_1)(\beta_3 - \beta_1)$$
$$= E[x_2x_3 - x_1x_3 - x_1x_2 + x_1^2] - (\beta_2 - \beta_1)(\beta_3 - \beta_1)$$
$$= \beta_2\beta_3 - \beta_1\beta_3 - \beta_1\beta_2 + E(x_1^2) - \beta_2\beta_3 + \beta_1\beta_3 + \beta_1\beta_2 - \beta_1^2$$
$$= \sigma^2_{true} + \beta_1^2 - \beta_1^2 = \sigma^2_{true}$$

This finding derives from the following

$$E(xy) = E(x)E(y), if \ x \ and \ y \ are \ independent$$
$$E(x^2) = \sigma_x^2 + E(x)^2, based \ on \ rearrangement \ of \ variance \ formula$$

Based on the variance and covariance terms shown above, we have a multivariate normal distribution for the seven differences, with each variance term (along the diagonal) being $2\sigma^2_{true}$ and each covariance term (off the diagonal) being $\sigma^2_{true}$.

We can evaluate the multivariate normal distribution to determine the probability that all differences will be between $-0.2$ logs and $0.2$ logs. We have determined that the concentrations are sufficiently homogenous (and close to the Location 1 concentration) if this requirement is met.

This evaluation is simple with the pmvnorm() function from the mvtnorm package in R:

```r
pmvnorm(lower = -x, upper = x, mean = betaDiff, sigma = covMatrix1), where betaDiff is a vector of the seven expected differences and covMatrix1 is a covariance matrix with the variance and covariance terms described above.
```