Statistical Modeling 3

Bias correction and normalization

Five technical replicates
Five technical replicates

More than location and scale changes!

smooth histogram
Median shifts do not solve the problem!

There are non-linear effects!
Quantile normalization

<table>
<thead>
<tr>
<th>Original</th>
<th>Order</th>
<th>Averaged</th>
<th>Re-order</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 4 4 5</td>
<td>2 4 3 5</td>
<td>3.5 3.5 3.5 3.5</td>
<td>3.5 3.5 5.0 5.0</td>
</tr>
<tr>
<td>5 14 4 7</td>
<td>3 8 4 5</td>
<td>5.0 5.0 5.0 5.0</td>
<td>8.5 8.5 5.5 5.5</td>
</tr>
<tr>
<td>4 8 6 9</td>
<td>3 8 4 7</td>
<td>5.5 5.5 5.5 5.5</td>
<td>6.5 5.0 8.5 8.5</td>
</tr>
<tr>
<td>3 8 5 8</td>
<td>4 9 5 8</td>
<td>6.5 6.5 6.5 6.5</td>
<td>5.0 5.5 6.5 6.5</td>
</tr>
<tr>
<td>3 9 3 5</td>
<td>5 14 6 9</td>
<td>8.5 8.5 8.5 8.5</td>
<td>5.5 6.5 3.5 3.5</td>
</tr>
</tbody>
</table>

Densities are forced to be identical

Five technical replicates

```
log2(pms[, 2])
Frequency
```

Five technical replicates

```
log2(qpms[, 2])
Frequency
```
Differential expression can be preserved
“Essentially, all models are wrong, but some are useful”

George E.P. Box

A biochemical experiment

Michaelis-Menten equation

\[ V = \frac{V_{\text{max}} \times C}{K + C} \]

- \( V \) = initial velocity
- \( C \) = concentration
- \( V_{\text{max}} \) = maximum velocity
- \( K \) = rate constant
A biochemical experiment

\[ V = \frac{V_{\text{max}} \times C}{K + C} \]

\[ \Rightarrow \quad \frac{1}{V} = \frac{K + C}{V_{\text{max}} \times C} \]

\[ = \frac{K}{V_{\text{max}} \times C} + \frac{1}{V_{\text{max}}} \]

\[ \Rightarrow \quad \frac{1}{V} = \left( \frac{1}{V_{\text{max}}} \right) + \left( \frac{K}{V_{\text{max}}} \right) \times \left( \frac{1}{C} \right) \]

Model:

\[ \frac{1}{V} = \beta_0 + \beta_1 \left( \frac{1}{C} \right) + \text{error} \]

Intercept \quad 0.00697
Slope \quad 0.00022

\[ \hat{V}_{\text{max}} = \frac{1}{\text{Intercept}} = \frac{1}{0.00697} = 143 \]

\[ \hat{K} = \text{Slope} \times \hat{V}_{\text{max}} = 0.031 \]
A biochemical experiment

which is more reasonable?

\[
\frac{1}{V} = \beta_0 + \beta_1 \left( \frac{1}{C} \right) + \text{error}
\]

\[
V = \frac{V_{\text{max}} \times C}{K + C} + \text{error}
\]

Eleven probes from one spiked-in gene
\[ Y_{ij} = \beta_j + \theta_i \phi_j + \varepsilon_{ij} \quad \text{var}(\varepsilon_{ij}) \propto \theta_i \phi_j \]

Model fit to two probes

**probe 5**

\[ \log_2(311+21.2 \text{ concentration}) \]

**probe 7**

\[ \log_2(72.4+7.55 \text{ concentration}) \]