Academic Forum 35 (2017–18)

Assessing Statistical Models for a Stress Hormone in Saliva

Michael Lloyd
Professor of Mathematics

Emilie Beltzer
Associate Professor of Psychology

Abstract

Various models were assessed for predicting cortisol concentration in psychology students' saliva based on light absorbance values.

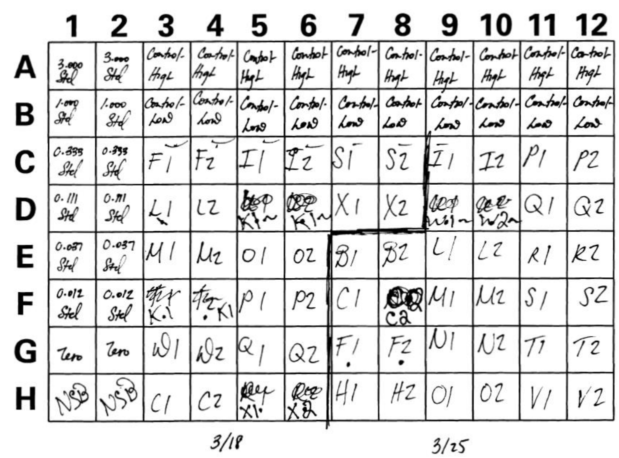
Institutional Review Board Statement

This paper uses data from the following studies which were approved by the Henderson State University Institutional Review Board:

- The analysis of cortisol and alpha-amylase responses to a laboratory acute stressor in relation to individual methylated DNA levels and personality traits.
- Psychological and Physiological Stress in Intermediate Algebra Students: Relating Anxiety, Preparation, and Performance.

The accompanying diagram shows the layout of the first microtiter plate from the Intermediate Algebra study conducted in spring 2016. The other plates from this study had problems as explained in Lloyd, M. (2016-17), “Stressed-out Intermediate Algebra Students,” *The Academic Forum*, 34, 13-23.

- The standard concentration values (Std) were in Cells A1..H2.
- The cortisol High and Low cells (A3..B12) are used for quality control.
- The student values were in Cells C3..H12.



Microtiter Math Plate 1 Layout

Here are the raw optical densities (OD) for every cell on this plate:

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.105	0.140	0.342	0.241	0.295	0.371	0.318	0.292	0.352	0.392	0.438	0.442
B	0.247	0.316	1.054	1.154	1.311	1.082	1.174	1.331	1.208	1.204	1.185	1.284
C	0.504	0.613	0.868	1.458	1.306	1.365	0.762	0.928	1.234	1.234	1.030	0.684
D	0.901	1.208	2.013	0.957	0.708	0.777	0.874	1.020	1.297	0.767	1.184	0.933
E	1.556	1.418	1.018	1.461	1.360	1.147	1.239	1.231	2.014	1.151	0.836	1.120
F	1.531	1.613	0.905	0.881	1.046	1.161	1.903	0.633	0.590	0.707	0.776	0.474
G	1.912	1.931	1.076	1.276	0.686	0.853	0.943	1.080	0.997	0.984	0.925	1.130
H	1.916	1.877	1.120	1.142	0.932	0.707	0.570	1.189	0.566	0.563	0.389	0.085

The zero concentration optical density (B_0) is estimated by averaging the two zero values:

$$B_0 = \frac{B_{01} + B_{02}}{2} = \frac{1.912 + 1.931}{2} = 1.9215$$

The non-specific binding optical density (NSB) is estimated by averaging the two corresponding values:

$$NSB = \frac{NSB_1 + NSB_2}{2} = \frac{1.916 + 1.877}{2} = 1.8965$$

All the other optical densities were converted to the fraction bound (Fb) by using the following formula. (Some references refer to Fb as the percent bound, but it is referred to as the "fraction bound" because we did not multiply it by 100.)

$$Fb = \frac{NSB - OD}{B_0} = \frac{1.8965\ - OD}{1.9215}$$

Here are the standard concentration values from Math Plate 1:

- There were two replications per standard concentration.
- $Conc$ is concentration of cortisol in micrograms per deciliter.
- OD is the raw optical density value from the plate.
- Fb is the fraction bound

Conc	OD	Fb
3.000	0.105	0.932
1.000	0.247	0.858
0.333	0.504	0.725
0.111	0.901	0.518
0.037	1.556	0.177
0.012	1.531	0.190
3.000	0.140	0.914
1.000	0.316	0.823
0.333	0.613	0.668
0.111	1.208	0.358
0.037	1.418	0.249
0.012	1.613	0.148

The most common models for predicting the concentration based on optical density are logistic, Gompertz, and the cubic spline. The *Bioassay Analysis using R* mentions the first two models as being widely used for sigmoidal dose-response curves; the *Good ELISA Practice Manual* only mentions cubic spline and 4-parametric logistic regression.

We considered these three R libraries for doing Enzyme-Linked Immunosorbent Assay (ELISA):

- Analysis of Dose-Response Curves (drc)
- n-Parameter Logistic Regression (nplr)
- Nonlinear Calibration (nCal)
  did not try, 23-page manual

We used the drc because it appeared to be the most popular and had the most features. We have used nplr before because of its simplicity, but, it did not work with the version of R we were using in fall 2017. We did not try the nCal library.

There are five possible Log-Logistic (LL) models. The variable x is the dosage (concentration), and $a..f$ are the parameters.

- 2-parameter $Fb = \frac{1}{1+\exp(b(\log x - \log e))}$
- 3-parameter lower $Fb = 0 + \frac{d-0}{1+\exp(b(\log x - \log e))}$
- 3-parameter upper $Fb = c + \frac{1-c}{1+\exp(b(\log x - \log e))}$
- 4-parameter $Fb = c + \frac{d-c}{1+\exp(b(\log x - \log e))}$

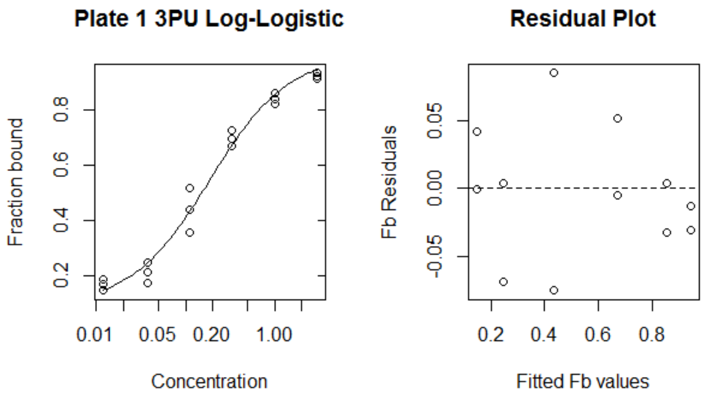
- 5-parameter $Fb = c + \frac{d-c}{\left(1+\exp\left(b(\log x - \log e)\right)\right)^{f}}$

First four models are nested in the 5-parametric, so they can be compared using the Akaike Information Criterion (AIC).

The optimal model from the LL family was the 3-parametric Upper Logistic.

Coefficients

$b = -0.98357$
$c = 0.09076$
$e = 0.18516$



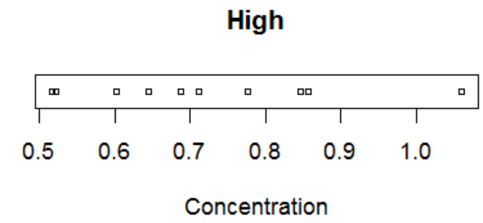
The accompanying table shows the Hi and Low Values from Math Plate 1. The concentrations were estimated using the above LL model. ConcSeHi and ConcSeLo are the standard errors for the high and low concentrations, respectively.

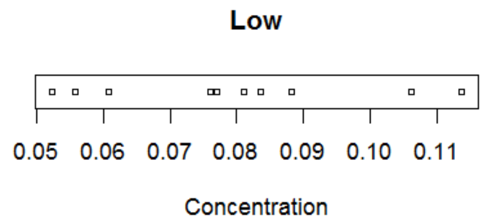
OdHi	OdLo	FbHi	FbLo	ConcHi	ConcSeHi	ConcLo	ConcSeLo
0.342	1.054	0.809	0.438	0.712	0.123	0.114	0.029
0.241	1.154	0.862	0.386	1.061	0.217	0.088	0.025
0.295	1.311	0.833	0.305	0.847	0.157	0.056	0.019
0.371	1.082	0.794	0.424	0.645	0.108	0.106	0.028
0.318	1.174	0.821	0.376	0.776	0.138	0.084	0.025
0.292	1.331	0.835	0.294	0.857	0.159	0.052	0.019
0.352	1.208	0.804	0.358	0.688	0.117	0.076	0.023
0.392	1.204	0.783	0.360	0.602	0.099	0.077	0.024
0.438	1.185	0.759	0.370	0.522	0.083	0.081	0.024
0.442	1.284	0.757	0.319	0.516	0.082	0.061	0.020

The salimetrics document in the references recommends assessing the quality control using the coefficient of variation of the predicted concentrations:

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

The Intra-assay High concentrations and Low concentrations for Math Plate 1 had CVs of 23% and 25%, respectively. These are considered unacceptably large because they are greater than 10%. Hence, we have concerns about the internal consitency of the student optical density values on this plate.

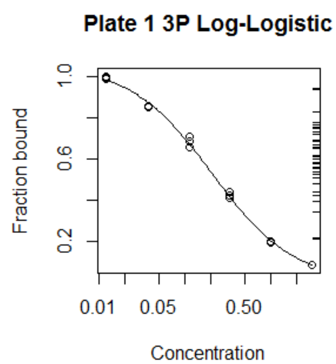


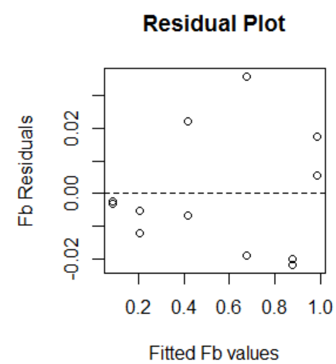


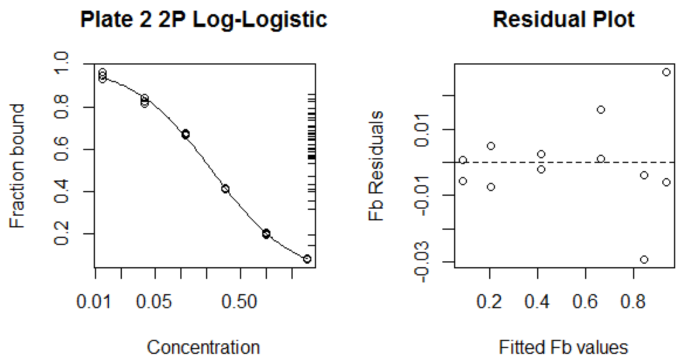
The following standard values were collected by Dr. Beltzer in fall 2016. We did not use the NSB cells, so we used the simpler transformation $FB = OD/B_0$ which will make her sigmoidal models decreasing instead of increasing. The B_0 values for Plate 2 were suspicious because they were both precisely 1.146. There were only two Hi and two Lo cells for her per plates, so an intra-assay could not be performed. However, an inter-assay could be done with at least ten plates.

Plate 1			Plate 2		
Conc	OD	Fb	Conc	OD	Fb
3.000	0.096	0.082	3.000	0.099	0.086
1.000	0.232	0.198	1.000	0.241	0.210
0.333	0.514	0.438	0.333	0.480	0.419
0.111	0.771	0.656	0.111	0.778	0.679
0.037	1.006	0.857	0.037	0.963	0.840
0.012	1.179	1.004	0.012	1.107	0.966
3.000	0.097	0.083	3.000	0.092	0.080
1.000	0.224	0.191	1.000	0.227	0.198
0.333	0.480	0.409	0.333	0.475	0.414
0.111	0.835	0.711	0.111	0.761	0.664
0.037	1.004	0.855	0.037	0.934	0.815
0.012	1.165	0.99	0.012	1.069	0.933

The optimal models for the Log-Logistic family for these two plates were the 3-parametric lower and 2-parametric, respectively. The rug plots on the edges of the scatter plots are the fraction bound student values.



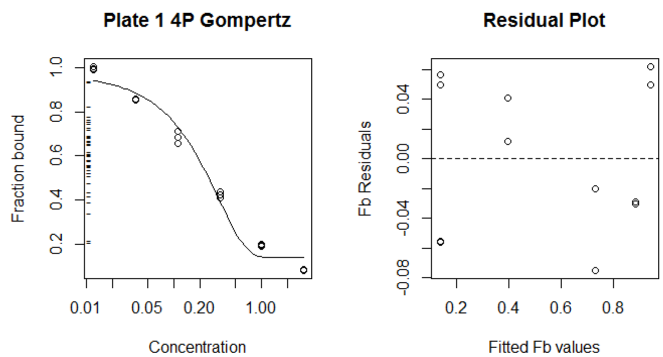


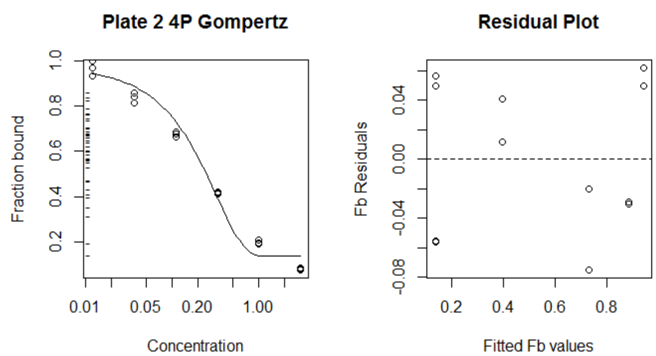


There are four possible Gompertz models: 2, 3, 3u, 4-parameters

$$Fb = c + (d - c)\exp(-\exp b(x - e))$$

The optimal models for both plates used all four parameters, but the shapes were described well by this model.

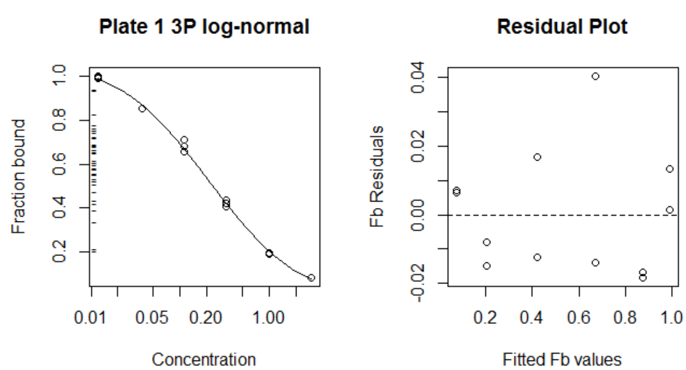




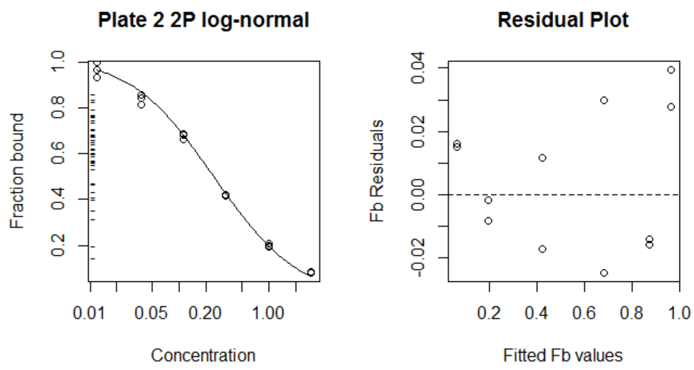
The Log-normal (LN) model has four possible forms: 2, 3, 3u, 4-parameters

$$Fb = c + (d - c)\Phi\left(b\,(\log x - \log e)\right)$$

The function Φ is the standard normal cumulative density function. The optimal models from this family for the two plates



were the 3-parametric lower and 2-parametric, respectively.

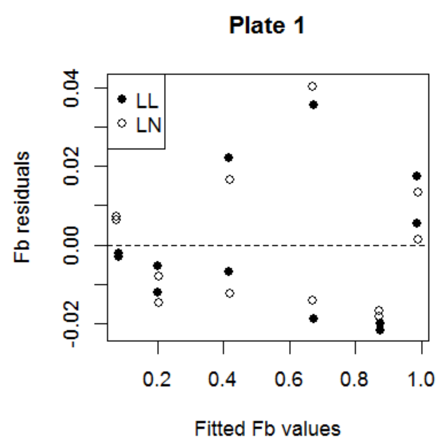


We also considered the Weibull1 and Weibull2 models (1.2, 1.3, 1.3U, 1.4, 2.2, 2.3U-parameters), but their optimal models each required four parameters per plate.

$$f_1(x) = c + (d - c)\exp(-\exp b(\log x - \log e))$$
$$f_2(x) = c + (d - c)\big(1 - \exp(-\exp b(\log x - \log e))\big)$$

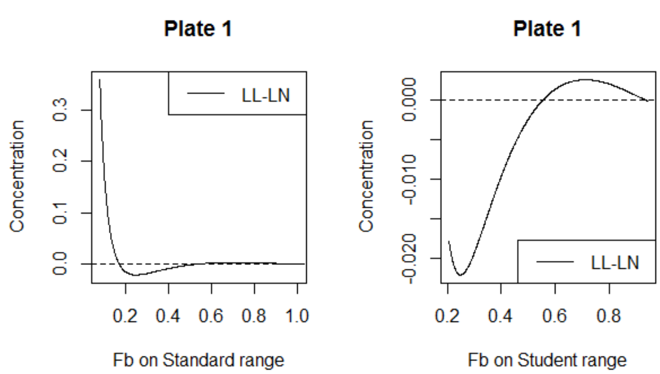
Of all the models considered, LL and LN were the best because they tended to use fewer parameters and have small residuals. Which of these two models is better? The residuals are about the same for Plate 1, but the LL model had a slightly smaller maximum absolute residual.

<table>
<tr><th></th><th>Median |Resid|</th><th>Max |Resid|</th></tr>
<tr><td>Log-logistic</td><td>0.015</td><td>0.036</td></tr>
<tr><td>Log-normal</td><td>0.014</td><td>0.040</td></tr>
</table>



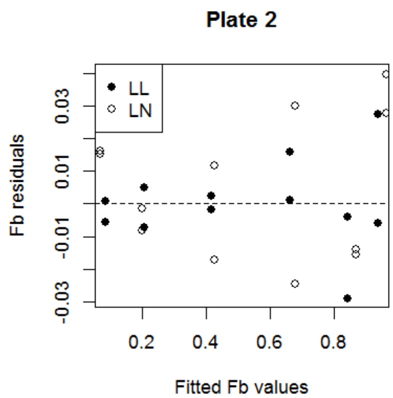
The difference between the models was relatively large when Fb was smaller than around 0.15, but there were no student values in that range.

The drc library supplied a function for computing the standard error for each model separately. However, we did not know how to compute it for the difference between the models ($LL - LN$).

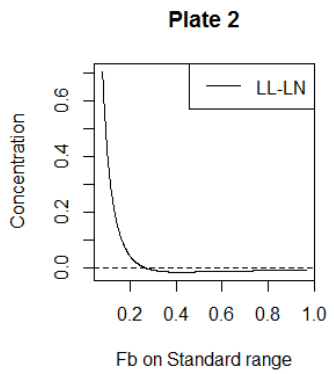


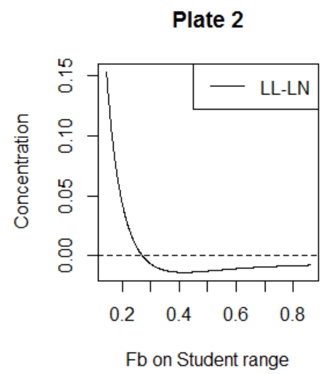
The LL model tended to have smaller residuals for Plate 2.

	Median |Resid|	Max |Resid|
Log-logistic	0.005	0.029
Log-normal	0.016	0.040



The difference between the models for Plate 2 was relatively large for Fb values less than about 0.2, but this only affected two of the 39 students on this plate. The models were probably more different for Plate 2 than Plate 1 because the model for Plate 2 used two parameters while the one for Plate 1 used three.





Observations

- The LL and LN families avoided overfitting compared to other models that we considered, yielding the simplest models with three and two parameters for Plates 1 and 2, respectively.
- The LL model tended to have smaller residuals than the LN for our data.
- The LL model is more versatile than the LN model because it has a 5-parametric version which includes the asymmetric parameter f.
- The difference between the LL and LN models was more pronounced for Fb values smaller than around 0.2. This corresponds to concentrations more than about 1 μg/dL cortisol, so we need more standard points for concentrations between 1 μg/dL and 3 μg/dL to determine which model is better.

Acknowledgments

- Aneeq Ahmed (coinvestigator for psychology study)
- Chemistry and psychology student workers
- David Bateman (coinvestigator for mathematics study)
- Holly Morado (principal investigator for mathematics study)
- Mathematics and Ellis College administrative assistants

References

- "Good ELISA Practice Manual" Food &Feed Analysis, a division of R-Biopharm AG, https://food.r-biopharm.com/
- "Inter- and Intra-Assay Coefficients of Variability" https://www.salimetrics.com/
- Ritz, C., Baty, F., Streibig, Jens C. (2005) "Bioassay Analysis using R," *Journal of Statistical Software*, 12(5) http://www.jstatsoft.org/
- Ritz, C., Baty, F., Streibig, J. C., Gerhard, D. (2015) Dose-Response Analysis Using R PLOS ONE, 10(12), e0146021
- "SALIVARY CORTISOL" https://www.salimetrics.com/

Biographical Sketch

Michael Lloyd graduated cum laude and in the honors program in Chemical Engineering with a B.S. in 1984. He accepted a position at Henderson State University in 1993 shortly after earning his Ph.D. in Mathematics (Probability Theory) from Kansas State University. He has presented papers at meetings of the Academy of Economics and Finance, the American Mathematical Society, the Arkansas Conference on Teaching, and the Southwest Arkansas Council of Teachers of Mathematics. He has been an active member of the Mathematical Association of America since 1993, earned 18 hours in computer science, and has been an Advanced Placement statistics consultant since 2002.