

Dr. Susan Love
Hormone Data
Final Results

From: Tiffany Himmel and Mahtash Esfandiari
UCLA Statistical Consulting Center

1. Data Preparation

The data presented are ducal lavage levels of overall protein as well as 4 hormones along with several demographic covariate variables. Our initial graphical analysis of the data showed that the outcome variables were highly skewed to the right, therefore a log base 10 transformation was needed on protein and all 4 hormones in order to meet our most basic normality assumption.

One of the covariate also needed transformation. The variable Gail was split into 3 groups: High Risk, Low Risk, Previously Had Breast Cancer. Additionally the variables menstrual status and menopausal status were combined to form a 3 factor variable: Pre-Menopausal Follicular-Phase, Pre-Menopausal Leutial-Phase and Post-Menopausal.

2. Kinds of Missing

The data presented an interesting challenge in terms of missing data. Most of our covariates have little missing data but the response variables had many missing. The first flavor of missing are those values which the machine determines to be "below detectable level" or BDL. These are the most influential missing data and have to be dealt with using a Monte Carlo (see section 5). The other missing data are either missing because they are lost (unavoidable but are very few and assumed to be at random because they are uncorrelated with outcome variables), because they were done by another lab (to be dropped), were lost to follow-up (drop both before and after to help offset bias).

3. Initial Model

The first model run was a basic multiple linear regression. These separate models allowed us to eliminate highly insignificant covariates early in the analysis. By slowly eliminating variables that had low significance, we were able to get down to 5 main effects (Age, Gail, NAF, Menopausal Status, and Visit). We were also interested in 3 two-factor interactions (Age vs Visit, Gail vs Visit and NAF vs Visit).

4. Random Effects Model

Because we have multiple samples from each woman, we need to model the relationship between them. Therefore, a random intercept model is used, such that each woman is assumed to have a random intercept that is the average of her data-points. The advantage that this give us is that we can now find the ICC or "Intra-Class Correlation". By looking at the proportion of the variance explained by the random intercepts, we can see how well the data "clumps" from woman to

woman. A large value of ICC means large amounts of clumping (each woman is very different). As small value means each woman is more or less the same.

5. Simulating BDL Data

Data that was BDL had to be dealt with using a Monte Carlo technique. By assuming the normal curve on our data, the BDL missing values can be sampled from the lower tail of the normal (below the detection limit). By simulating this 10,000 times, we are able to find the correct parameter estimates as well as correct standard errors. Because the standard error of the estimates is Gaussian and the distribution of the estimates in the Monte Carlo is Gaussian, the errors can be added in quadrature.

6.1 Results - Tables df=44

Protein	Estimate	Standard Error	t-value	p-value	
Intercept	-0.62455814	1.1831365	-0.52788342	0.30011625	
Ages	0.02576227	0.01971265	1.30689018	0.09901968	'
Gail (High Risk)	-0.32655188	0.38647294	-0.84495404	0.20135387	
Gail (Breas Cancer)	-0.06417714	0.45543962	-0.14091252	0.44429111	
NAF (Yes)	0.27513327	0.44893014	0.61286433	0.27156049	
MenStat (Pre-F)	0.0362124	0.37371339	0.09689887	0.46162337	
MenStat (Pre-L)	0.68868067	0.3832491	1.79695314	0.0396046	*
Visit 2	1.00195405	1.18490045	0.84560188	0.20117501	
Ages : Visit 2	-0.02248487	0.0214136	-1.05002738	0.14971938	
Gail (High Risk) : Visit 2	0.81502771	0.5272981	1.54566784	0.06467386	'
Gail (Breast Cancer) : Visit 2	0.28180065	0.62975522	0.44747647	0.32836318	
NAF (Yes) vs Visit 2	-0.46932332	0.57346533	-0.81839877	0.20876994	

ESS	Estimate	Standard Error	t-value	p-value	
Intercept	-1.1692215	1.68056137	-0.6957327	0.24512744	
Ages	-0.03560523	0.02812723	-1.2658635	0.10611182	
Gail (High Risk)	0.36683394	0.54206323	0.6767364	0.25105779	
Gail (Breas Cancer)	0.58456871	0.64161577	0.9110884	0.18360669	
NAF (Yes)	-1.18908996	0.64321755	-1.8486591	0.03561687	*
MenStat (Pre-F)	0.0719448	0.51555171	0.1395491	0.44482652	
MenStat (Pre-L)	0.22619692	0.53630557	0.4217687	0.33762372	
Visit 2	-0.22941835	1.65450339	-0.138663	0.44517459	
Ages : Visit 2	0.01719175	0.03008709	0.5713995	0.28531816	
Gail (High Risk) : Visit 2	-1.33507776	0.74136445	-1.8008386	0.03929241	
Gail (Breast Cancer) : Visit 2	-1.29786497	0.88459605	-1.4671838	0.074719	'
NAF (Yes) vs Visit 2	1.11807997	0.80488706	1.3891141	0.08589445	'

DHEAS	Estimate	Standard Error	t-value	p-value	
Intercept	-0.4994395	1.00543376	-0.4967403	0.310922525	
Ages	0.01935733	0.01668734	1.1600009	0.126152134	
Gail (High Risk)	-0.87688835	0.32241139	-2.719781	0.004658314	**
Gail (Breas Cancer)	-0.10697286	0.38128901	-0.2805558	0.390183031	
NAF (Yes)	-0.59929776	0.37521321	-1.5972192	0.058687377	'
MenStat (Pre-F)	0.63831415	0.31143497	2.0495905	0.023197188	*
MenStat (Pre-L)	0.5799349	0.32470776	1.7860211	0.040494187	*
Visit 2	0.09991593	0.99319785	0.1006002	0.460162411	
Ages : Visit 2	-0.01083587	0.0179282	-0.6044036	0.274339748	
Gail (High Risk) : Visit 2	1.22190651	0.44213382	2.7636576	0.004157533	**
Gail (Breast Cancer) : Visit 2	0.4763052	0.5300969	0.8985248	0.186898354	
NAF (Yes) vs Visit 2	0.97400961	0.47724587	2.0408969	0.023644091	*

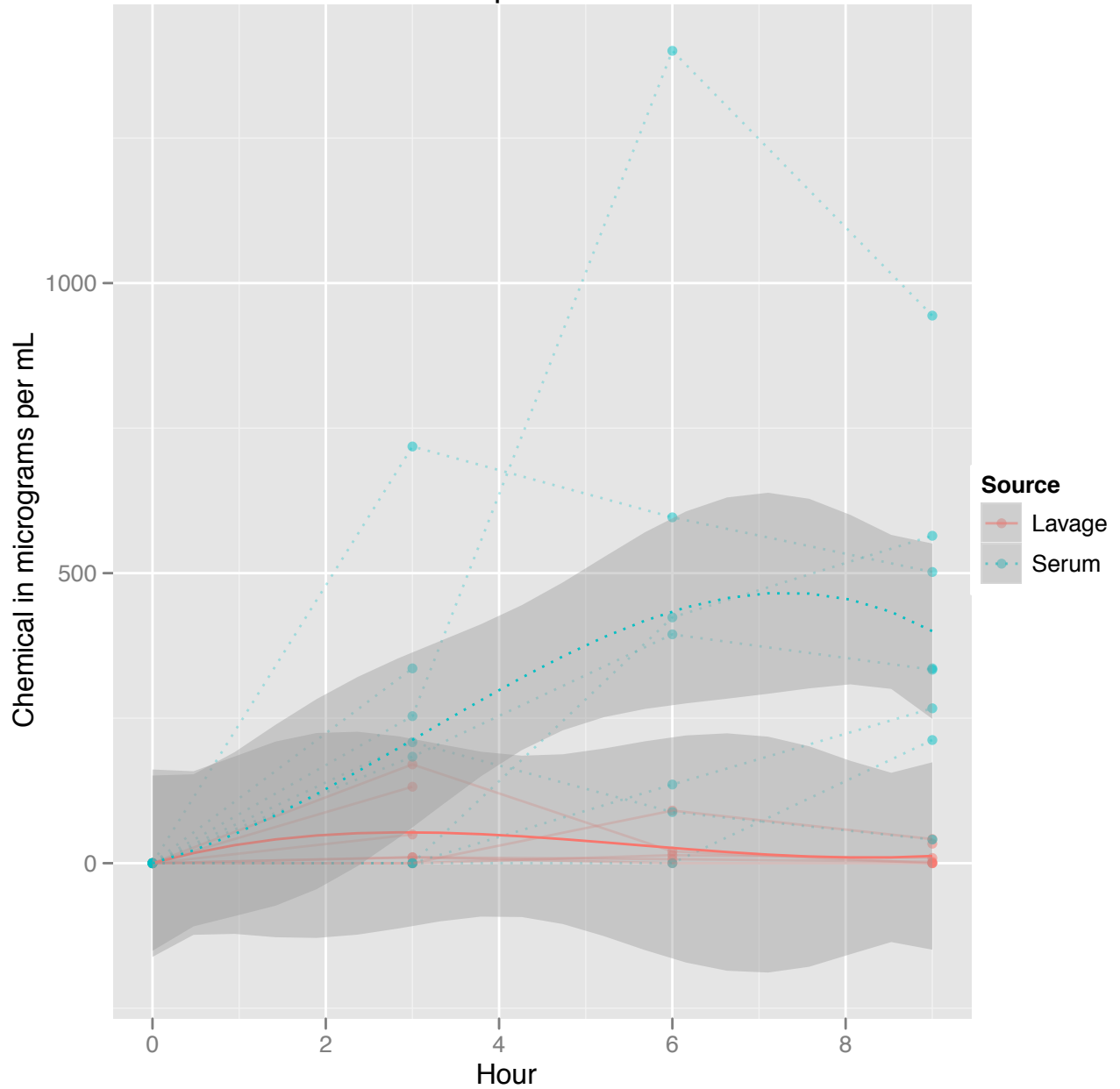
Progesterone	Estimate	Standard Error	t-value	p-value	
Intercept	-2.6712755	0.65975561	-4.0488864	.000103	***
Ages	0.04970362	0.01098731	4.5237281	.0000228	***
Gail (High Risk)	-0.2367464	0.21370518	-1.1078178	.137	
Gail (Breas Cancer)	-0.30938334	0.25419297	-1.21712	.115	
NAF (Yes)	0.0949626	0.24846002	0.3822048	.352	
MenStat (Pre-F)	0.39067037	0.20521177	1.9037425	.0317	*
MenStat (Pre-L)	0.56647431	0.21432623	2.643047	.00567	***
Visit 2	1.41544702	0.65680323	2.1550549	.0183	*
Ages : Visit 2	-0.0312354	0.01187511	-2.6303249	.00586	**
Gail (High Risk) : Visit 2	0.26353579	0.29238301	0.9013375	.186	
Gail (Breast Cancer) : Visit 2	-0.03892143	0.35184392	-0.1106213	.456	
NAF (Yes) vs Visit 2	-0.16306865	0.31624537	-0.5156397	.304	

Estrodiol	Estimate	Standard Error	t-value	p-value	
Intercept	1.259940222	1.82861482	0.689013458	.247	
Ages	-0.012920662	0.03049042	-0.423761397	.337	
Gail (High Risk)	0.488081879	0.58322041	0.836873803	.204	
Gail (Breast Cancer)	0.609490841	0.69392876	0.878319037	.192	
NAF (Yes)	0.586607781	0.68164843	0.860572338	.197	
MenStat (Pre-F)	-0.001786596	0.56275647	-0.003174723	.499	
MenStat (Pre-L)	-0.417369089	0.58516873	-0.713245718	.240	
Visit 2	3.352274992	1.81568886	1.84628273	.0358	*
Ages : Visit 2	-0.049900632	0.03283953	-1.519529407	.0679	'
Gail (High Risk) : Visit 2	-1.062431906	0.80311062	-1.322896101	.0964	'
Gail (Breast Cancer) : Visit 2	-1.247727456	0.96733172	-1.289865136	.102	
NAF (Yes) vs Visit 2	-1.300974095	0.87080403	-1.493991824	.0712	'

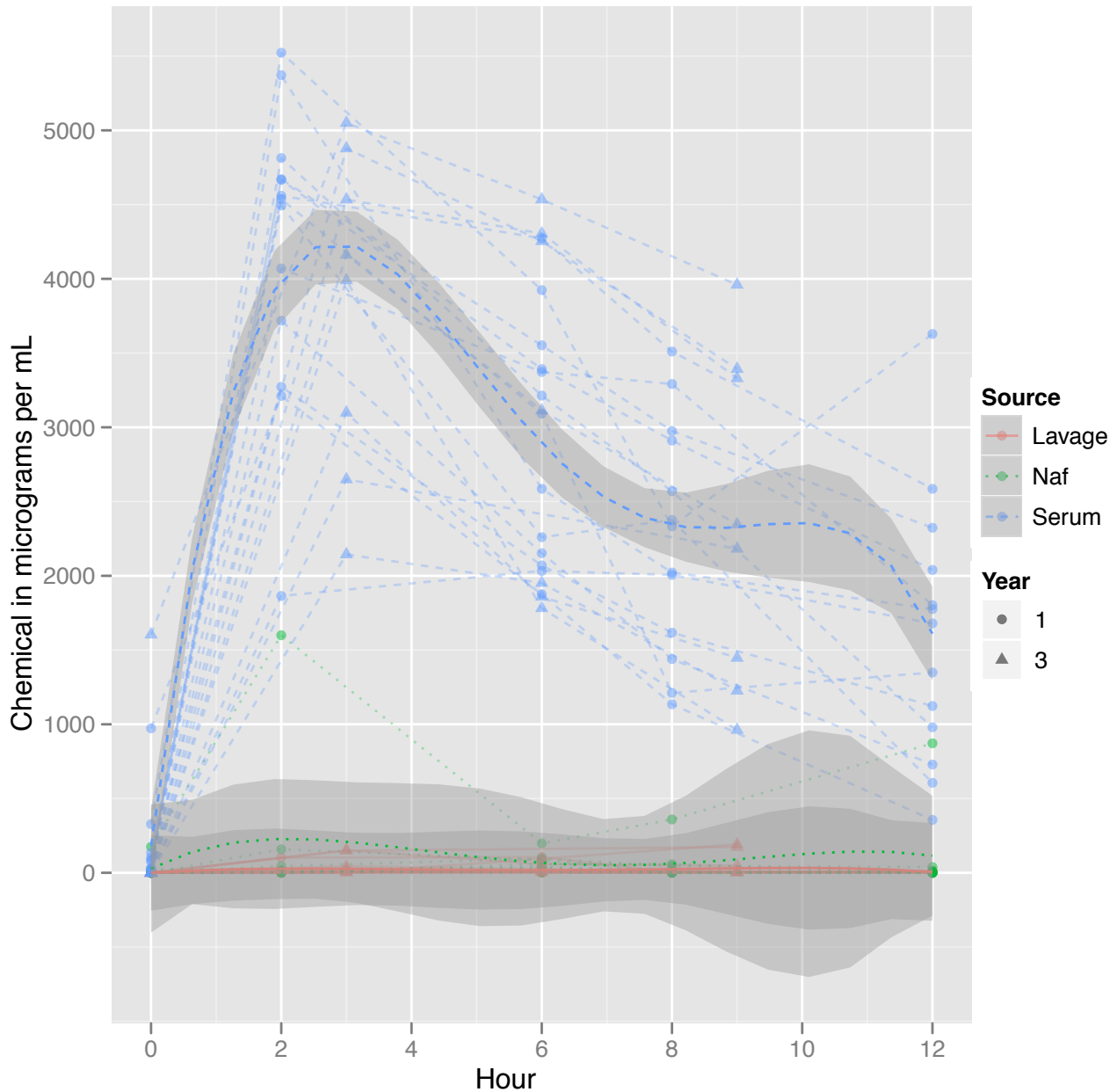
6.2 Results- In Context

The coefficients in the tables above cannot be interpreted directly because of the log base 10 transformations. Here we will give an example result in context so that other coefficients can be interpreted. Age is a significant factor for Progesterone. So to evaluate it's effect, we take $10^{0.04970362} = 1.121253$. This tells us by what factor Progesterone increases for every one year increase in age. That is for every year above the average a woman is, you increase her progesterone level by 12%

Aspirin



Caffeine



Cimetidine

