Biometry Take Home Exam 2

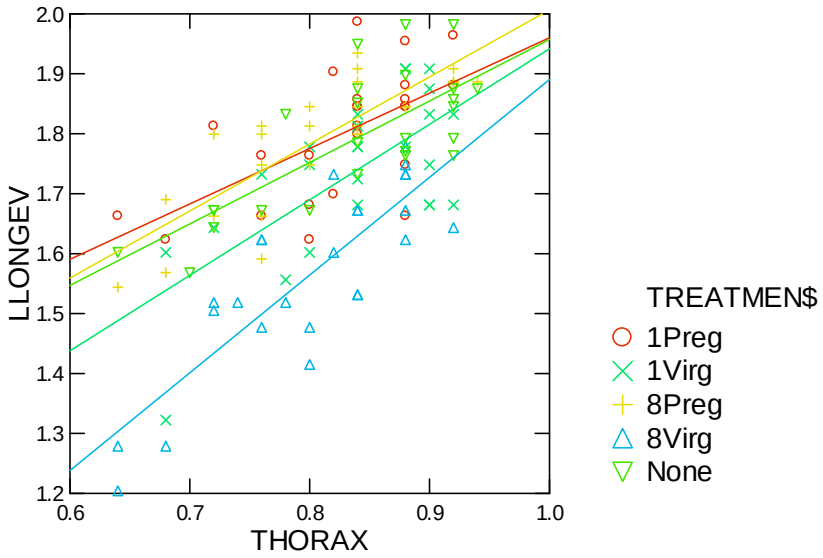
Spring 2010

Name:_____

Instructions: A copy of this exam (exam2.pdf) and data sheets (exam2data.xls) are on the class website. The worksheets are in the same order, left to right, as the exam questions below, top to bottom. In writing up each question, pretend that these are your data and that you're preparing your analysis for a talk at the upcoming CEEB symposium. Attending the symposium are several prominent keynote speakers from outside UK, and you would very much like to impress them with the fact that you would make an excellent future postdoc in their laboratories. With that in mind, be sure to make your analysis as clear as possible. Include in your write up your model statements (from full to whatever model you're left with, reduced or not), your summary statistical tables, graphs that drive home your results, and your conclusions based on the data and hypotheses at hand.

1. Partridge and Farquhar (1981) undertook a study of costs of male reproduction in the fruit fly, *Drosophila melanogaster*. They hypothesized that males would invest more energy in reproduction the more females there were to mate with, and if those females were virgins as opposed to pregnant, and that this increased energy expenditure would shorten their reproductive lifespan. They used 5 experimental treatments: No Females, 1 Pregnant Female, 8 Pregnant Females, 1 Virgin Female, and 8 Virgin Females. Previous research had indicated that male longevity is positively correlated with male body size (thorax length), so they measured male thorax length as a covariate in their data. Your task is use ANCOVA to analyze their data, which are contained in the Longevity vs Mating Number worksheet in your excel file.

1. First, plot the data of Longevity versus Thorax Length for the 5 treatments.



The 5 lines seem roughly parallel. The control and the two pregnant treatments seem on top of one another, with 1Virg below them, and 8Virg lower still. Next, we begin an ANCOVA with the interaction term included, which tests for heterogeneity among regression slopes. Our model statement is:

Log (longevity) = Treatment + Thorax Length + Treatment*Thorax Length

…where Treatment is a fixed categorical variable, and Thorax Length is a fixed continuous variable.

Analysis of Variance					
Source	**Type III SS**	**df**	**Mean Squares**	**F-ratio**	**p-value**
TREATMEN$	0.082	4	0.021	2.991	0.022
THORAX	0.991	1	0.991	144.363	0.000
TREATMEN$*THORAX	**0.043**	**4**	**0.011**	**1.561**	**0.189**
Error	0.790	115	0.007		

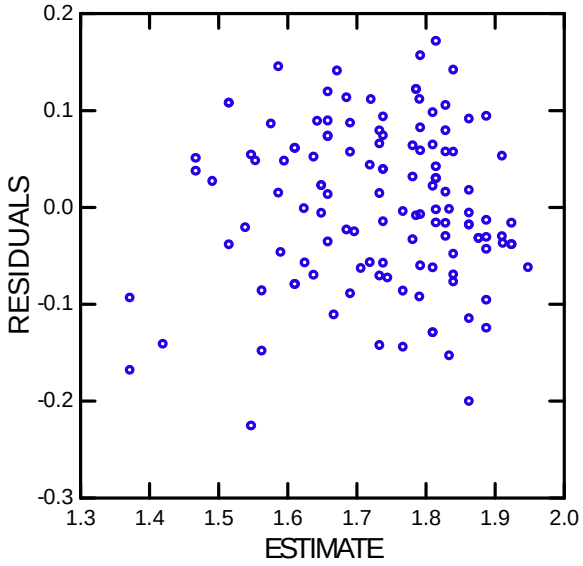
The interaction term is not significant, so it is dropped from the model, which is then refitted and analyzed again.

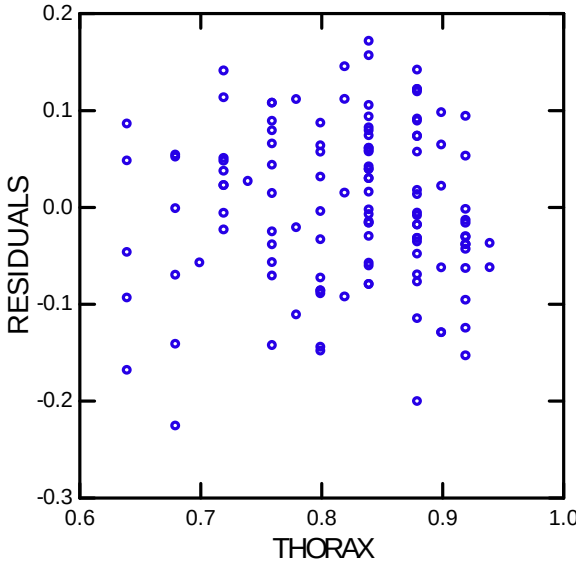
Analysis of Variance					
Source	**Type III SS**	**df**	**Mean Squares**	**F-ratio**	**p-value**
TREATMEN$	0.783	4	0.196	27.970	0.000
THORAX	1.017	1	1.017	145.435	0.000
Error	0.833	119	0.007		

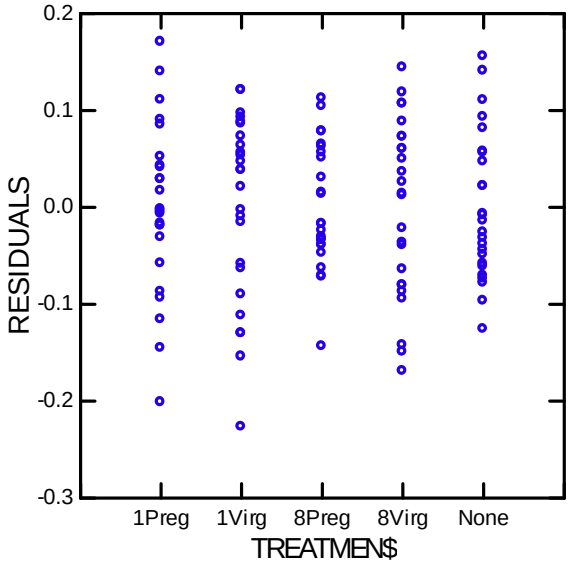
Both Treatment and Thorax Length are significant. Dropping the interaction term lowers AIC

from -256.322 to -257.712. Thus dropping the interaction seems justified (the lower the AIC, the better). Besides, ANCOVA assumes parallel regression slopes, and we cannot test adjusted means unless we drop the interaction. Note, if we drop either Treatment or Thorax Length, AIC increases; thus, both of these terms should remain in the model.

Although not expected for this exam, an examination of the residuals (sensu Zuur et al's 2006 model verification) reveals no patterns to worry about.

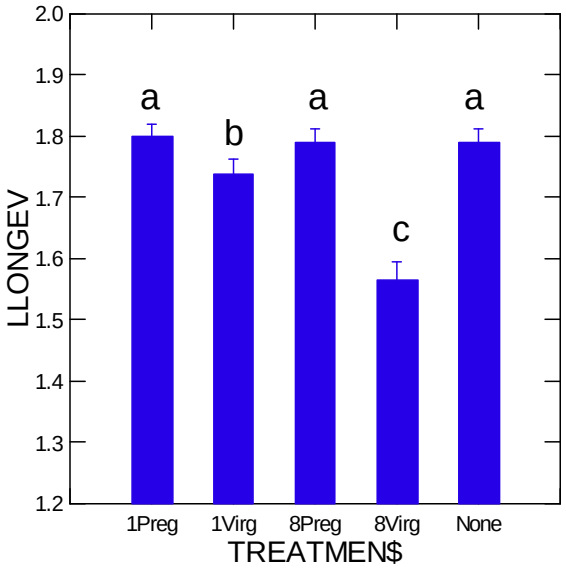






Now we examine the hypothesis that female number and reproductive state both contribute to male longevity. We test all pairwise comparisons among adjusted treatment means. Doing this gives the following Table and Graph…

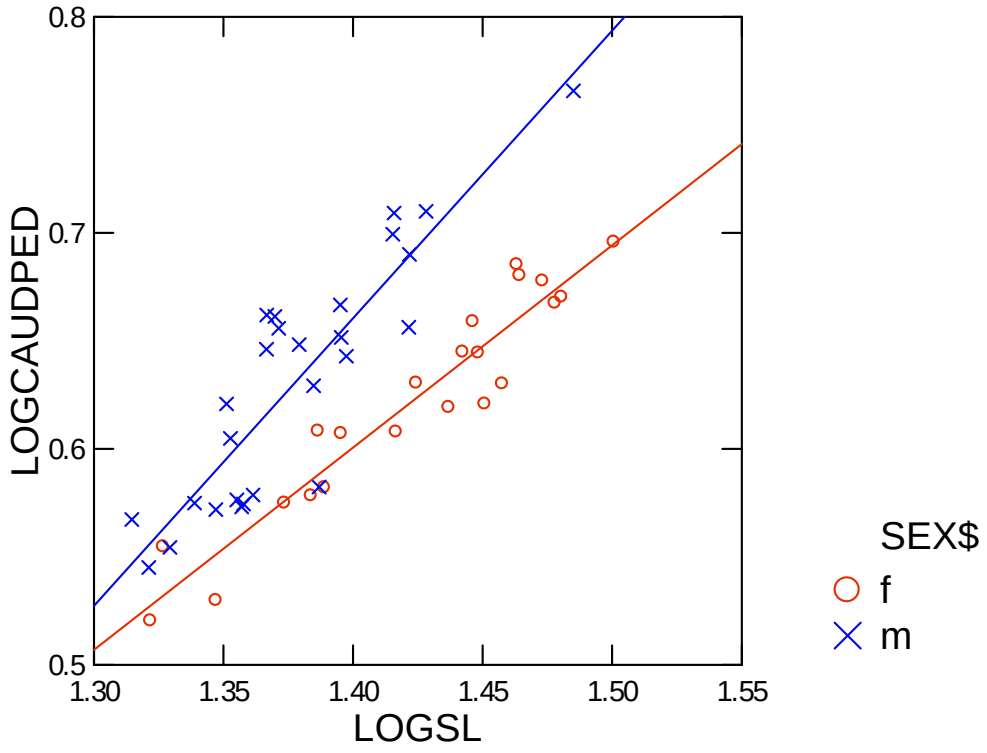
Fisher's Least-Significant-Difference Test					
TREATMEN$(i)	TREATMEN$(j)	Difference	p-value	95.0% Confidence Interval	
				Lower	Upper
1Preg	**1Virg**	**0.076**	**0.002**	0.030	0.123
1Preg	8Preg	-0.014	0.560	-0.061	0.033
1Preg	**8Virg**	**0.204**	**0.000**	0.157	0.251
1Preg	None	0.023	0.342	-0.024	0.069
1Virg	**8Preg**	**-0.090**	**0.000**	-0.137	-0.043
1Virg	**8Virg**	**0.128**	**0.000**	0.081	0.175
1Virg	**None**	**-0.054**	**0.025**	-0.101	-0.007
8Preg	**8Virg**	**0.218**	**0.000**	0.171	0.265
8Preg	None	0.036	0.129	-0.010	0.083
8Virg	**None**	**-0.182**	**0.000**	-0.228	-0.135



The control and two treatments with pregnant females form a non-significant subset; males with 1 virgin female have less longevity than this group, and males with 8 virgin females have less longevity than males with 1 virgin female.

2. A high school intern (Paul Laurence Dunbar Math, Science and Technology Center) undertook a project that examined the ontogeny of sexual dimorphism in the livebearing fish, *Limia perugia*, and quantified this as caudal peduncle depth versus standard length. She was particularly interested in modeling this sexual dimorphism with the allometric equation: $Y = b_0X^{b1}$, where Y is caudal peduncle depth and X is standard length. B1 is called the "allometric coefficient," and its magnitude indicates positive allometry ($b_1 > 1$), isometry ($b_1 = 1$) and negative allometry ($b_1 < 1$). Although these parameters (b_0, b_1) can be estimated by nonlinear regression, Eakin took advantage of the fact that taking logs of both sides of this equation makes it linear: $\log Y = \log b_0 + b_1 \log X$, and that the "allometric coefficient," b_1, can be estimated as the slope of the log-log linear regression line. A snapshot of her data are in the Sexual Dimorphism worksheet in your excel file, where she presents data for a cohort of juveniles about halfway to their final adult body size. Your job is to determine whether or not for this snapshot of her data the two sexes are dimorphic in their allometric coefficients for caudal peduncle depth versus standard length.

2. First, plot the data of Log (Peduncle) versus Log (SL) for the two sexes.



In class, the way we addressed testing for differences among regression slopes was as the preliminary test in ANCOVA for homogeneity of regression slopes, which is included in the ANCOVA model as a Treatment by Covariate interaction. For these data, the model statement is…

Log(Caudal Peduncle Depth) = Sex$ + Log (Standard Length) + Sex$*Log(SL)

…where Sex is fixed and categorical, and Log(SL) is fixed and continuous. The Systat ANOVA table is as follows…

Analysis of Variance

Source	Type III SS	df	Mean Squares	F-ratio	p-value
SEX$	0.003	1	0.003	5.514	0.023
LOGSL	0.114	1	0.114	227.853	0.000
SEX$*LOGSL	**0.003**	**1**	**0.003**	**6.904**	**0.012**
Error	0.022	45	0.000		

The significant interaction term means the male slope is significantly different from (i.e. steeper than) the female slope.

If we examine the individual regression lines and their analyses, we get…

Females

Regression Coefficients B = $(X'X)^{-1}X'Y$

Effect	Coefficient	Standard Error	Std. Coefficient	Tolerance	t	p-value
CONSTANT	-0.711	0.095	0.000	.	-7.517	0.000
LOGSL	0.937	0.066	0.953	1.000	14.106	0.000

Confidence Interval for Regression Coefficients

Effect	Coefficient	95.0% Confidence Interval		VIF
		Lower	Upper	
CONSTANT	-0.711	-0.908	-0.513	.
LOGSL	0.937	0.798	1.075	1.000

Males

Regression Coefficients B = $(X'X)^{-1}X'Y$

Effect	Coefficient	Standard Error	Std. Coefficient	Tolerance	t	p-value
CONSTANT	-1.203	0.190	0.000	.	-6.340	0.000
LOGSL	1.331	0.138	0.888	1.000	9.664	0.000

Confidence Interval for Regression Coefficients

Effect	Coefficient	95.0% Confidence Interval		VIF
		Lower	Upper	
CONSTANT	-1.203	-1.594	-0.812	.
LOGSL	1.331	1.048	1.615	1.000

Note that this is a case where slopes differ significantly, in spite of slight overlap in 95% confidence intervals.

The t-test for each regression slope is a variation on the single sample t-test (which compares a sample mean against a parametric mean e.g. μ = 0)..,

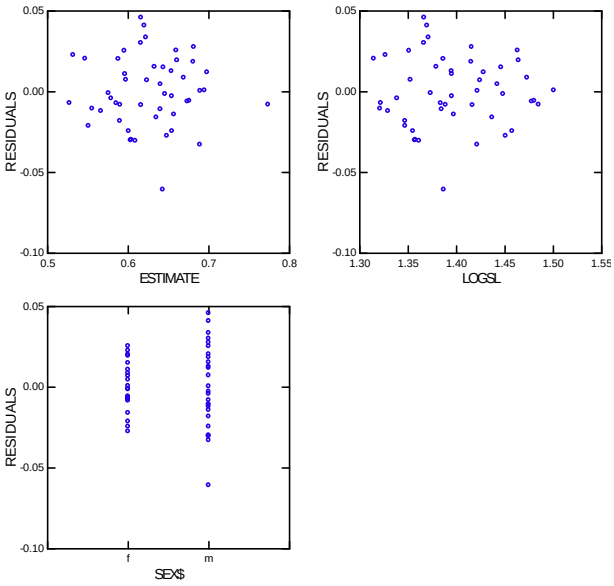
$$t = \frac{b_1 - \beta_1}{SE_{b_1}}$$

…where the null hypothesis is $\beta_1 = 0$, so the t statistics in the tables are regression slopes divided by their standard errors. We can do the same test against the null hypothesis of isometry, $\beta_1 = 1$. Doing these t-tests, we obtain,

Females: t = (0.937-1)/.066 = -0.955, df = 20, p = 0.351 (two-tailed)

Males: t = (1.331-1)/.138 = 2.399, df = 25, p = 0.024 (two-tailed)

…thus males show significant positive allometry; whereas females do not differ significantly from isometry. Note that we can compare male and female regression slopes with a two sample t-test, and if we do so, the t we obtain is the square root of the F for the Sex*Log(SL) interaction in the ANOVA table above.



Residuals appear OK; although, the variances appear heterogeneous between the sexes.

3. Paruelo and Laurenroth (1996) are interested in whether or not the relative abundance of C3 versus C4 plant species (logC3) depends on latitude, longitude and their interaction. Your task it analyze their data with multiple regression (note that General Linear Model procedures easily accommodate interaction terms, and use OLS to estimate your partial regression coefficients). Is there collinearity in your analysis? Can you deal with it, and still answer their question?

3. This problem is one of multiple regression, which we can analyze with a General Linear Model (that uses OLS estimation), such as Systat's GLM. Our model statement is…

Log (C3) = Latitude + Longitude + Lat*Long

…where Lat and Long are both fixed continuous variables. Multiple regression gives the following Table with all 3 slopes significant.

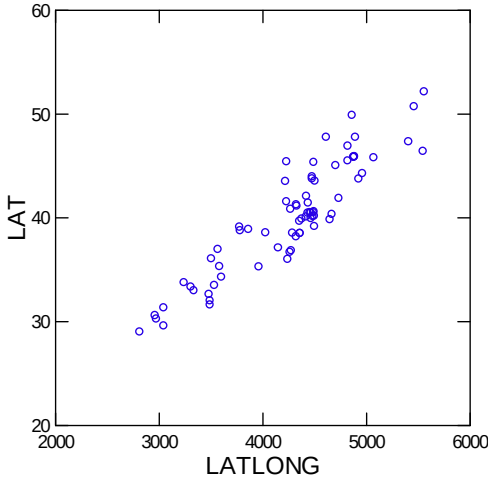
Regression Coefficients B = $(X'X)^{-1}X'Y$

Effect	Coefficient	Standard Error	Std. Coefficient	Tolerance	t	p-value
CONSTANT	7.391	3.625	0.000	.	2.039	0.045
LONG	-0.093	0.035	-1.824	**0.015**	-2.659	0.010
LAT	-0.191	0.091	-3.095	**0.003**	-2.101	0.039
LONG*LAT	0.002	0.001	4.323	**0.002**	2.572	0.012

Pearson Correlation Matrix

	LAT	LONG	LAT*LONG
LAT	1.000		
LONG	0.097	1.000	
LAT*LONG	**0.914**	**0.489**	1.000

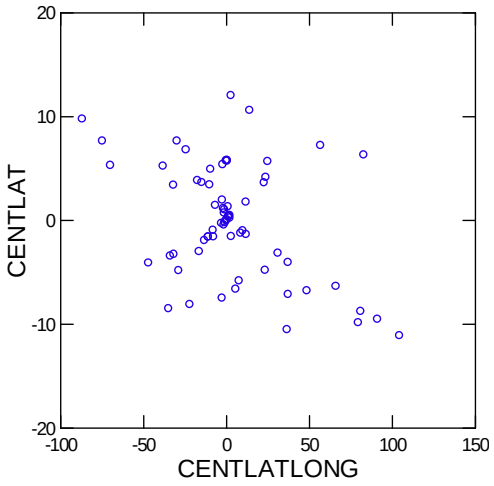
The extremely low tolerance levels indicate collinearity due to high correlations between main effects and their interaction, especially between Lat and Lat*Long.



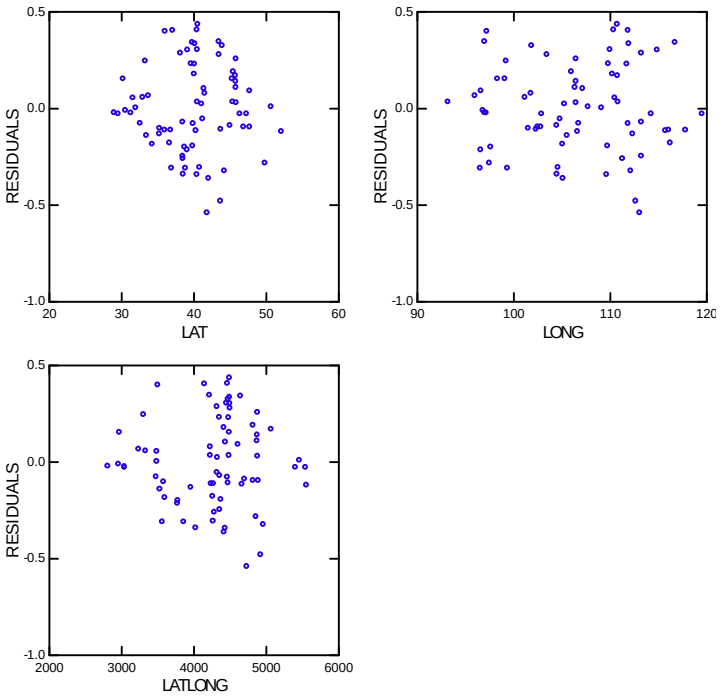
We can address this problem by mean centering both Latitude and Longitude, computing a new interaction based on these mean centered main effects, and refitting the model. Doing this, we get the following:

Regression Coefficients B = $(X'X)^{-1}X'Y$						
Effect	**Coefficient**	**Standard Error**	**Std. Coefficient**	**Tolerance**	**t**	**p-value**
CONSTANT	-0.553	0.027	0.000	.	-20.131	0.000
CENTLONG	-0.003	0.004	-0.051	**0.980**	-0.597	0.552
CENTLAT	0.048	0.006	0.783	**0.827**	8.484	0.000
CENTLONG*CENTLAT	0.002	0.001	0.238	**0.820**	2.572	0.012

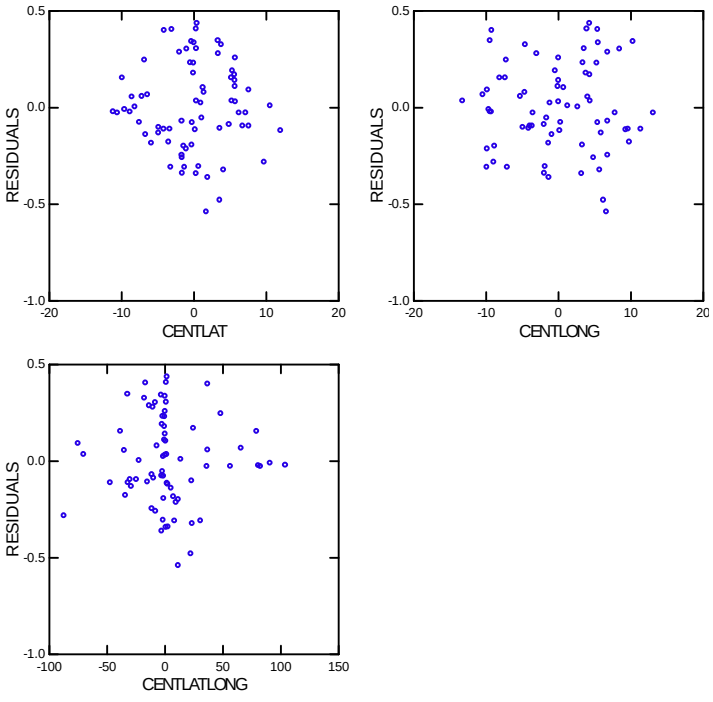
Pearson Correlation Matrix			
	CENTLAT	**CENTLONG**	**CENTLATLONG**
CENTLAT	1.000		
CENTLONG	0.097	1.000	
CENTLATLONG	**-0.414**	**-0.134**	1.000



Now the tolerances are at acceptable levels as are the correlations between the interaction and main effects, and we can interpret our regression parameters. Both Lat and Lat*Long are significant; whereas, Long is no longer significant. Its apparent significance in the first analysis was due to collinearity. We retain the full model, because if an interaction is significant, we retain it along with its main effects.

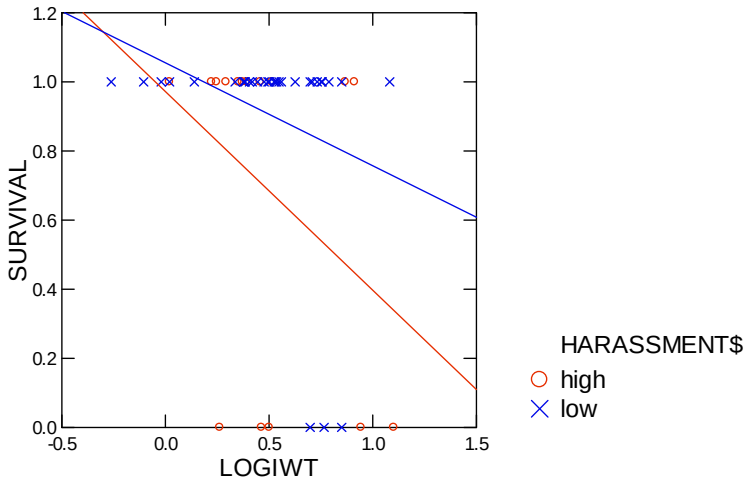


Residuals before (above) and after (below) mean centering versus the 3 axes. Both sets of patterns are similar. There seems to be some variance heterogeneity, with increased variances toward the centers of the Lat and Lat*Long distributions.



4. Jafaar et al (unpublished) are interested in whether or not male sexual harassment (forced male mating attempts per minute per female) affects female survival in the western mosquitofish, *Gambusia affinis*. They divided females into two treatments: high harassment (1.7 forced mating attempts/min) versus low harassment (0.9 forced mating attempts per min), and included body size (log initial weight) as a covariate. Use logistic regression, or a Generalized Linear Model with proper linking function to analyze your data.

4. First we plot the binomial survival data, 0 or 1, versus body size for our two treatments.



Next, we perform a logistic regression (or a generalized linear model with a logit or binary linking function). Our model is…

Survival = Harassment + Log (Weight) + Harassment*Log(Wt)

…where Harassment is a categorical fixed variable and Log(Wt) is a continuous fixed variable.

Parameter Estimates

Parameter	Estimate	Standard Error	Z	p-value	95 % Confidence Interval	
					Lower	Upper
1 CONSTANT	5.993	2.681	2.235	0.025	0.738	11.248
2 HARASSMENT$	-3.758	2.954	-1.272	0.203	-9.547	2.031
3 LOGIWT	-5.814	3.571	-1.628	0.104	-12.814	1.186
4 HARASSMENT$*LOGIWT	3.087	4.079	0.757	0.449	-4.907	11.081

The interaction term is not significant; therefore, we will drop it and refit the model..,

Survival = Harassment + Log (Weight)

…which gives the following results.

Parameter Estimates						
Parameter	Estimate	Standard Error	Z	p-value	95 % Confidence Interval	
					Lower	Upper
1 CONSTANT	4.440	1.328	3.344	0.001	1.837	7.042
2 HARASSMENT$_high	-1.710	0.898	-1.904	0.057	-3.469	0.050
3 LOGIWT	-3.612	1.727	-2.092	0.036	-6.997	-0.228

To check further on whether dropping the interaction was justified, we employ the log-likelihood test on Deviances (Quinn & Keough 2002, 13.15, p. 367):

G = -2(log-likelihood reduced – log-likelihood full)

The log-likelihood of the full model is -16.891. If we drop the interaction from the model, the log-likelihood is -17.204. Thus, G = -2*(-17.204-[-16.891]) = 0.626. This is distributed as χ^2 with 1 df (for one less parameter estimated between the full and reduced models), p = 0.429. Thus, we are justified in dropping the interaction term. Dropping the interaction gives a model in which the covariate, Log(Wt) is significant, and the treatment, Harassment, is almost significant. Should we also drop Harassment from the model? The log likelihood of the model without the interaction, is -17.204, and the log likelihood of a further-reduced model without Harassment..,

Survival = Log (Weight)

…is -19.155. G = -2*(-19.155-[-17.204]) = 3.904, which is distributed as χ^2 with 1 df, p = 0.048. Thus, we keep Harassment in the model. Our final model is..,

Survival = Harassment + Log (Weight)

In this case, we'd cautiously suggest that male Harassment may lower female survival, calculate the power of this test and sample size necessary to detect this effect size in a future experiment. Below is a plot of the predicted survivals versus body size for the two treatments.