Module 4: Analyzing Data That Involve a Significant Group by Body Mass Interaction

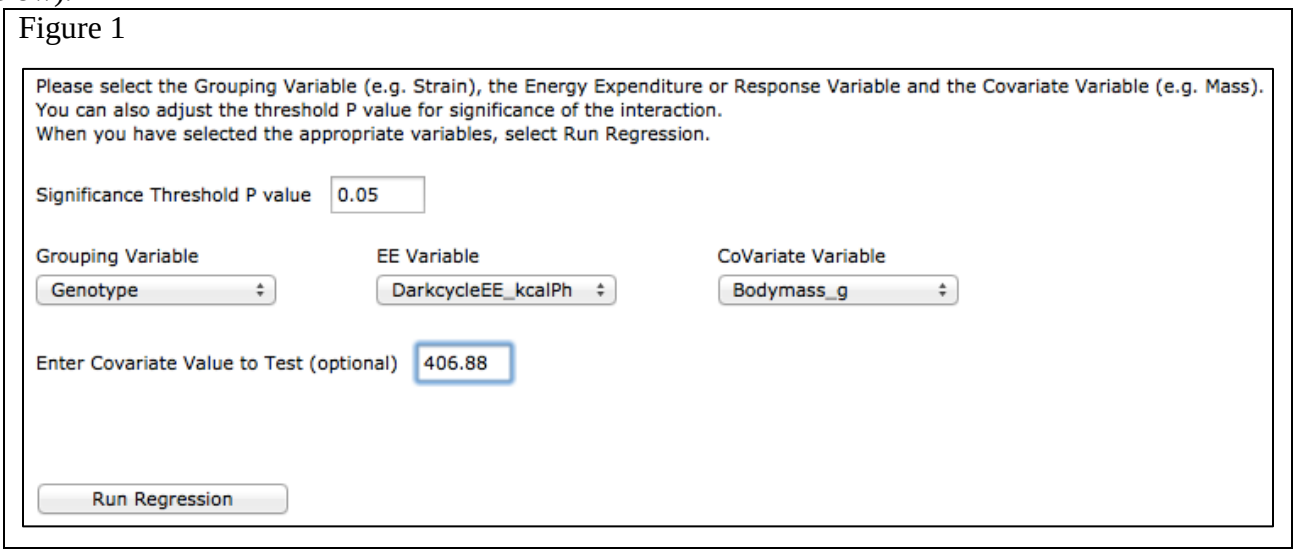
This module explains how to use a multiple linear regression extension of ANCOVA to adjust EE for differences in a body mass covariate when the slopes of energy expenditure on mass differ significantly between groups. A brief explanation of why the slopes may differ is in Module 2.

About the dataset The rat data in **CCK1RKO.csv** were measured in the Rodent Metabolic Phenotyping Center, VA Puget Sound Health Care System, Seattle WA, and are used with permission. The EE data are a component of a study [1] reporting phenotypic characteristics of a cholecystokinin (CCK) knockout rat. These animals lacked the CCK 1 receptor, but retained the CCK 2 receptor and endogenous CCK signaling attributable to CCK R2.

Groups are CCK knockouts (CCK-/-; n=8) and CCK controls (CCK+/+; n=8). The dataset includes dark photoperiod EE in units of kcal/h and mass variables in g.

Analysis of dark cycle EE with total body mass as the covariate Enter the variables to be analyzed as shown in Figure 1, using the MMPC analysis web site (http://www.mmpc.org/shared/regression.aspx). Leave the Significance Threshold P value at 0.05 for the test of whether the slope of EE on Mass differs significantly by group. Enter 406.88 into the Enter Covariate Value to Test box (the reason is explained below).

Figure 1



Click **Run Regression** to obtain the output shown in Figure 2.

Figure 2

Program= MMPC Multiple Linear Regression Output					
Datafile=CCK1RKOdata.csv					
Response Variable=DarkcycleEEkcalPh					
Covariate=Bodymass_g					
Grouping Variable=Genotype					
The interaction Bodymass_g:Genotype is significant.(p=0.001712)					
16 total cases. CCK-/- = 8, CCK+/+ = 8					
R Squared = 78.7497	Adjusted R Square = 73.4371				
Residual Standard Error 0.0551 with 12 degrees of freedom					
Source	Sum of Squares	df	MeanSquare	F-ratio	P value
Regression	0.134798	3	0.0449327	14.8233	0.000244191
Residual	0.0363747	12	0.00303123		

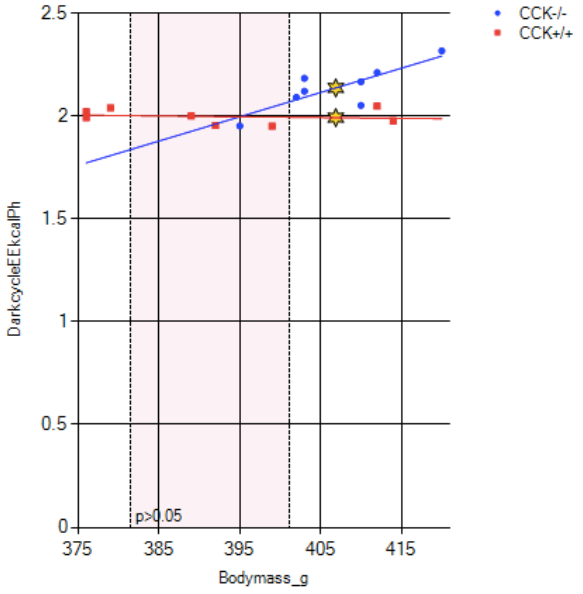
Variables	Estimate	StdError	P Value
(Intercept)	-2.66714	1.10263	0.0323828
Bodymass_g	0.0118041	0.00270957	0.000934186
Genotype	4.81405	1.22612	0.00201225
Bodymass_g:Genotype	-0.0121867	0.00303473	0.00171249

Basic Statistics - Avg (StDev)

Groups	CCK-/-	CCK+/+
DarkcycleEEkcalPh	2.14 (0.103)	2 (0.034)
Bodymass_g	406.88 (7.184)	392.13 (14.243)

Model-based Statistics - Avg DarkcycleEEkcalPh (StErr)

	Groups		
Used: Bodymass_g	CCK-/-	CCK+/+	P value
Overall Mean	2.05 (0.028)	1.99 (0.022)	0.1502
Group Mean	2.14 (0.019)	2 (0.019)	0.0002877
Bodymass_g=406.88	2.14 (0.019)	1.99 (0.028)	0.00116
Johnson-Neyman Cutoffs	<=381.6 , >=401.1		
Residual Variance	0.00454	0.00153	



Interpretation Note: All p-values are two-tailed.

Topmost information box:

- Includes basic information on variable selection and sample sizes.
- The body mass by genotype interaction term is denoted as Bodymass_g:Genotype.
- The statistical significance of the Bodymass_g:Genotype interaction term is based on the p-value, p=0.0017, which indicates the slope of EE on body mass does differ between groups at the designated significance threshold value for the interaction term. Accordingly, the ensuing regression model includes the interaction term.
- The R Squared value indicates that the ANCOVA model accounts for 78.7% of the total variance in dark cycle EE (p=0.00024; shown in the line labeled Regression in the table). An explanation of Adjusted R Square of 73.4 % is beyond the scope of this module.

Variables information box and graph of 24h EE vs. total mass by group:

- The inclusion of the interaction term makes the mathematical expression for the regression model more complex than is the case for standard ANCOVA. However, the regression lines in the graph show the regression model in visual form. To interpret the model, just realize that the regression lines show the estimated mean dark cycle EE values for each group given a particular value of body mass on the x-axis.
- The estimates in the "Variables…" information box provide the numerical information used to determine the equations of the lines that estimate DarkcycleEE_kcalPh for each group. Because two groups are included, two separate regression lines with different slopes and intercept values are estimated. One must figure out how the groups correspond to the lines, but this is not difficult.
- One group is the reference group while the other group is the alternative group. The equation of the line for the reference group equals the sum of the estimates for "(Intercept)" and "Bodymass_g" times body mass. The equation for the alternative group equals the sum of the estimates for "(Intercept)", "Bodymass_g" times body mass, "Genotype", and the "Bodymass_g:Genotype" interaction term times body mass. Note that the y-intercept term for the alternative group equals the sum of the "(Intercept)" and "Genotype" terms, while the slope parameter equals the sum of the "Bodymass_g" and "Bodymass_g:Genotype" estimates.

  Accordingly, the two estimated lines are:

  Reference group line:

  $$\text{DarkcycleEE_kcalph} = -2.667 + 0.0118 \times \text{Bodymass_g}$$

  Alternative group line:

  $$\text{DarkcycleEE_kcalPh} = -2.667 + 4.814 + 0.0118 \times \text{Bodymass_g} - 0.0122 \times \text{Bodymass_g}$$

  $$= 12.667 + 4.814 + (0.0118 - 0.0122) \times \text{Bodymass_g}$$

  $$= 2.147 - 0.0004 \times \text{Bodymass_g}$$

  The task now is to determine which equation corresponds to which genotype. Here the equation for the reference group has a larger slope than the equation for the alternative group (0.012 vs. -0.0004), which follows from the negative estimate for the Bodymass_g:Genotype interacton (-0.0122). The graph of the lines in Figure 2 shows that the CCK -/- genotype has the larger slope, which indicates that this genotype must be the reference group. The second equation therefore corresponds to the one for the CCK +/+ genotype.
- Important notes about the parameter estimates:
  - The parameter estimate for Genotype corresponds to the estimated "group difference" in DarkcycleEE_kcalPh, but only at the point where Bodymass_g is equal to zero (the y-intercept). Since the two lines are not parallel, this difference would only have meaning when the bodymass is zero, which is not biologically relevant. Thus, the estimate for the group difference and its associated p-value are invalid at non-zero values of the mass covariate in interaction models (but not in standard ANCOVA). (Remember in standard ANCOVA the lines are parallel so the group difference when body mass is zero (the difference in y-intercepts) pertains at all body masses). Many users of regression have failed to understand this fact [2].
  - The appropriate line should be used to calculate the predicted DarkcyleEE_kcalPh. For example, to compute estimated EE for the CCK+/+ rats at their mean body mass of 406.88 g:

$EE\ for\ CCK\ control\ rats\ at\ 406.88\ g = 2.147 - 0.0004\ \times\ 406.88 = 1.98$ kcal/h

(Owing to rounding error, this result differs by 0.01 from the result based on entering 406.88 in the topmost information box, and shown in the Model Based statistic box and described below).

- Because the parameter estimate for Genotype reflects the difference between the genotype groups only when Bodymass_g is zero, one needs a way to estimate the difference between groups when Bodymass_g equals a particular value within the observed range of this covariate. The WEB interface makes it easy to obtain a valid estimate for the group difference at any value of Bodymass_g by specifying the value of the covariate to test in the WEB interface. We did this at the beginning of this module by entering 406.88 into the "Enter Covariate Value to Test box." The result is discussed below.

Basic Statistics box

- Shows basic (unadjusted) means and standard deviations for dark cycle EE and body mass by group.

Model Based Statistics box

- A key feature of analyses that include the group by body mass interaction is that the difference between the groups depends on the body mass. Thus, the groups may show no difference for certain values of body mass (i.e. in the region where the lines intersect), while the groups will show some difference for other values of body mass. A major task of the analysis is therefore to determine the values of body mass for which EE differs. The "Model Based Statistics" box, along with the ability to specify values of the covariate to use to predict EE, can be used for determining how the groups differ.
  - Note: the distribution of body mass is not the same for both groups in the example here, with the CCK -/- genotype not having smaller body masses. As such, the difference in body mass potentially contributes to the difference in mean EE between the groups.
- The row labeled "Overall Mean" shows the regression-adjusted group means, standard errors and significance of the group comparison when the comparison is adjusted to the overall mean of the body mass covariate (the 'overall mean' is the mean of body mass for the entire sample taken over both groups). Stated differently, the Overall Mean information gives the estimated EEs and standard errors that the groups would have if every animal had the mean body mass of the entire sample. In the present example, the EE difference adjusted to the mean body mass of the entire sample is not significant (p=0.15). The overall mean body mass is at a point close to where the two lines intersect , which is why the two groups do not differ at this "average" body mass.
- The row labeled "Group Mean" shows the adjusted means, standard errors and p-value for the group comparison when the comparison is made using the individual group means of the body mass covariate. Thus for the CCK-/- group EE for every animal is adjusted to 406.88 g, while for the CCK+/+ group EE for every animal is adjusted to 392.13 g (see Basic Statistics box), and the group comparison is performed at these values. This method retains the effect of the group body mass difference on the group EE difference and yields the same group difference as obtained in the numerator of a Student's independent samples t-test. The advantage of the regression approach is that the variance of the group EE estimates and the variance of the EE difference between groups is reduced by accounting for variance in EE due to variance in body mass. This yields a more precise estimate of the group difference in an analysis that is intended to embody both the effect on EE of a group difference in body mass and a mass-independent EE component. In the example here, the group means do differ at these different body masses (p =

0.0003). This result indicates that the different body masses do contribute to a difference in mean EE.

- The row labeled Bodymass_g=406.88 is produced based on having specified 406.88 as the value of the covariate to test when the analysis was originally run. This row gives the regression-adjusted group means, standard errors and significance of the group comparison when the comparison is adjusted to the CCK-/- group's mean value of the body mass covariate. As the slope of EE on a body mass covariate differs between groups (i.e. there is an interaction), asking whether groups differ in EE when compared at the mean mass value of the experimental group (or control group) can be a logical question. In the present example, the CCK-/- rats exhibit higher EE compared to CCK+/+ rats when EE is adjusted to the mean value of the experimental group (p=0.001). The star symbols on the graph show the regression-estimated EE values for each group.
- The Johnson-Neyman procedure [3] provides another approach to determining the range of body mass for which mean EE differs. The row labeled Johnson-Neyman Cutoffs shows the results for the Johnson-Neyman procedure, providing the range values of body mass for which mean EE differs significantly between groups. The shaded region on the graph shows the region of no significant difference in mean EE based on the chosen significance threshold p-value for the interaction term.

"Teachable moments" raised by this dataset We chose this dataset primarily to illustrate how to use the web tool and interpret its output when there is a group by body mass interaction. The goal was not to convince you that CCK -/- mice have lower dark cycle EE below a body mass of 381.6 g and have higher EE above 401.1 g. The data, however, allow us to raise some issues that you should be aware of when interpreting EE phenotyping studies.

- **Be careful about generalizing beyond the data** Because there are no CCK-/- mice below the lower Johnson-Neyman cutoff, you should not conclude that the groups actually differ when the animals mass is less than the cutoff. In regression analyses, it is best to not extrapolate to regions that do not include actual data. Accordingly, the analysis is more believable when considered in the region that includes rats from both groups above the upper Johnson-Neyman cutoff.
- **Evaluate the data holistically** Analyses involving interactions should be interpreted very thoughtfully using all the information one has at hand. The CCK-/- rats were significantly and substantially more active than controls during the dark phase, which provides a biologically plausible explanation for their higher dark cycle EE when considered relative to the control rats in the shared range of body mass. When the analysis is run using LBM instead of body mass, the interaction term is not significant (p=0.47) yet the group comparison remains in the direction of CCK-/- rats having higher dark cycle EE than controls (by 0.01± 0.067 kcal/h; p = 0.098), when you make the comparison at the mean weight for each group. When the analysis is run using FM, the interaction term is not significant (p=0.06) and the group comparison is significant in favor of the CCK-/- rats (0.14±0.041 kcal/h ; p=0.004). You are encouraged to run these analyses yourself. Taken together, the data suggest that CCK-/- rats have an elevated dark cycle EE phenotype that could be secondary to a higher activity phenotype. Consistent with this interpretation, EE and activity did not differ between groups during the light cycle photoperiod.
- **Interactions can reflect sampling error** It may be that the interaction occurred solely because of sampling error and that no interaction actually exists in the populations from which the samples were drawn. This possibility is a particular concern in studies involving modest sample sizes (the current sample size is small) as just a few observations can have a big effect on a regression slope. A "false interaction" should be regarded as a possibility in this dataset due to the complete lack of the expected (i.e. EE should be higher in larger animals) positive

relationship between EE and body mass in the control rats despite a substantial range of body mass (almost 100 g).

The explanation of the dfbeta plots is considered in detail in module 3.

Acknowledgements: This module was written by Karl J. Kaiyala[1], Charles F. Spiekerman[2], Robert H. Podolsky[3] and Owen P. McGuinness[4]. The MMPC ANCOVA web tool was created by Richard McIndoe[5].

[1]School of Dentistry, Department of Oral Health Sciences, University of Washington, Seattle, WA; [2]School of Public Health and Department of Biostatistics, University of Washington, Seattle, WA; [3]Department of Family Medicine and Public Health Sciences, Wayne State University School of Medicine, Detroit MI; [4]Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, TN; [5]Coordinating and Bioinformatics Unit, Georgia Regents University, Augusta, GA.

The authors were supported by funding from the National Institute of Health (KJK: U24DK076169, P30DK035816; CFS: U24DK076169; RHP: R03-DA033453; OPM: DK59637, DK20593, DK078188, DK043748; RM: U24DK076169) and the NIDDK Mouse Metabolic Phenotyping Center Program.

References

1. Blevins JE, Moralejo DH, Wolden-Hanson TH, Thatcher BS, Ho JM, Kaiyala KJ, Matsumoto K (2012) Alterations in activity and energy expenditure contribute to lean phenotype in Fischer 344 rats lacking the cholecystokinin-1 receptor gene. Am J Physiol Regul Integr Comp Physiol 303: R1231-1240.
2. Engqvist L (2005) The mistreatment of covariate interaction terms in linear model analyses of behavioural and evolutionary ecology studies. Animal Behaviour 70: 967-971.
3. White CR (2003) Allometric analysis beyond heterogeneous regression slopes: use of the Johnson-Neyman technique in comparative biology. Physiol Biochem Zool 76: 135-140.