Activity 13: Why do birth weights vary?

LEARNING GOALS

- Explain the benefits of running one model with all explanatory variables of interest, over separate one-variable models
- Explain the benefits of running a model with all explanatory variables of interest and interactions, over one-variable model with all the treatment combinations
- Describe what a statistical interaction is.
- Explain how the effects of other variables are "accounted for" via the design of the study in well-designed experiments

Research question: *How do maternal body temperature and food availability during gestation affect the birthweight of baby skinks?*

Background and data: Researchers (Itonga et al. *PlosOne*, 2012) in the School of Zoology at the University of Tasmania in Australia investigated the effects of **maternal body temperature** and **food availability** during gestation on the offspring of the southern grass skink. The southern grass skink is a type of lizard endemic to colder regions of Australia. It has a lifespan of about 5 to 6 years and grows to be about 7.5 cm long (not including the tail).

The researchers obtained 160 pregnant skinks of approximately the same age. They manipulated both the length of time the expectant mother was given the opportunity to bask in the sun (4 hours vs. 12 hours per day) and the amount of food that was available per day (high vs. low) during the gestation period. Forty skinks were randomly assigned to each treatment, that is, each of the 4 combinations of time spent basking in the sun and amount of food. Each skink was kept in its own enclosure and apart from the treatment the skink was exposed to, the environmental conditions were the same for all enclosures. The researchers measured several response variables on the offspring, one of which was the body mass (in mg).

1. What do you think are some variables (besides maternal basking and food quantity) that may predict a baby skink's birth weight?

Observed variation in: birthweight (mg.)		
Inclusion criteria & design	Sources of explained variation	Sources of unexplained variation
	Maternal basking Food availability	

Materials and Methods

Maternal conditions during gestation

One hundred and sixty female *P. entrecasteauxii* were collected from one population at the Peter Murrell Reserve in Kingston, southern Tasmania, Australia (41°50′S, 146°36′E) between 15 and 31 October 2007, shortly after ovulation took place. We do

Females were randomly assigned to one of the following four treatment groups (2 by 2 factorial design):

1. 12 hours basking opportunity per day with high food availability (three mealworms per lizard, three times per week);
2. 12 hours basking opportunity per day with low food availability (two mealworms per lizard, once a week);
3. four hours basking opportunity per day with high food availability; and
4. four hours basking opportunity per day with low food availability.

2. Based on the above excerpts from the research paper, what were the inclusion criteria? What is the type of study? Update the Sources of Variation diagram based on these.

Let's explore the data: [Multiple Variables applet]

- Click on the **Clear** button to clear the pre-filled data
- Copy all the columns from the [skinks] data Google sheet
- Paste data into the **Enter data** box
- Click on the **Use Data** button
- Drag and drop variables

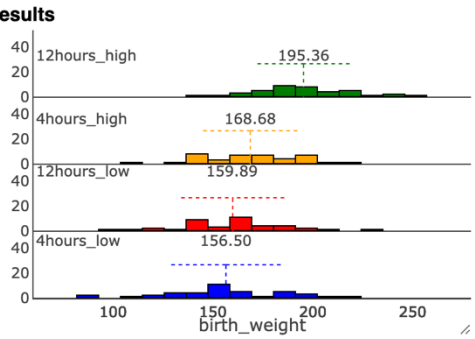
3. Use the applet to obtain and record a prediction equation to predict birth weight from food availability alone. What proportion of variation in birth weights is this model explaining?

4. Next, use the applet to obtain and record a prediction equation to predict birth weight from basking time alone. What proportion of variation in birth weights is this model explaining?

What if we wanted to include both food availability and basking time in the same model?

Approach 1: One-variable (with four treatments or levels) analysis

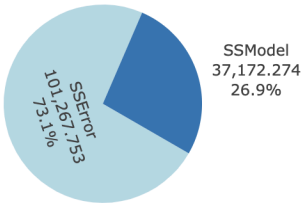
5. You can use the applet again, this time with the treatment variable as the explanatory variable.



$$predicted\ birth\ weight\ (mg.) = \begin{cases} 156.50 & if\ low\ x\ 4\ hours \\ 168.68 & if\ high\ x\ 4\ hours \\ 159.89 & if\ low\ x\ 12\ hours \\ 195.36 & if\ high\ x\ 12\ hours \end{cases}$$

with *SE = 26.024mg.*

SSTotal = 138440.027



a) What proportion of variation in birth weights is this model explaining?

b) Is this proportion of variation bigger or smaller than the two proportions of variation reported in (3) and (4)? What about compared to their sum?

c) This model (from Approach 1) is explaining more variation than just basking time or just food availability individually! So, it seems to be a better model than the two separate one-variable models in (3) and (4). However, this model does have a disadvantage. What is the disadvantage of Approach 1?

Approach 2: Two-variable analysis

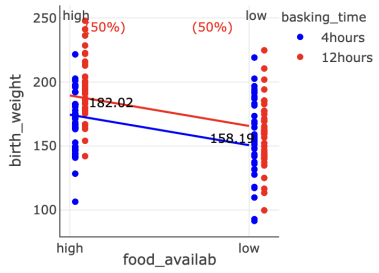
6. You can use the applet again, this time with the treatment variable as the explanatory variable.

Build model

Variables: treatment

Response: birth_weight

Explanatory: basking_time, food_availab



☑ Show equation

☑ Show equation

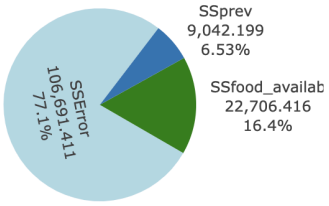
4hours: *predicted birth_weight = 162.59 + −11.91 × food_availab*

12hours: *predicted birth_weight = 177.63 + −11.91 × food_availab*

☐ Adjust values

☑ Separate Lines

SSTotal = 138440.027



a) Does Approach 2 overcome the disadvantage of Approach 1?

b) Are you satisfied with Approach 2? Why or why not?

Hint: What proportion of variation in birth weights is this model in Approach 2 explaining? Is this proportion of variation bigger or smaller than that explained by the model in Approach 1?

7. Let us compare the Analysis of Variance (ANOVA) tables to see how the Sums of Squares are being partitioned.

Approach 1:

ANOVA Table: ☑

Source	df	SS	MS	F-stat	p-value
treatment	3	37172.27	12390.76	19.09	< 0.0001
Error	156	101267.75	649.15		
Total	159	138440.03			

Approach 2:

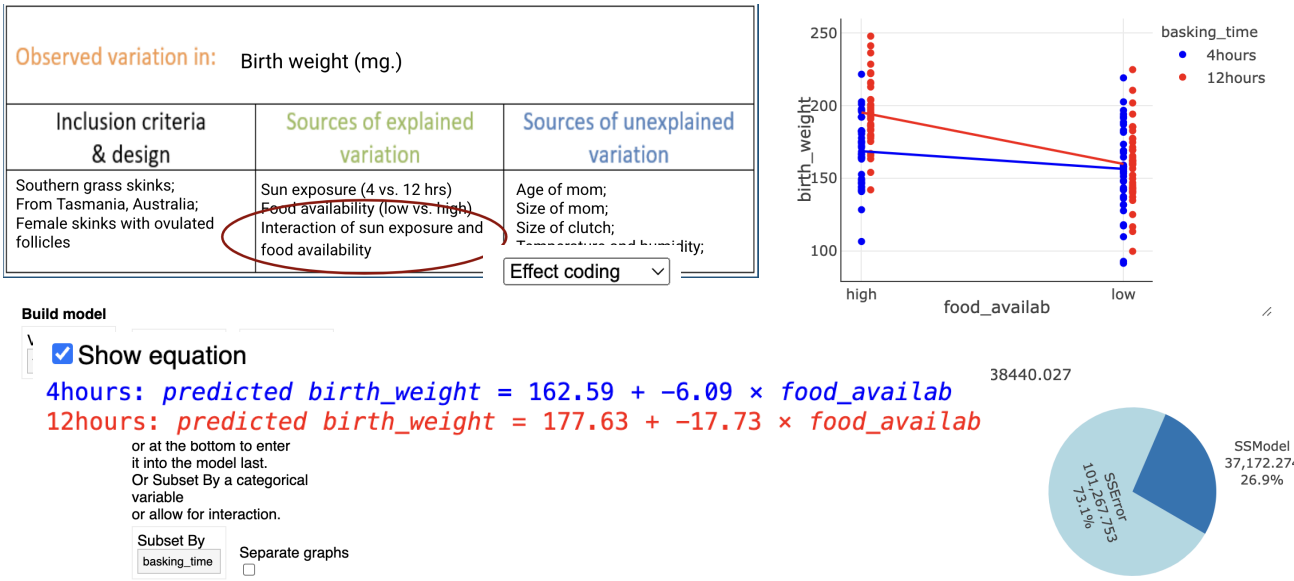
ANOVA Table: ☑

Source	df	SS	MS	F-stat	p-value
model	2	31748.62	15874.31	23.36	< 0.0001
basking_time	1	9042.20	9042.20	13.31	0.0004
food_availab	1	22706.42	22706.42	33.41	< 0.0001
Error	157	106691.41	679.56		
Total	159	138440.03			

As you may have noticed already The SSModel for Approach 2 is *lower* than Approach 1! By how much? What accounts for that?

Approach 3: An interaction model

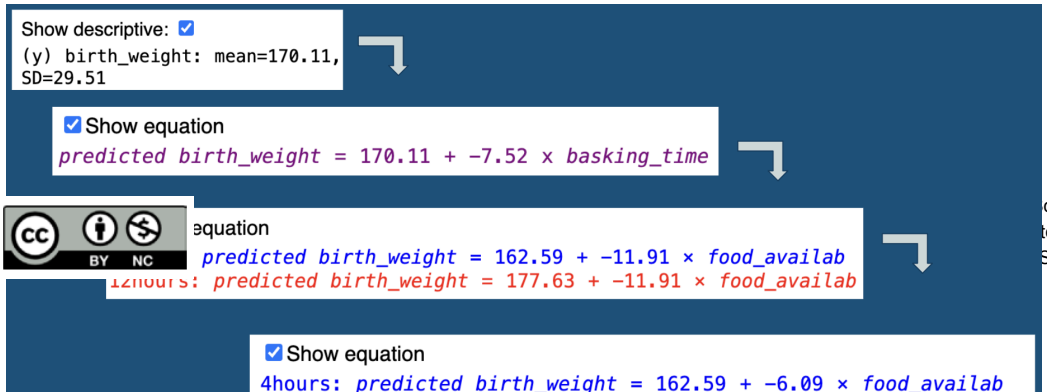
8. Let's add an interaction of basking time and food availability to the model!



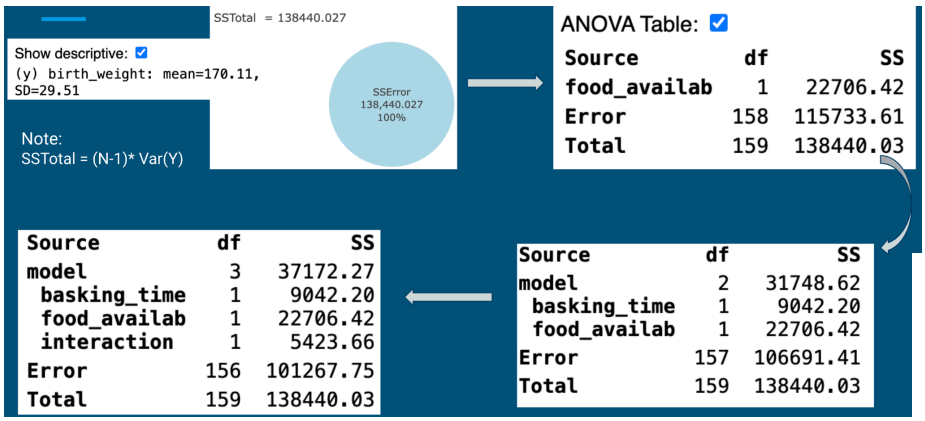
a) Does Approach 3 overcome the disadvantages of Approaches 1 and 2?

b) Are you satisfied with Approach 3? Why or why not?

9. What do you notice about how the effects get *updated* as we include ("adjust/account for") other explanatory variables in the model?



10. What do you notice about the *Total Sums of Squares* get partitioned as we include ("adjust for") other explanatory variables in the model? Why does it make sense given how the treatments were administered?



Key idea: Typically, in well-designed experiments, the effects of the various explanatory variables are *orthogonal* to each other, meaning the effects and sums of squares pertaining to individual explanatory variables remain the same regardless of the presence of other explanatory variables. So, we don't need to say "*after adjusting for…"* when we interpret the coefficients.