Exploration 3.2: What determines a baby skink's weight?

Researchers 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).

K. Itonaga, S. Jones and E. Wapstra, 2012, "Effects of Maternal Basking and Food Quantity during Gestation Provide Evidence for the Selective Advantage of Matrotrophy in Viviparous Lizard," *PlosOne*, https://doi.org/10.1371/journal.pone.0041835

In this exploration, we will focus on how time basking in the sun and the amount of food available affect the body mass of the offspring.

1. Identify the experimental units (how many are there?) and the explanatory variables (aka "factors") in this study. Identify and count the levels of each explanatory variable/ (This is a ____ x ____ factorial design.)

2. Fill in the possible Sources of Variation diagram based on what you know about the study already.

Observed variation in:	Sources of explained variation	Sources of unexplained variation
	• •	• •
Inclusion criteria • *Design* •		

3. Identify the four treatments in this study. Is this a balanced design?

The data are in the ***skinks*** file. Copy and paste the data into the **Two-variable ANOVA** applet, press **Use Data.** Check the **Show Means** box to determine the treatment means.

4. Enter the four treatment means into the table below. Then compute the row means and the column means and the overall mean. (Keep in mind that with a balanced design, the "row" and "column" means will be the same as the means of all the corresponding values in each row and each column.)

		Food Availability		
		Low	**High**	**Mean**
Maternal Basking	**4 hours**			
	12 hours			
	Mean			

5. From your table, calculate and interpret the maternal basking effects and the food availability effects and write out the prediction equation.

maternal basking effects: food availability effects:

interpretation:

Predicted offspring body mass =

6. Use your prediction equation to predict the outcomes for each treatment.

Observed

	Low	High
4 hours	156.50	168.68
12 hours	159.89	195.36

Predicted

	Low	High
4 hours		
12 hours		

7. The graph below shows the observed treatment means. Add the predicted treatment means to the graph. Do the predicted means closely match the observed means?

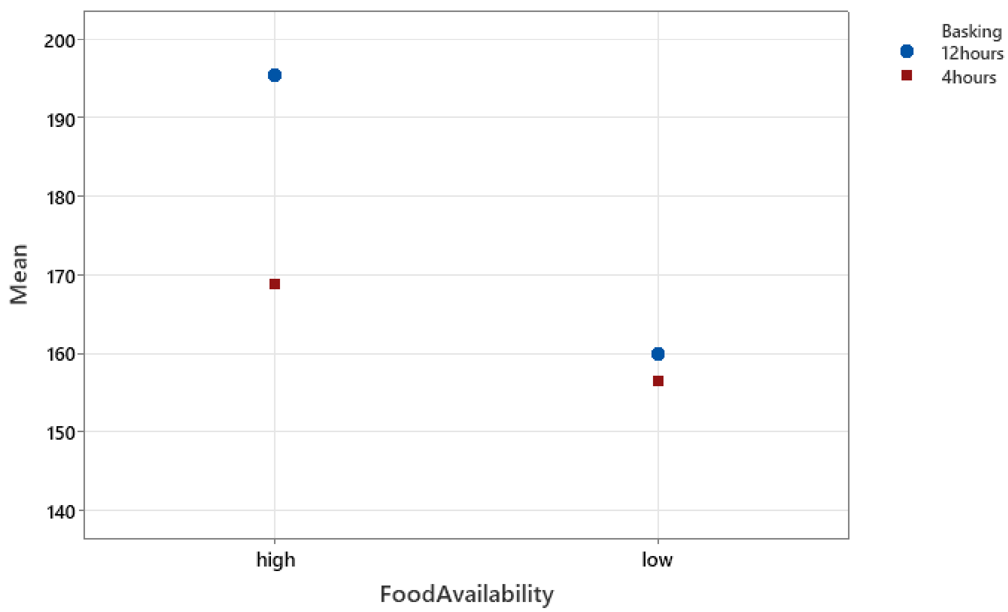


FIGURE 1: Observed treatment means of offspring body mass in the four treatment groups.

8. Now, calculate how different each observed mean is from its prediction. To do this compute the differences (observed – predicted) and put them in the table below.

Differences = Observed – Predicted (use two decimal places)

	Low	High
4 hours		
12 hours		

9. Ideally, if our model is doing a good job of predicting, what value would we like the values in the table in #9 to be (or at least be close to)? What interesting feature do you notice about these values?

Interestingly, the model predictions above are not particularly close to the observed values! Let's dig deeper into what's happening in this study by looking at two different ANOVA tables.

Approach #1: One-Variable Analysis

One analysis approach for a 2 x 2 factorial design is a *separate means model* which estimates each of the four treatment means separately. The ANOVA table just has one row representing the variation explained by for the four treatments.

Source	DF	SS	MS	*F*-statistic	p-value
Treatment	3	37170.89	12390.3	19.0867	< 0 .0001
Error	156	101268.81	649.2		
Total	159	138439.70			

The model is shown below. Notice that it uses the observed treatment means and thus exactly predicts each treatment mean.

$$\text{Predicted body mass} = \begin{cases} 156.50 & if\ low\ x\ 4\ hours \\ 168.68 & if\ high\ x\ 4\ hours \\ 159.89 & if\ low\ x\ 12\ hours \\ 195.36 & if\ high\ x\ 12\ hours \end{cases}$$

10. What is the null hypothesis for this model? What proportion of total variation in offspring body mass is explained by this model?

Approach #2: Two-Variable Analysis

A second approach is a two-variable model which estimates the treatment effects for each variable and separates the two sources of variation in the ANOVA table.

Source	DF	SS	MS	*F*-statistic	p-value
Model	2	31747.14	15873.6	23.36	< 0.0001
Maternal Basking	1	9042.34	9042.34	13.31	0.0004
Food Availability	1	22704.79	22704.79	33.41	< 0.0001
Error	85	106692.57	679.6		
Total	87	138439.70			

11. What is an advantage of Approach #2 over Approach #1 in terms of your conclusions about the study?

12. What does the p-value corresponding to the overall two-variable model tell us? Describe in the context of the study.

13. What proportion of the total variation in offspring body mass is explained by each variable? What are the two null hypotheses for these p-values for the individual variables?

14. What proportion of the total variation in offspring body mass is explained by this model (the model R^2)? How does this compare to what was found with Approach #1? Is the Approach #2 overall model statistically significant?

Key Idea: The two-variable analysis allows us to make distinct conclusions about the individual variables (the *main effects*). However, in this case, the two-variable model is not able to explain as much variation in the response as the separate means model using the treatments. This happens because the two-variable model estimates the main effects by combining the information across the other variable (making some assumptions), which can sometimes miss some key patterns in the responses.

Statistical Interaction

A key advantage of a factorial design over a one-variable-at-a-time design is the ability to estimate the effects of both of the individual variables (e.g., maternal basking and food availability) within the same experiment. A second advantage is the factorial design allows us to consider whether the effect of one variable could change depending on the category/value of the other variable. When this is the case, we say that the explanatory variables *interact*.

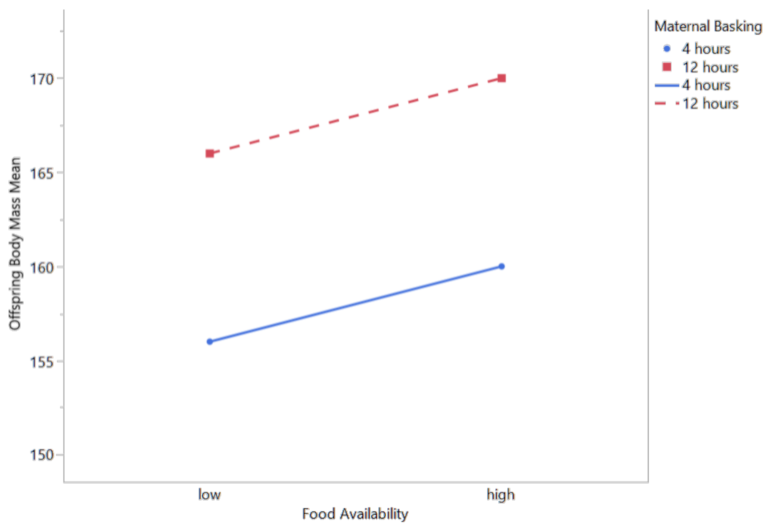
Definition: A ***statistical interaction*** occurs between two explanatory variables when the effect (or association) of one explanatory variable on the response variable changes based on the value of the other explanatory variable. In other words, one of the explanatory variables *modifies the effect* of the other explanatory variable on the response variable (and vice versa). An ***interaction plot*** displays the treatment means on the vertical axis, with one explanatory variable or factor on the horizontal axis and the other explanatory variable or factor identifying the treatment means with different colors or styles (it doesn't matter which factor is which).

15. To begin to investigate the interaction between maternal basking and food availability on offspring body mass, start by updating your possible Sources of Variation diagram from #2 by adding in a source of explained variation called "Interaction between maternal basking and food availability" on a separate line.

Observed Variation in: Offspring body mass (mg)	**Sources of explained variation**	**Sources of unexplained variation**
Inclusion criteria *Design*	• Maternal basking • Food availability •	•

To graphically explore the presence (or not) of an interaction, we plot one of the explanatory variables on the horizontal axis (doesn't matter which one), and then we plot the treatments means, using separate lines to connect within the categories of the other explanatory variable. For example, consider Figure 2.

FIGURE 2 Hypothetical interaction plot.



We can first consider the individual main effects.

16. Estimate the mean offspring body mass with 12 hours maternal basking (averaging the two red values in the graph). Estimate the mean offspring body mass with 4 hours maternal basking (averaging the two blue values in the graph). Do these means indicate a main effect of maternal masking?

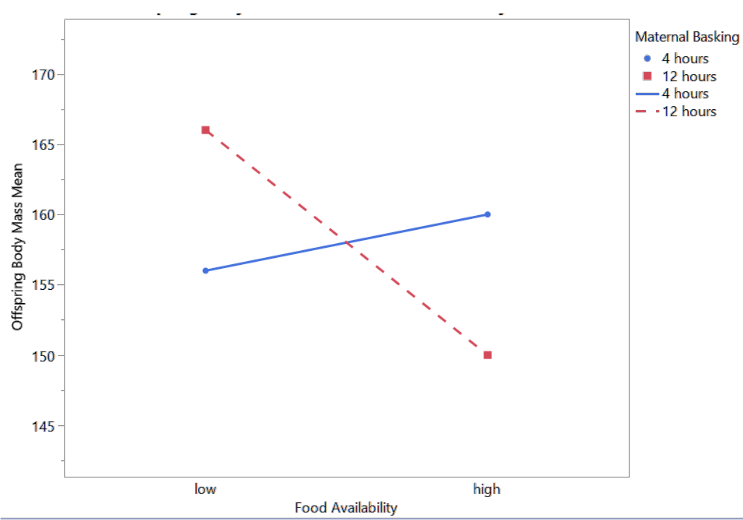
17. Estimate the mean offspring body mass with low food availability and with high food availability. Do these values indicate a main effect for food availability?

Next we can consider the interaction.

18. Estimate the difference in mean offspring body mass between 4 hours and 12 hours of basking time when food availability is low? Is this similar to the difference in mean offspring body mass between 4 hours and 12 hours of basking time when food availability is high? If so, then we don't have evidence of an interaction – the effect of maternal basking is the same for both food availability categories.

Now consider another hypothetical interaction plot in Figure .3

FIGURE 3: Hypothetical interaction plot.



19. Is there a main effect of maternal basking on offspring body mass? A main effect of food availability on offspring body mass? Explain how you are deciding in each case.

20. Is there evidence of an interaction between maternal basking and food availability? In other words: Does the effect of maternal basking appear to differ for the two food availability groups? Does the effect of food availability appear to differ for the two maternal basking groups? Explain how you are deciding in each case.

Notice from the hypothetical interaction plots that sometimes there are both main effects and an interaction, but sometimes there is only an interaction, or only one main effect and an interaction.

Now reconsider the graph for the actual data, adding the connecting lines.

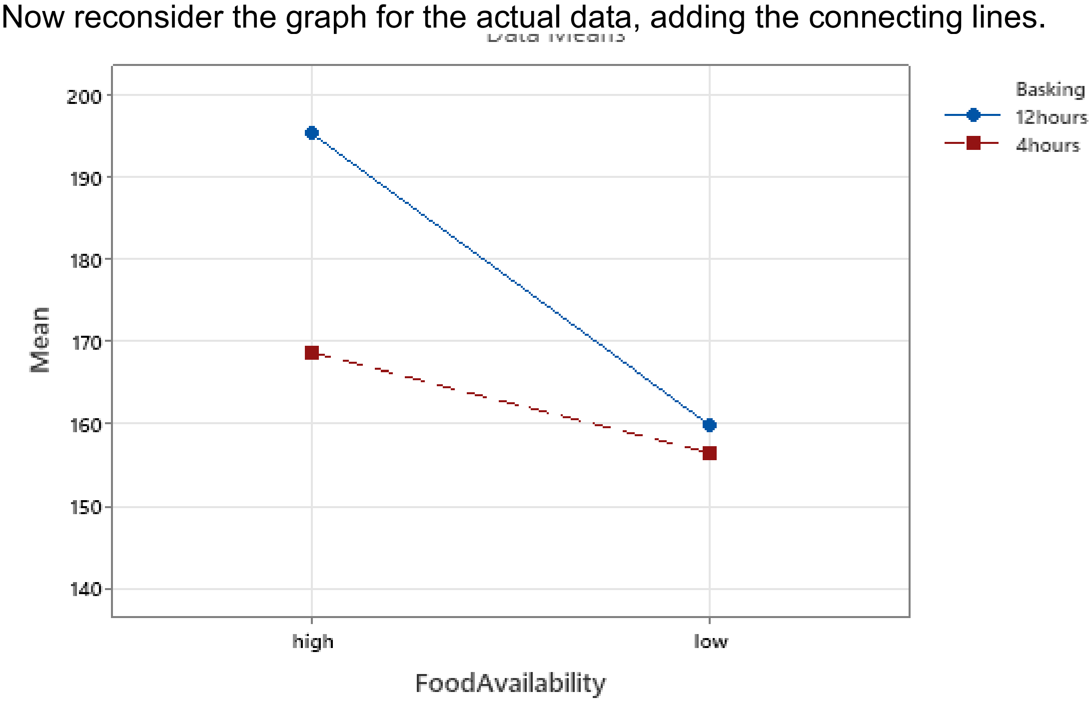


FIGURE 4: Interaction plot for the skink study

21. Describe the nature of the interaction revealed by this graph. That is, how does the food availability variable impact the effect of the basking variable or vice versa?

22. Figure 5 swaps the roles of the explanatory variables in the interaction plot. How would you describe the nature of the interaction from this plot?

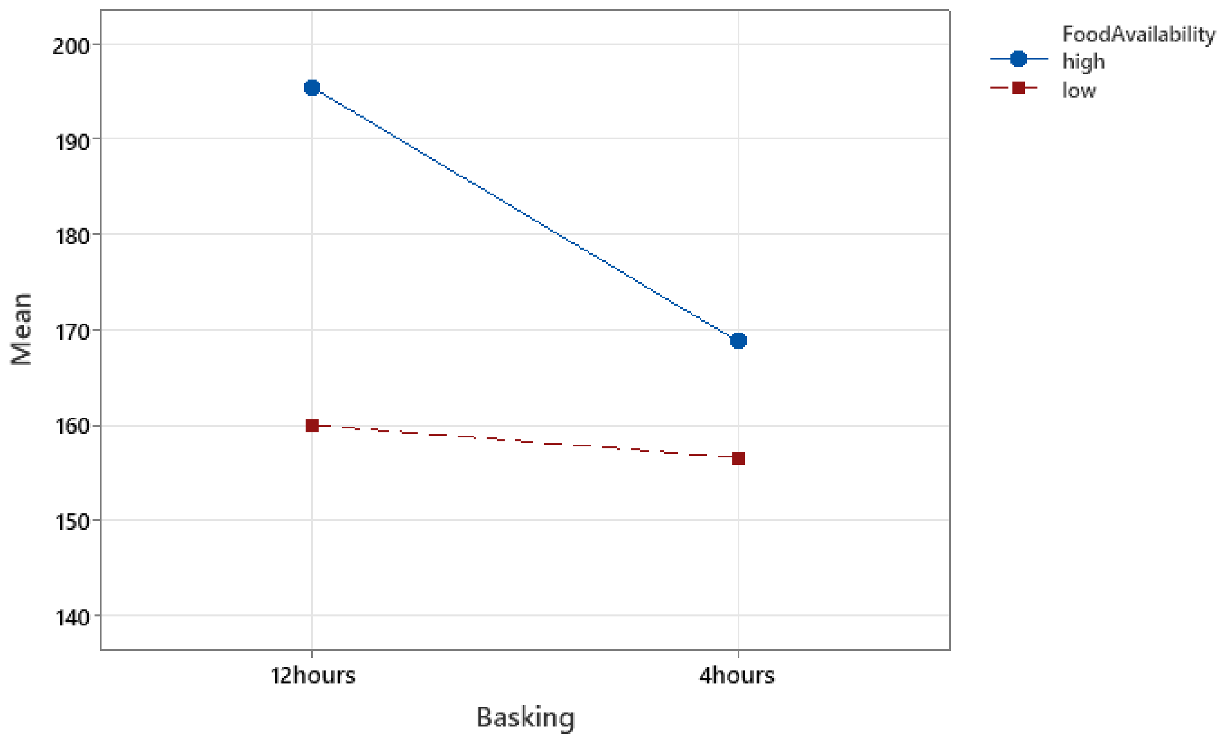


FIGURE 5: The other interaction plot for the skink study

To interpret the interaction in this study, we can say that the amount of food available to the pregnant skink modifies (or changes) the effect of the amount of maternal basking time on the offspring body mass. In particular, the difference in the mean offspring body mass between 12 hours and 4 hours basking time is much larger for high food availability than for low food availability. Equivalently, we can say that the difference in mean offspring body mass for high and low food availability is much larger when maternal basking time is 12 hours than when maternal basking time is 4 hours. When we have a statistical interaction in the data, our analysis tends to focus on the interpretation and the significance of the interaction, rather than the interpretation and significance of the main effects.

Is the Interaction Statistically Significant?

Because the observed means in Figure 4 show some evidence of an interaction (Figure 4 is somewhere in between the two hypothetical cases in Figures 2 and 3), we now want to decide whether the observed interaction is statistically significant. Keep in mind that the interaction plot tells us nothing about the amount of within treatment variation or the sample sizes. The interaction plot is also scaled so that it does not start at 0 on the *y*-axis so small differences appear large!

To test the significance of the statistical interaction, the first thing to do is to state the null and alternative hypotheses about this interaction. There are several ways to do so, focusing on H_0: no interaction vs. H_a: is an interaction.

23. Fill in the blanks below to indicate a two-sided alternative hypothesis for each of two equivalent ways of starting the hypotheses about the interaction.

H_0: There is no interaction between _________ and _____________ on the offspring body mass in the population of Southern Grass skinks.

H_a: There _________________between__________ and _____________ on the offspring body mass in a population of Southern Grass skinks.

Or

H_0: The effect of ___________________ on the offspring body mass is the same regardless of ____________ in a population of Southern Grass skinks.

H_a: The effect of _________________ on the offspring body mass is not the same regardless of _______________in a population of Southern Grass skinks.

24. Outline a simulation-based approach for deciding whether the interaction effect is statistically significant. In particular, what statistic could you use to measure the strength of the interaction in the observed data? Regardless of what statistic you choose to use, how would you shuffle the data (simulate the null hypothesis)? Why? How will you evaluate the strength of evidence?

Choice of Statistic

As always, there are different options for a choice of statistic. Let's consider one intuitive option now.

25. For the actual study data:

    For low food availability, what is the difference in mean offspring body mass for 12 hours and 4 hours (12 hours – 4 hours)?

    For high food availability, what is the difference in mean offspring body mass for 12 hours and 4 hours (12 hours – 4 hours)?

    Are the two previous values the same? What is the "difference of the (two) differences" you just computed (low – high)?

26. Explain how the difference in the differences is measuring the potential interaction in the data. *(Hint*: What would the value of the difference in the differences be if the null hypothesis is true? If the null hypothesis is not true?)

Definition: The ***difference in the differences*** is an intuitive statistic that summarizes the size of a statistical interaction by evaluating the impact of one explanatory variable within each group of the second explanatory variable separately and then seeing how different those impacts are across the separate groups.

27. How does the value of the difference in the differences in Question #26 compare to the value of the difference in the differences in the hypothetical interaction plots in Figures 3.2.18 and 3.2.19?

Return to the Two-variable ANOVA applet and verify your "observed difference in differences" (difference in the differences) calculation. Check the **Show Shuffle Options** box. Press **Shuffle Responses** a few times, with the **Data** radio button selected. Notice that the applet re-assigns the response outcomes to the same factor-level combinations. Now press **Shuffle Responses** a few times with the **Graphs** radio button selected.

28. What do you notice about the behavior of the interaction plot for the re-randomized data compared to the original data?

29. Generate at least 1000 shuffles and sketch a graph of the re-randomized distribution of the difference in differences statistic.

30. Use the simulation results to decide whether the observed difference in the differences (the interaction between maternal basking and food availability) is statistically significant. (Is your p-value one-sided or two-sided? Why?) Summarize your conclusions in context.

An important thing to recognize about the full factorial design is that it allows for maximum statistical power, not only in evaluating potential main effects, but also when testing interactions.

Key Idea: A full factorial design is an optimal design because it maximizes the ability to find the effects of each explanatory variable on its own (rather than dividing the number of subjects into separate studies for each variable), while also maximizing the ability to find statistical interactions.

Estimating Interaction Effects

The small p-value tells us that when predicting offspring body mass, it's not enough to "add together" the effect of food availability and the effect of maternal basking. Instead, we need a model that reflects that the effect of maternal basking changes with the level of food availability and vice versa. Thus, we need to include additional "interaction effects" in our prediction equation to improve our predictions.

Calculating these *interaction effects* turns out to be straightforward in a balanced, factorial design like we have here.

31. The model below shows the two-variable model for predicting offspring body mass without including the interaction (from #5). The predicted treatment means for offspring body mass from this model are shown in the "Predicted" table and the difference between the observed and predicted treatment means are shown in the "Differences" table.

predicted offspring body mass = $170.11 + \begin{cases} -7.52 & if\ 4\ hours \\ 7.52 & if\ 12\ hours \end{cases} + \begin{cases} -11.91 & if\ low \\ 11.91 & if\ high \end{cases}$

The interaction effects are the differences in the observed treatment means and the treatment means predicted by the two-variable model without interaction (from #8):

Interaction effects = Differences in differences
= Observed treatment mean – Predicted treatment mean

Observed

	Low	High
4 hours	156.50	168.68
12 hours	159.89	195.36

Predicted

	Low	High
4 hours	150.68	174.50
12 hours	165.72	189.53

Differences

	Low	High
4 hours	5.82	-5.82
12 hours	-5.82	5.82

Key Idea: In a balanced design, with the same number of observations in each cell, the interaction effects will sum to zero across the rows, down the columns, and overall.

32. Write a new prediction equation that uses the overall mean, the main effects for maternal basking and the main effects of food availability, *and the interaction effects.* Confirm that this prediction equation produces the four observed treatment means.

Key Idea: A two-variable model (with no interaction) assumes that the effects of one variable do not depend on the values of the other variable. A two-variable model with an interaction allows those effects to change depending on the level of the other explanatory variable.

The ANOVA table and *F*-statistic for the interaction

Use the **Two-variable ANOVA** applet and check the **Show ANOVA Table** box to yield an ANOVA table that includes a row for the interaction. Include a copy of the table below.

33. What is the Sum of Squares for the interaction effects? What percentage of all variation is explained by the Interaction term? Is this statistically significant? How are you deciding?

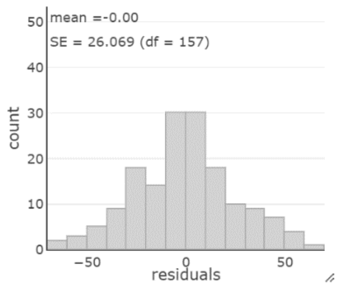
34. Considering how many observations were in each cell, calculate the (weighted) sum of squares for the interaction effects. Confirm that this matches the *SSInteraction* you found in the applet.

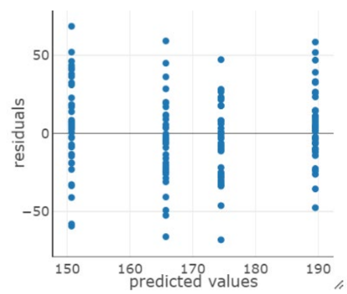
To help decide whether we can trust the theory-based p-value reported in the ANOVA table for the interaction term, we can use simulation to produce a null distribution for the *F*-statistic on the interaction. This simulation assumes any of the observed body mass values could be assigned to any of the 4 treatments.

In the applet, use the Statistic pull-down menu on the far left to change the simulated null distribution from difference in differences to *F*-statistic for the interaction.

35. Use the null distribution of the *F*-statistic to estimate the simulation-based p-value and compare to the theory-based p-value. Does the theory-based p-value adequately approximate the simulation-based p-value?

As we've seen before, theory-based p-values from the *F*-distribution will be similar to those obtained by simulation when certain validity conditions are met. The residual plots for the two-variable model including the interaction are as follows.





36. Does the distribution of the residuals appear symmetric? In the graph of the residuals vs. the predicted values, are the residuals equally spread for each fitted value?

Follow-up Confidence Intervals

In the analysis so far, whether from the simulation analysis or the theory-based analysis, we can conclude there is a significant main effect of maternal basking (p-value = 0.0003), a significant main effect of food availability (p-value < 0.0001), and a significant interaction between maternal basking and food availability on offspring body mass (p-value = 0.0044).

When we have a significant interaction, we need to be cautious in interpreting "main effects." Instead of comparing "low to high" or "4 hours to 12 hours," we will compare the treatments to each other. With 4 treatment groups, we have six possible pairwise comparisons. However, not all of these pairwise comparisons may be of interest.

Here are 95% confidence intervals for all six different treatment group comparisons:

```
4hours_low - 12hours_low: (-14.64, 7.86)
4hours_low - 4hours_high: (-23.43, -0.93)*
4hours_low - 12hours_high: (-50.11, -27.61)*
12hours_low - 4hours_high: (-20.04, 2.46)
12hours_low - 12hours_high: (-46.72, -24.22)*
4hours_high - 12hours_high: (-37.93, -15.43)*
```

37. Based on the treatment means (question 4), which treatment(s) saw the highest mean offspring body mass? Based on the confidence intervals, is this treatment significantly different from any of the other treatments?

38. Based on the confidence intervals, do you conclude about the difference in mean offspring body mass for 4 hours and 12 hours basking if food availability is low?

39. Based on the confidence intervals, what do you conclude about the difference in mean offspring body mass for 4 hours and 12 hours basking if food availability is high?

40. Summarize the conclusions you would draw from this study. Be sure to talk about significance, estimation (confidence intervals), generalizability, and causation.