Exploration 3.1: Pig Growth

Section 3.1 Learning Goals

- Design an experiment with more than one variable of interest.
- Explore the benefits of a two-variable study where the levels of both variables are assigned by the researcher.

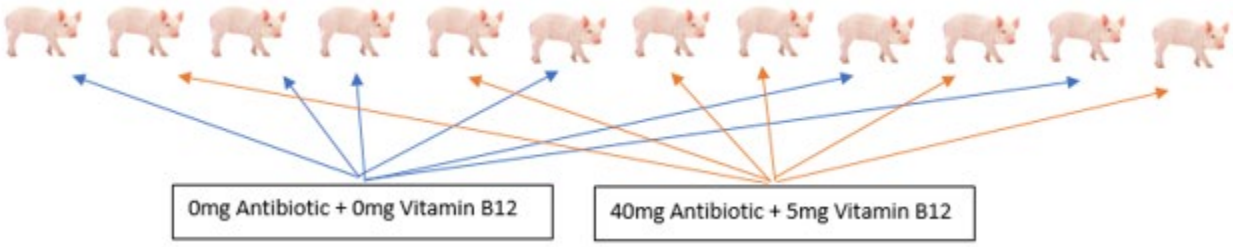
Livestock farmers want their animals to be healthy and grow under the best conditions. It makes sense that adding vitamins to the pigs' diet should produce larger pigs faster. However, perhaps an antibiotic can also impact growth by controlling the bacteria in the pigs' intestines. An experiment conducted with 12 baby pigs assessed the effects of adding 0 mg or 40 mg of an antibiotic to the pig diet per day, as well as the effects of 0 mg or 5 mg of vitamin B12 per day. After 4 weeks, the average daily weight gain (ADG), in pounds per day, was determined for each pig. For pigs, higher ADG is better.

Step 1: Ask a research question. Does adding either an antibiotic and/or vitamin B12 to pig diets affect average daily weight gain of baby pigs?

Step 2: Design a study and collect data.

1. What is the response variable? What is/are the explanatory variable(s) (aka "factor(s)") of interest? Do you recommend an experiment or an observational study to investigate this research question?

2. Consider this proposed study design: Randomly assign 6 pigs to receive 0 mg of antibiotic per day and 0 mg of vitamin B12 per day (the control group), and 6 pigs to receive 40 mg of antibiotic per day and 5 mg of vitamin B12 per day (as depicted in the following picture).

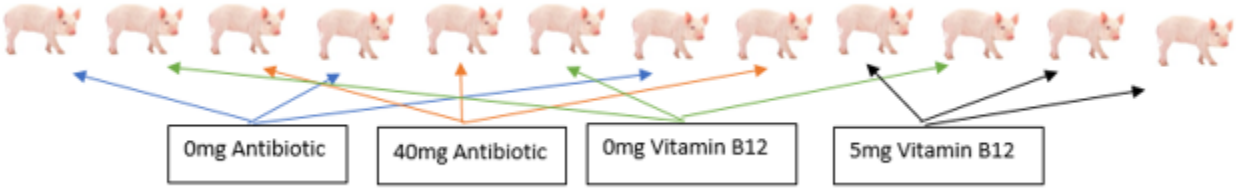


Explain why this is not a very good design.

These materials were developed by the STUB Network and supported by the National Science Foundation under Grant NSF-DBI 1730668. They are covered under the Creative Commons license BY-NC which allows users to distribute, adapt, and build upon the materials for noncommercial purposes only, and only so long as attribution is given to the STUB Network.

Key Idea: One key component of any good experimental design is to make sure that *only one variable changes at a time*. Otherwise, we can't isolate the individual impact of different variables.

3. Consider this proposed study design: Randomly assign 3 pigs to receive 0 mg of antibiotic per day; 3 pigs to receive 40 mg of antibiotic per day (no vitamins); 3 pigs to receive 0 mg of vitamin B12 per day and 3 pigs to receive 5 mg of vitamin B12 per day (no antibiotic) (as depicted in the following figure).



Identify one way this study is better than the first study. Would you recommend this study design? Explain why or why not.

Key Idea: A second key component of a good experimental design is maximizing the number of observations involved in each comparison.

While the previous design in #5 allows us to estimate the impact of each variable individually, these comparisons are only made on 6 pigs, 3 in each group. To increase the sample size, we can explore changing both variables on the same pigs, but we need to ensure by our design that the explanatory variables are not associated with each other (e.g., the vitamin B12 comparison isn't impacted by which pigs are receiving antibiotic). The most efficient study design to explore both variables at once is a ***factorial design***. In this case, each pig is assigned one of four ***treatments*.** The four treatments are the four possible factor-level combinations of the antibiotic and vitamin B12:

- Treatment 1: Antibiotic 0 mg, vitamin B12 0 mg
- Treatment 2: Antibiotic 40 mg, vitamin B12 0 mg
- Treatment 3: Antibiotic 0 mg, vitamin B12 5 mg
- Treatment 4: Antibiotic 40 mg, vitamin B12 5 mg

This type of study design is called a "2 x 2 factorial design." (The first "2" indicates the first explanatory variable has 2 levels and the second "2" indicates the second explanatory variable also has 2 levels.)

4. How many of the 12 pigs would you assign to each treatment? How would you make this assignment?

Definition: With multiple explanatory variables, also called factors, a ***full factorial design*** examines all possible ***factor*** (variable)-level combinations (called ***treatments***). A ***balanced*** design assigns the same number of experimental units (***replications***) to each treatment; this ensures that the two explanatory variables will not be associated with each other. The experimental units should be randomly assigned to each treatment.

Consider the fabricated data in ***pigADG*** as the results from one such study.

Step 3: Explore the data: Copy and paste the data (all four columns) into the **Multiple Variables** applet. (Or **Clear** the initial data and type pigADG.txt in the window and press **Use Data.**) Drag *ADG* to the Response box. Check the **Show Descriptive** and **Show boxplot** boxes.

5. Sketch the boxplot of the ADG for these 12 pigs. Report the overall mean ADG, along with the standard deviation of ADG.

Drag the *Treatment* variable to the **Subset By** box.

6. How many pigs were assigned to each treatment? Does which treatment the pigs receive appear to explain variation in ADG? What proportion of variation in ADG do the treatments explain?

7. Record the mean and standard deviation of ADG for the four treatments. Which treatment appears to be the best? The worst?

	Antibiotic 40 mg, Vitamin B12 5 mg	**Antibiotic 40 mg, Vitamin B12 0 mg**	**Antibiotic 0 mg, Vitamin B12 5 mg**	**Antibiotic -0 mg, Vitamin B12 0 mg**
Mean ADG (lbs/day)				
SD of ADG (lbs/day)				

Best: Worst:

Remove *Treatment* from the Subset By box. Drag the *Antibiotic* variable to the Explanatory box and the *VitaminB12* variable to the Explanatory box. Place the VitaminB12 variable above the Antibiotic variable in the Explanatory box. Select Show 2-variable graphs.

8. What proportion of pigs in each Antibiotic group were given 5 mg of vitamin B12? What proportion of pigs in each Antibiotic group were given 0 mg of B12? Explain why the two explanatory variables *Antibiotic* and *VitaminB12* are not associated in this study design.

Step 4: Draw inferences beyond the data. The above analysis tells us that, in this sample, the best pig diet appears to use both vitamin B12 and the antibiotic, that is the 5mg x 40mg treatment combination. But our next questions should be whether there are any statistically significant differences in the treatment means, especially considering the small group sizes, and if so, how do we estimate the differences in ADG across the treatments.

Analysis #1: One-Variable Model Using the Treatments

Remove the Antibiotic and VitaminB12 variables from the Explanatory box and return *Treatment* to the **Subset By** box. Check the box to display the ANOVA table.

9. Does the treatment variable explain a statistically significant amount of variation in the pig growths? Cite a test statistic and p-value to support your answer. What is the residual standard error for this one-variable model?

One disadvantage to this analysis is we can't tell the researchers whether there is a "vitamin effect" or an "antibiotic effect." We can only talk about the effects of each of the four treatments.

It turns out that by carefully setting up the balanced design, we will be able to estimate the individual effects as well.

Analysis #2: Two-Variable Model Using *Antibiotic* and *VitaminB12*

To begin, we will conduct one-variable analyses using each variable separately.

Fit the one-variable model using *Antibiotic*: Remove *Treatment* and place the *Antibiotic* variable in the Explanatory variable box.

10. From the graph: What is the mean ADG for the 0 mg antibiotic pigs? What is the mean ADG for the 40 mg of antibiotic pigs? From the chart: What proportion of variation in pig growth is explained by the antibiotic? From the ANOVA table: What is *SSantibiotic?*

    Mean ADG for 0 mg antibiotic: Mean ADG for 40 mg antibiotic:

    Proportion of variation explained: Sum of squares for antibiotic:

Check the **Statistical Model** box.

11. What is the prediction equation for mean pig growth based on the antibiotic treatment given?

$$\textit{Predicted ADG} = \_\_\_\_\_ + \begin{cases} if\ given\ 40mg\ antibiotic \\ if\ given\ 0\ mg\ antibiotic \end{cases}$$

Fit the one-variable model using vitamin B12: Remove *Antibiotic* and place the *VitaminB12* variable in the **Explanatory variable** box.

12. What is the mean ADG for the 0 mg vitamin B12 pigs? What is the mean ADG for the 5 mg vitamin B12 pigs? What proportion of variation in pig growth is explained by vitamin B12? What is the *SSvitaminB12*? What is the prediction equation?

    Mean ADG for 0 mg vitamin B12: Mean ADG for 5 mg vitamin B12:

Proportion of variation explained: Sum of Squares for vitamin B12:

$$\textit{Predicted ADG} = \_\_\_\_\_ + \begin{cases} if\ given\ 5\ mg\ vitamin\ B12 \\ if\ given\ 0\ mg\ vitamin\ B12 \end{cases}$$

Definition: In a balanced factorial design, we can calculate ***main effects*** by comparing the mean of each level of a factor (averaged over the categories of the other factor) to the overall mean. In other words, we can use the row means and the column means (see table below) to calculate these effects. In a balanced design, these row and column averages are exactly what we would find if we simply averaged the groups within each of the variables, ignoring the other variable.

13. The table below shows the four treatment means from #7. Compute the column means for Antibiotic by averaging across the two B12 values. From these means (and their overall mean), calculate the Antibiotic effects. How do these values compare to what you found in #11? Compute the row means for Vitamin B12 and the effects (from the average of the row means). How do they compare to what you found in #12? How does the average of the row means compare to the average of the column means?

Vitamin B12	**Antibiotic**		**Row Means**	**Effects**
	0 mg	**40 mg**		
0 mg	0.957	1.197		
5 mg	1.400	1.587		
Column Means				
Effects				

When there is no association between the two explanatory variables, you can simply combine the two one-variable models into a two-variable model.

Now use the **Multiple Variables** applet to conduct a two-variable analysis by putting both *VitaminB12* and *Antibiotic* in the Explanatory variables box.

14. Fill in the ANOVA table below.

Source	**DF**	**SS**	**MS**	**F**	**p-value**
Model					
VitaminB12					
Antibiotic					
Error					
Total					

14. Compared to the one-variable models, did the sums of squares change for VitaminB12 or for Antibiotic?

15. Compared to the one-variable models, did the *F*-statistics change for VitaminB12 or for Antibiotic? Why? What is a critical difference between the one-variable models and the two-variable model?

16. What proportion of the total variation in ADG is explained by the two-variable model? That is, what proportion of the total variation in ADG is explained by these variables together (the "model R^2")? How does this compare to the sum of the R^2 values from the two one-variable models?

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

Key Idea: Because of the balanced study design, the "adjusted effects" are the same as the unadjusted effects and we can "add together" the effects of each variable in predicting the response.

Check the **Statistical model** box.

18. Use the output to create the combined prediction equation. (*Hint*: Focus on the values in the "Coef" column; they should look familiar!)

Predicted ADG = ___ + $\begin{cases} if\, given\ 40mg\ antibiotic \\ if\, given\ 0mg\ antibiotic \end{cases}$ + $\begin{cases} if\ given\ 5mg\ vitamin\ B12 \\ if\ given\ 0mg\ vitamin\ B12 \end{cases}$

What is the residual standard error for this model?

19. How does the residual standard error for this model compare to the values from the one-variable models?

20. Use your model to estimate the predicted ADG for a pig given both the antibiotic (40mg) and vitamin B12 (5mg). Include units in your answer.

21. How does your model prediction from the previous question compare to the actual treatment mean?

Of course, not every pig will grow the same amount as the mean of the treatment.

22. Calculate the residual for the first pig in the dataset. Interpret this residual in the context of this study.

Validity Conditions

For the theory-based p-value to be a valid p-value, we need three validity conditions to be met.

1. The responses are independent between and within the treatment groups.
2. The standard deviations of the responses are approximately equal among the treatment groups.
3. The distributions of the responses are approximately symmetric for all treatment groups or the samples sizes are at least 20 without strong skewness or outliers for all treatment groups.

To check the second and third validity conditions we will use the model residuals.

Key Idea:

- If the responses in each treatment are approximately symmetric, then a graph of the residuals will be approximately symmetric.
- If the responses in each treatment have a similar standard deviation, then a graph of the residuals vs. the predicted values for each treatment will exhibit a similar spread in each treatment.

Click on the **Show residuals** button in the **Multiple Variables** applet.

23. One of the validity conditions for the *F*-tests is that the distributions of the responses within all four treatments are symmetric, meaning after adjusting for the group means, all the residuals should be coming from the same symmetric distribution. Using the distribution of residuals. does this condition appear to be met? Be clear about how you are deciding.

24. Another validity condition is that the responses in each treatment have a similar standard deviation. Scroll down to examine the residuals vs. predicted values graph. Does this condition appear to be met? (Also consider your output in Question #6.)

25. The third validity condition is that the responses between and within each of the treatments are independent. Is this validity condition met? Why or why not?

Step 5: Formulate conclusions.

26. For these fabricated data, which variable (*VitaminB12* or *Antibiotic*) has a larger effect on average ADG? Which explains a greater proportion of variation in ADG?

Step 6: Look back and ahead. These were fabricated data, so let's focus on evaluating the study design.

27. Would these be causal "effects" or is a cause and effect conclusion not possible here? Explain.

28. Summarize the main advantages of the study design (balanced factorial design) and the two-variable analysis using both the antibiotic and vitamin B12 simultaneously over one-variable designs or using the treatment variable.