

**STA 302 Fall 2024 Final Project Part 2**  
**Analysis Plan Flowchart**  
**Due: October 25, 2024, by 8:00PM ET**  
**Latest Acceptance: November 1, 2024, by 8:00PM ET**

---

*Please note that if you intend to use **one of the three NQA extensions**, **you should not submit any documents prior to the posted deadline** as Quercus will not allow any changes or additions to the submission after the initial deadline. Instead, make sure you have all your documents prepared and ready to submit all at once.*

---

Goal of the Assessment:	Learning Outcomes being Assessed:
<ul style="list-style-type: none"><li>• To create a comprehensive analysis plan for completing the final project</li><li>• To receive feedback regarding appropriate use of linear regression tools</li><li>• To create a scaffold of the necessary information for reporting methods and results of an analysis.</li><li>• To synthesize the course material to better understand each method as part of the whole linear regression analysis cycle.</li></ul>	<ul style="list-style-type: none"><li>• Report the results of a residual plot analysis and recommend a course of action.</li><li>• Select and apply an appropriate transformation to correct a model violation in a given dataset/model.</li><li>• Explain the limitations of hypothesis tests and confidence intervals arising from violated assumptions.</li><li>• Integrate decomposition-based hypothesis tests into decision process when building models.</li><li>• Develop a detailed analysis plan in which assumptions are verified and corrected at appropriate places</li></ul>

**Overview of Assessment:**

Your group is tasked with creating a [flowchart](#) that outlines the steps you will take in your final project analysis to arrive at an answer to your proposed research question. Since you have already fit a preliminary model in your proposal, **your flowchart should start with this first model** and outline the steps you will take to arrive at a final model that answers your research question, based on the content covered thus far.

**You will not need to write code or do what you describe in your flowchart for this part of the assessment.** Rather focus on describing how you will build models and make decisions to reduce or increase the size of your model, as well as how you will incorporate model diagnostics. This means you will need to think about the **order** (i.e., where in the analysis it makes sense to do these steps) in which you use certain linear regression methods and what **process you will follow/decisions you will make** depending on the results of these methods.



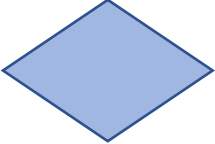


The plan that you and your group are outlining will be fairly general – while you should refer to specific variable names and properties of your dataset where appropriate, you will not necessarily know the outcome of any decision step. Rather you should consider all possible outcomes of these decisions and show us in each case what your steps would be. **You should focus on making each element of your flowchart (e.g. process, decision, etc.) clear and specific**, with enough detail for someone to follow your steps and get to the same answer as you.

For example, instead of writing “check assumptions”, be specific about how (which specific residual plots) and what you’re checking (looking for what specific patterns)). **Your audience should have enough information to understand each step without needing to fill in gaps with their own knowledge of regression.** Be sure that all your steps and decisions are written using content that is consistent with the course materials.

## What is a Flowchart:

A flowchart consists of different shaped boxes in which you write steps/decisions corresponding to a process in your analysis. Arrows connect these boxes and point in the direction that you would proceed after each step. You can create loops and branches in your flowchart if you need to repeat a certain process until the desired result is achieved (loop) or if, depending on the result, you will take a different approach (branch). There are specific shapes for the boxes that have specific meanings which you should use (see below, table 1). These boxes should contain text that describes what is occurring at this step (e.g., check assumption A by creating residual plot B and looking for pattern C).

Table 1 - Flowchart Elements and their meaning/purpose

Flowchart Elements	Meaning and Function
	This shape indicates the starting point or ending point in your flowchart. You should ensure it is clearly written inside the box what the starting point of your process is and exactly where it ends.
	This shape indicates that you are applying a certain method/process in your analysis (often called an action). The text inside should specify exactly what action you are taking at this step.
	This shape indicates you must make a decision. The text inside this shape should refer to what question you are asking at this step and should have arrows leading out indicating what you would do depending on whether the answer is yes/no, for example.
	If you need to continue your chart on another page/on another part of the same page, you can use a circle to indicate that there is a break in your flowchart at the end of one page/step and use the same circle as the starting spot on the new page/area. You should include a label inside that is the same on both pages/areas so that it's obvious this links the two halves of the flowchart together.
	The arrow is what connects the steps together and shows how you will proceed through your analysis. Arrows are unidirectional but can be used to refer to a previous step (if looping is needed).

Some additional references you may refer to for help in using the elements of a flowchart are:

- [details and examples of flowchart elements](#)
- some [examples of the types of flowchart structures](#) you may wish to consider (e.g., how to write an iterative or branching step)

These can be used to help you decide how you might want to draw your flowchart (these websites refer to paid software, but don't purchase it). An example flowchart has been provided on the Quercus assignment page, which is not an example of a good flowchart, but shows you how you might want to think about organizing your own analysis.

## How to Create your Flowchart

Your group may create your flowchart in any way that you are comfortable. Examples include:

- creating it in Word (if your group saves it on someone's OneDrive, all members would be able to work on it online) or some other word processing program,
- creating it in powerpoint, keynote, or other program for creating presentations/slides
- graphics-based software if you have access, or
- free version of the online flowchart creator at [lucidchart.com](https://lucidchart.com)

Some general guidelines for designing a flowchart:

- Start with a rough outline to decide the overall structure/order of steps in your flowchart.
- Be sure to consider what step or action you will take depending on if you answer Yes or No to a decision step.
- If you plan to perform a specific action multiple times, consider using a loop (i.e., having your flowchart circle back to an earlier step) to make your flowchart more concise.
- Fill in your outline with plenty of detail, describing exactly what you would do and how you would come to a decision or conclusion at each step (e.g., how will you check a specific assumption, with what tool and how would a conclusion be made)
  - **Don't use R function names** to describe how each step is being conducted – instead, refer to the information in the lecture slides and outline the process in words.
- Consider colour-coding your flowchart elements to correspond to your legend for easier reading.

## Submission Requirements:

Your final flowchart should be **submitted to Quercus** by the deadline above. Your group submission should have the following characteristics:

- Be no more than 2 pages (of standard letter-sized paper, 8.5x11 inches) in length
- Should be saved as a PDF file
- All components and text should be legible and easy to read.
- A legend is included displaying the Module of course material each step in the flowchart covers.
- Your flowchart should include at least one topic from each of the below modules of course material:
  - **Module 1&2** – Fitting/Estimating Linear Models (simple or multiple), categorical predictors.
  - **Module 3** – Assumptions of linear regression, how to check them, additional conditions.
  - **Module 4** – Correcting violated assumptions with transformations.
  - **Module 5** – Hypothesis test/Confidence intervals on coefficients and mean response, prediction intervals.
  - **Module 6** – Decomposition of sum of squares, ANOVA test of overall significance, partial F test.
- **Each group member's name is listed and a description of their contribution to the flowchart is outlined.**
- Your research question from part 1 is listed at the top of your flowchart and your flowchart ends with a clear means of answering your research question.

Failure to meet the submission requirements will result in a 1-point deduction on the grade of this assignment.

**\*\*Group members who do not participate in the creation of the flowchart will receive a grade of zero\*\***

Criteria of Assessment	Excellent (3 points)	Satisfactory (2 points)	Revision needed (1 point)	Incomplete (0 points)
Flowchart appearance:				
<ul style="list-style-type: none"> <li>Does not exceed two 8.5x11 inch pages in length.</li> <li>Legend is provided correctly distinguishing each module of course material required.</li> <li>Legible and easy to read without excessive magnification</li> </ul>	All three criteria are met.	Only two criteria are met.	Only one criterion is met.	No criteria are met.
Use of flowchart elements (see Table 1):				
<ul style="list-style-type: none"> <li>All decision steps are questions answerable by yes/no and denoted using correct shape.</li> <li>All arrows are unidirectional, point in the correct direction and are properly labelled when exiting decision step.</li> <li>All action steps are denoted using correct shape.</li> <li>Start/End points (and, where applicable, continuations) are denoted using correct shape with appropriate labels.</li> </ul>	All four criteria are met.	Only three criteria are met.	Only two criterion is met.	Only one or fewer criteria are met.
Modules 1 & 2 Content:				
<ul style="list-style-type: none"> <li>An initial/preliminary model is explicitly provided and is described in the context of the research question.</li> <li>Steps that involve changing the initial model explicitly indicate that a new model has been fit to reflect this change with the exact change noted.</li> <li>Categorical (not individual indicator) variables in each model should be treated appropriately throughout the steps of the flowchart (e.g., tests applied correctly, plots appropriate to variable type, etc.)</li> </ul>	All three criteria are met.	Only two criteria are met.	Only one criterion is met.	No criteria are met.
Module 3 Content:				
<ul style="list-style-type: none"> <li>All assumptions/conditions are assessed and explicitly mentioned.</li> <li>Tools used to assess each assumption/condition are appropriate and based on instructor-created materials and are explicitly described using appropriate course terminology.</li> <li>How violated assumptions/conditions are identified is explicitly described using appropriate course terminology and is correct.</li> </ul>	All three criteria are met.	Only two criteria are met.	Only one criterion is met.	No criteria are met.

Module 4 Content:				
<ul style="list-style-type: none"> <li>Each violated assumption is addressed.</li> <li>Choice of tool to address each violated assumption is appropriate.</li> <li>How tool will be used to address each violated assumption is described correctly using appropriate course terminology.</li> </ul>	All three criteria are met.	Only two criteria are met.	Only one criterion is met.	No criteria are met.
Modules 5 & 6 Content:				
<ul style="list-style-type: none"> <li>Inferential tool being used is explicitly described using correct terminology from the course and is identifiable by the reader.</li> <li>How conclusions are made using inferential tool are correct and explicitly stated.</li> <li>Conclusions from inferential tool are valid based on previous steps.</li> </ul>	All three criteria are met.	Only two criteria are met.	Only one criterion is met.	No criteria are met.
Research question:				
<ul style="list-style-type: none"> <li>The research question being addressed (like you did in final project part 1) is explicitly stated with identifiable response and predictor variables at the beginning of the flowchart.</li> <li>For each endpoint listed in the flowchart, it is explicitly stated how the steps leading to the endpoint could yield an answer to your research question.</li> <li>The potential answers to the research question are appropriate based on previous steps.</li> </ul>	All three criteria are met.	Only two criteria are met.	Only one criterion is met.	No criteria are met.
<b>TOTAL:</b>	<b>/21</b>			