Project 2 Prospectus

STA 315 W21

For Project 2, you will design, carry out, and analyze the data from an experiment (like you did in Project 1). The additional requirement for Project 2 is that your treatments come from crossing two factors. Your Project 2 experiment may be about the same topic as your project 1 experiment if you wish, but it does not have to be.

Fill in your answers to the items below to describe your proposed experiment for Project 2. Please leave the original questions in your document. (The amount of space I’ve left after each item is not indicative of how long your answers should be. Generally, the more detailed the answers, the better.)

General Question

* What is the general question you seek to answer?  
    
  On my laptop, how can I get the best synthetic benchmark scores for my CPU depending on the voltage allowed to my CPU (this is measured in an offset based on the standard voltage to the CPU) and clock speeds for the CPU (3.3,3.4 ect).

What is your motivation for studying this question?  
My motivation for studying this is I love playing with computers and trying to overclock (OC) and maximize performance on my laptop.

3 Choices

Identify your **response**. (Guidelines: needs to be measurable, univariate and of interval/ratio type, have validity and reliability)  
  
Use the synthetic benchmarks ‘Cinebench R23’ to objectively measure CPU performance.

* Identify your conditions/treatments:
  + What **2 factors** are you going to study and what are the **levels** of each factor? (Note: blocking factors do not count here; both of the 2 factors should be “factors of interest”.)  
     The factors of interest that I am going to be investigating in this study is going to **voltage to the CPU** (measured in an offset so I will actually be ‘down-volting’ the CPU by giving it less power so it wont over heat which should give extra performance to the CPU) and the **clock speed** of the CPU (3.4, 3.5 ect)
  + Identify each of these factors as an **o­­­bservational factor** or an **experimental factor**. For an observational factor, the levels are “built into” the units, like the conditions in an observational study; for an experimental factor, the levels are assigned to the units, like the conditions in an experiment. **It is required that at least one of the factors be an experimental factor**.  
     Both factors in my experiment are going to be experimental factors as I am going to be artificially setting the speed of the CPU clock speed and the number for the offset voltage, which will be negative since I’m going to be down-volting my CPU.
  + Replace the entries of the table below with the factors, levels, and treatments in your experiment (adding rows and/or columns for more levels if necessary):

|  |  |  |  |
| --- | --- | --- | --- |
| Offset Voltage (Mv) of CPU | CPU Clock Speed (GHz) | |  |
| 3.3 | 3.4 | 3.5 |
| 0 (essentially default voltage) | 0, 3.3 | 0, 3.4 | 0, 3.5 |
| -80.1 | -80.1, 3.3 | -80.1, 3.4 | -80.1, 3.5 |
| -160.2 | -160.2, 3.3 | -160.2, 3.4 | 3.5 |

* Identify the experimental units. (This is what you assign the treatments to.)  
    
  In this case the experimental unit would be the intel core i7 8750H on my MSI GS 65 Stealth thin Laptop
* What is the specific question you seek to answer?(This comes from applying your choices of the response, conditions, and the material to the more general question you posed at the beginning.)

How is the performance of the CPU, measured through synthetic benchmark ‘Cinebench R23’, affected by it’s core clock speed and voltage offset when the fan speed is constant

* It is required that you complete an experiment, not an observational study. Explain why your study will be an experiment and not an observational study, as defined on Notes 1.
  + My study will be an experiment and not an observational study because I would be assigning the conditions.

Anticipating sources of variation, bias, and confounding

* Collect a few preliminary response values. (Do not include these values in the dataset you analyze for your project 2 report.) [If this is not possible (e.g. if it takes a week to collect your data), try to plan/think/learn about your data collection process.] What issues in carrying out the experiment did this process help you consider that you did not consider before?   
    
  Making sure I just leave the fan speed constantly at max at all times during testing should help get rid of any temperature variations between tests when I am changing the clock speed and CPU voltage.
* Describe your measurement process (i.e. of obtaining the response values), including what steps you will take to minimize measurement error.  
    
  The synthetic program for the CPU, Cinebench R23, will give me a performance score (like 6800) and as long as I saturate the computer with heat before testing and make sure I set the clock speed and voltage offset correctly there should no chance of a measurement error.
* To what degree are the units representative of the question you seek to answer? (It may be beyond the scope of your study to make them truly representative, but it is at least important to acknowledge that.)

While the CPU and the laptop I am using may have similar or exact components to other laptops with the same name and specifications, there is a thing in the industry called ‘Silicon lottery’ where the performance of each CPU and therefore computer can differ because the materials that make each CPU might be slightly different that might allow the same CPU in one computer to achieve higher scores than the same CPU in another computer. So, in theory my testing with my CPU should be representative with all other CPUs of the same spec but unforutnatly due to the ‘Silicon lottery’ that’s not really the case.

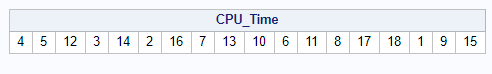
* To what degree are the units uniform (to limit variability)? What can you do to make them more uniform?  
    
  Since I’ll be using my laptop for these runs I’ll keep the clock speeds and the voltage of the CPU the same while making sure fan speeds do not change from run to run as well as having the computer sit on my desk every time I get test results so that my desk surface should heat up at the same pace every time. And I will be measuring the ambient temperature to make sure its roughly the same from run to run.
* Does blocking of units make sense for your experiment? Explain why or why not. If so, how you will block?  
    
  No, blocking of units does not make sense for my experiment since I do not have multiple experimental units
* What are other sources of variation and/or bias in the response? How will you control for them?  
    
  Ambient temperature could be something that affects performance scores so I have a thermometer I can use to measure ambient temperature to make sure it is similar every time I conduct these tests.
* Identify at least once nuisance influence that you will record in your dataset. After your experiment is over, you will be required to look back and analyze its effect. (Common options are trial number or block but it could also be other variables related to the data collection.)  
    
  The nuisance influence that I am going to record in my dataset is going to be CPU temperature because as temperature rises the CPU will clock down in (and therefore result in less performance) in order to cool itself down.

Random assignment / the Design

* How many treatments do you have? How many replications (units) per treatment? How many total observations? (You are required to have enough replications that your degrees of freedom of residual error is at least 6.)

|  |  |  |  |
| --- | --- | --- | --- |
| Offset Voltage (Mv) of CPU | CPU Clock Speed (GHz) | |  |
| 3.3 | 3.4 | 3.5 |
| 0 (essentially default voltage) | 0, 3.3 | 0, 3.4 | 0, 3.5 |
| -80.1 | -80.1, 3.3 | -80.1, 3.4 | -80.1, 3.5 |
| -160.2 | -160.2, 3.3 | -160.2, 3.4 | -160.2,13.5 |

Based on this diagram I will have 9 different treatments (3 for CPU and 3 for voltage) doing 2 replications per treatment would leave me with a total of 18 observations

* What is the design name? Your options are:
  + **two-way completely randomized (CR[2]) design – if no blocking**
  + two-way complete block (CB[2]) design – if one blocking factor
  + Other (you probably won’t use these):
    - Two-way Latin square (LS[2]) design – if two blocking factors
    - Split plot / Repeated measures (SP/RM) design – if units of two different sizes
* Show output from PROC PLAN that you will use in randomly assigning treatments to units. Do not supply a seed in your PROC PLAN code. Instead, look at the Log Window to see which seed SAS used. Write down this value below.  
    
  

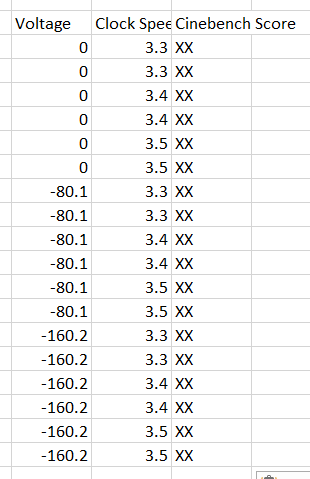
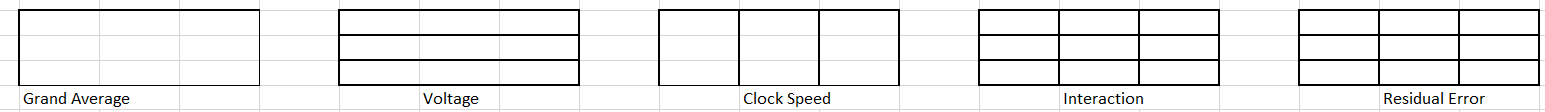
Seed = 164351913

* Explain how you will use this PROC PLAN output to randomly assign treatments to units.  
    
  Okay so using this chart I marked off each treatment with 2 numbers. And depending on where those numbers come up in the sequence in the chart above I will be testing that combination. Example, since the first number is 4 I will be test the CPU performance by using a voltage at 0 and CPU clock speed at 3.4

1-2 3-4 5-6

|  |  |  |
| --- | --- | --- |
| 0, 3.3 | 0, 3.4 | 0, 3.5 |
| -80.1, 3.3 (7-8) | -80.1, 3.4 (9-10) | -80.1, 3.5 (11-12) |
| -160.2, 3.3 (13-14) | -160.2, 3.4 (15-16) | -160.2,13.5 (17-18) |

Looking ahead to the data analysis

* Draw a table (without response values) for the data you plan to get from your experiment. This should be similar to those on Assignment 3. Each cell will correspond to a unit in the experiment; you should label the data table so it is clear which levels of the factor(s) correspond to each cell.   
    
  
* Based on this data table, draw the **factor diagram** for your design as in Assignment 3.   
    
    
  



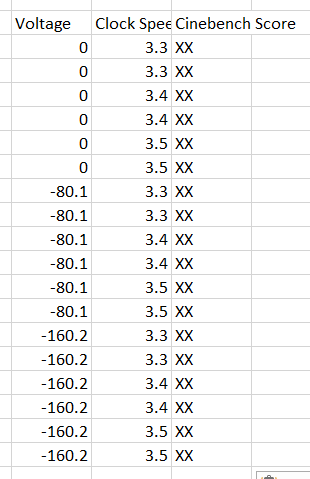
* What are the degrees of freedom associated with each factor? **You are required to have at least 6 degrees of freedom for residual error.** (If you don’t, you may need more replications).

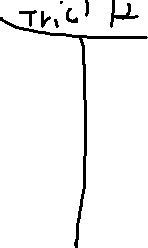


|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| * Factors | Outside factors | # Levels | Sum of df for outside factors | df |
| Grand avg | None | 1 | 0 | 1 |
| Clock Speed | Grand Average | 3 | 1 | 2 |
| Voltage | Grand average | 3 | 1 | 2 |
| Interaction | Voltage, Clock Speed, Grand average | 9 | 5 | 4 |
| Residual Error | Grand average, interaction, voltage, clock speed | 9 | 9 | 0 |

* Identify a **variable** that you will use when checking the **independence assumption** by plotting the residuals vs. this variable (and looking for patterns). Trial number is a common choice for this variable but there are other possibilities. (This variable should not appear in the tables above, but should be part of the Excel data table below.)

Trial number unless I wanted to use temperature of the CPU for this

* Give a data table that shows how the Excel file will look (other than missing response values) that is inputted into SAS for statistical analysis.  
    
  



* Write the code of PROC GLM you will use to obtain the ANOVA table.  
    
  Proc GLM data = cpu;

Class voltage clock\_speed;

Model score = voltage clock\_speed voltage\*clock\_speed ;

Run;

