RMHI/ARMP Assignment 2022

Hello everyone! This is the description for the assignment, which is due on Canvas on Monday May 2, 2022 before 8am. You'll need to submit a Word-knitted version of the completed R Markdown file found in this zip file, according to the following instructions:

1. Rename the document called assignment.Rmd as studentID-assignment.Rmd. (Replace studentID with your student ID number). This is your R Markdown file, where you'll be putting all your code and answers.

2. Replace "Your name, Student ID" in the header of the R Markdown file with your name and student ID. (Keep the quotes)

3. While we encourage collaboration in tutorials and learning in general, *you should not be collaborating with anybody AT ALL for this assignment.* **That means sharing code privately or publicly; even talking in the abstract about problems will effectively be collusion.** You should be completing it independently, with no help from any other person in any capacity. Of course, as always, you are free to use any of the resources from the class to help you, and you're also free to google or look anything up that you like (as long as you aren't asking anybody, including anybody on discussion boards, specific questions about this assignment).

4. Plagiarism check is enabled and you can check the similarity report on your submission. In previous years we have found people who tried to cheat, so please don't risk it! That said, understand that we will not be naively looking at the overall % figure: with this sort of assignment a certain amount overlap is inevitable, so don't worry if you get what looks like a high % score. Probably most people will. We will be using the plagiarism check for the parts of the assignment where we'd expect some variability, and to give a general sense of the overall gestalt.

5. Complete all of the problems below in the R Markdown document. *Do not change any of the arguments to the code chunks, like the names of the code chunks or where it says echo=FALSE or whatever.* If a problem asks you to display a tibble or variable so it shows up in the knitted version, **make sure that you do** as the marker cannot evaluate it without seeing it, and if they can't see it then they won't be able to award you points for it! Remember that to display a tibble (or any variable) you just type its name on a line of its own within the R chunk.

6. I've structured this so that, as much as possible, questions do not build on each other. That means that if, say, you can't get Q5 then you can still get Q6. Try to do all of them.

7. Go for partial credit! Many of these questions have some form of partial credit possible. What that means is that if it is asking for some R code, break down the problem into pieces. Even if you can only do some of the pieces, or do them part of the way, that will be worth something. [Note that there is no question-by-question rubric available because designing one would mean giving away the answers. In general we will give full credit for responses that correctly address all of the parts of the question.] Short answer questions (SAQs) are also worth partial credit and are generally asking for some thoughtful interpretation. If it is based on a previous graph or test you've done, if you did the first part wrong but discuss it well, you can still get most or all points for the SAQ part. If your code does not run but you want to include it for possible partial credit, just comment it out (using the # sign) so that it shows up in the knitted document. If you include a lot of commented-out code and some is correct and some isn't, we will not give you credit for the commented-out code; put the thing in there that you think is the closest to the correct answer, don't just include everything.

8. We are not overly worried about to what decimal place you round answers to and you will not lose credit for this unless you round so much that your answer is impossible to discern (e.g., don't round p-values to the nearest integer!). Similarly, you will not lose points for trivial presentation things like using parentheses instead of commas around statistical references, as long it's clear. That said, for those who want a guideline, I'll suggest that you follow APA format or round p-values to three decimal places, degrees of freedom to one, and test statistics and probabilities to two.

9. If the question is an SAQ, it specifies a word count. You need to either calculate it from the knitted document or type up your answer in Word[1] and then cut and paste it into the R Markdown file. (Please put your answer in between the word ANSWER and [Word count: XX]; needless to say, those two bits do not count towards your word count.) I know that's annoying; sorry. Anything else I thought of, like specifying a number of sentences or having no limit, was worse in terms of equity across students. The word counts I've specified in each question are designed to give you a guideline about the maximum amount of words you should need answer completely and correctly. So don't feel like you must use all of the words; if you can answer it fully with less, that's fine; in fact, the total word count for the solution set I wrote up is around 850. It's okay to go over the word limit for individual questions as long as the total word count for all of the questions is less than 1650 words (i.e., less than 1500+10%, with the standard penalty if it is 1500+10% or over. See the student manual for details on word count penalties).

10. There is no word count for code chunks. Word count only applies to the SAQs. **Remember to report** your total word count for the assignment as a whole at the top of the document. Your total word count is the sum of the word counts for all of the SAQs.

10. You'll be turning in the knitted output of your R Markdown file. We prefer that you knit to Word but if you can't get Word to knit then html is okay. In the worst case, you can turn in the completed Rmd file. **I highly, *highly* recommend that you knit as you go**: (a) knitting can identify problems in your code that you would have otherwise missed; and (b) you do not want to get close to the deadline and think you're done only to find that you're having troubles knitting. Save yourself the panic and knit often.

11. Similarly, you can turn in the assignment multiple times before the deadline, so I highly encourage you to turn it in even before it's perfectly polished. That will save you last-minute panic or computer issues. Also, take a screenshot for proof of having turned it in just in case you need it.

The story of LFB and Foxy

For this assignment, we're going to go back to meet up with LFB and Foxy and hear their story. As you'll recall, a few weeks ago they, Doggie, and Flopsy went on a mission to Otherland to steal some of their data. The mission was successful but LFB and Foxy went missing! In your assignment we get to see what happened to them.

* * *

LFB and Foxy are acting as lookout as Doggie and Flopsy enter the building. Standing on one side of the building, LFB is squinting through the darkness trying to see when she hears a rustle. Then another one, then another one, coming ever closer. Not wanting to raise the alarm prematurely, LFB holds still, but when she hears another rustle only meters from where she is, she whistles, giving the signal.

Foxy dashes around the building as quickly as she can, just in time to see the bushes near LFB part. She catches a hasty glimpse as three very large shapes -- bears? dogs? something else? -- jump out at LFB. Startled, LFB whistles as loud as she can, but it is abruptly cut off by one of the animals covering her mouth.

Forgetting her normal shyness, Foxy shouts "stop!" as loud as she can and charges at the two creatures. She growls at them, surprising herself, and they turn. She trembles: one of them is the largest bear she has ever seen, but it's too late to go back now. She growls again, and then the bear rushes at her and hits her and she is knocked unconscious.

---

[1] I know different software calculates word count in slightly different ways, so we are using Word as the standard.

After a short and frightening journey through the dark, LFB and Foxy are put into a small, bright room. LFB is relieved to see Foxy start to stir after a moment — she hasn't been hit that hard — and the two of them cuddle together in fear. The room they are in looks like a library, but it has very large chairs and books. LFB has to jump just to get down from the high sofa and reach the door handle. It is locked.

After about a half an hour of worry, the door opens and seven people come in. The first is an enormous bear, larger than anybody LFB has ever seen; she suspects that is the main person who subdued them. He is followed by an owl, a small unicorn with a rainbow mane, a hippo, and a cute penguin carrying a snake. The entire group is trailed by what looks, to their astonishment, to be a sentient guitar (that's right, a musical instrument that can walk and talk). *This is a very strange place*, thinks LFB.

Q1 [4% of total mark]

The seven strangers introduce themselves. The tibble do, which has been loaded for you in the R Markdown document, contains the information about them that LFB and Foxy have gleaned. Each row is one of the seven individuals, and the five columns are as follows:

name: the name of that person
species: the species of that person
height: the height of that person in centimeters
scariness: how scary that person seems (1 = not scary at all, 10 = extremely scary)
loudness: how loud that person seems (1 = not loud at all, 10 = extremely loud)

Take do and use the functions select(),filter(), and mutate() to create a new tibble that contains two new variables. One, called *personality*, is the mean of that animal's scariness and loudness. The other, called *morescary*, is TRUE if that animal's scariness is greater than its loudness. Then remove the *scariness* and *loudness* columns, as well as all animals less than 5cm tall. Assign this to a tibble called do_new and make sure that the tibble is visible when you knit your document. do_new should look like this:

name <chr>	species <chr>	height <dbl>	morescary <lgl>	personality <dbl>
super size	bear	164	TRUE	8.5
big wol	owl	32	TRUE	6.0
rainbow	unicorn	18	FALSE	3.5
hugo	hippo	22	TRUE	5.5
little blue	penguin	18	FALSE	4.0
kevin	guitar	25	FALSE	4.0

* * *

LFB and Foxy are both feeling a little calmer now that it appears nobody is going to try to kill them on sight. Still, the Others seem rather suspicious (not that that is surprising, really).

"What are your names?" the giant bear, Super Size, asks.

"LFB," says LFB, trembling.

"What kind of name is LFB?" asks Kevin, the guitar.

LFB bites her tongue and narrowly avoids asking what kind of guitar is named Kevin, and just says "It stands for Lovable Fluffy Bunny. My mum named me."

The unicorn shakes her tail and says "It's a lovely name. I like it," and glares at Kevin.

"How about hers?" the snake hisses, pointing at Foxy.

"*She* can answer for herself," Foxy says, bristling a bit. "My name is Foxy. Because I am a fox."

"Okay, okay," says Hugo the hippo. "That's fine. Are you okay? We didn't mean to hurt you when we captured you, we just didn't want to let you get away."

Mollified, Foxy nods. "Head hurts a bit but I'm okay."

"What are you doing here?" the giant owl interrupts.

Trading back and forth, LFB and Foxy tell everyone the whole story -- how they fear they are running out of food, and they wanted to see if the Others were stealing it (at this point LFB trembled a little bit more, and Foxy gave her a reassuring hug) or were having similar problems. As they get into the story, they can't help but noticing that most of their listeners seem stunned. The penguin whispers to the unicorn and the giant bear several times during the explanation. When they stop, there is a long silence.

"How do we know you're telling the truth?" the snake, Sissily, finally asks.

"I... don't know," LFB says. "We are, I swear."

After a pause, Little Blue raises her wing. "I have an idea," she says. "We can give them the HONOUR scale that I recently developed – it measures how honourable somebody is. That will tell us whether we should trust them."

"Oh dear," says Foxy. "I always overthink these things and mess them up."

"I can take it," LFB volunteers. "I mean, I don't know how it works, but I know that we're telling the truth and we're both honourable."

"Is this scale even normed appropriately though?" asks Kevin, glaring as well as a guitar can glare. He is obviously still very suspicious. "How do we know what 'good' is on it?"

Little Blue looks crestfallen. "Good point. I haven't had time to norm it yet. It has a good collection of questions with strong internal validity, and I know how to interpret them, but I haven't yet run it on lots of different people so I don't have a good sense of what the population behaviour on it would look like."

Everyone is crestfallen, but then Sissily ventures timidly, "Maybe we don't need that?"

"What do you mean?" asks Super Size.

"I mean, all we really need to do is compare LFB to somebody we know is honourable. If LFB's scores are around the same, we can conclude that she is probably honourable as well."

"That's a great idea," says Hugo. "How about Rainbow?"

All of the Others agree enthusiastically; apparently everybody there considers Rainbow a virtuous and truthful person. *I hope this works*, LFB thinks nervously. *I know I'm honourable, but am I as honourable as the most honourable of the Others? What if I mess this up? What if their test isn't very good after all?*

But there is no choice – she can't think of anything else that would persuade them better, so she nods and tries to look confident. She can't help but notice that Rainbow, too, looks nervous, but the colourful unicorn nods as well, and the two of them go to separate rooms to take the test.

It turns out that the HONOUR scale consists of 60 questions, each of which yields a score between 0 and 50 (higher = more honourable). All of the questions reflect different things so people are interested not only in comparing their performance overall, but also specific questions.

The tibble `dh`, which has been loaded for you in the R Markdown document, contains the coding of the answers to the questions LFB and Rainbow were asked. Each row corresponds to one question. The four columns are as follows:

question: the question number
lfb: LFB's answer to that question (min 0, max 50, higher=more honourable)
rainbow: Rainbow's answer to that question (min 0, max 50, higher=more honourable)
diff: The difference between LFB's and Rainbow's answer on that question (LFB-Rainbow)

Q2 [4% of total mark]

Use a pivot command to convert `dh` to a tibble called `dh_new` that looks like the one below and is the same as `dh2`, which has already been loaded for you. Make sure that the top rows of the tibble are visible when you knit your Markdown document.

```
> dh_new
# A tibble: 180 × 3
   question what      score
   <chr>    <chr>     <dbl>
 1 q1       lfb        49.7
 2 q1       rainbow    40.9
 3 q1       diff        8.8
 4 q2       lfb        40.3
 5 q2       rainbow    35.8
 6 q2       diff        4.5
 7 q3       lfb        29.7
 8 q3       rainbow    26.8
 9 q3       diff        2.9
10 q4       lfb        38
# … with 170 more rows
```

Q3 [5% of total mark]

Let's look at our data! Make a figure that shows the 1D distribution of *diff, lfb,* and *rainbow* using whichever geom seems appropriate to you. There should be three facets/panels in the single figure corresponding to each of the measures (one for *diff,* one for *lfb*, and one for *rainbow*), with the x axis corresponding to the score of each measure. Make sure the different measures have a different fill colour (using a palette of your choosing, as long as it's not the default) and are outlined in black. Title and label the axes appropriately (there is no need for a subtitle). Remove the legend if it is redundant, use a nice theme, and make sure that the scale of the x axis is different on each panel based on the range of scores (i.e., the scale shouldn't be fixed in the same way for all panels).

Q4 [6% of total mark]

Perform a statistical test to determine whether *lfb, diff,* and *rainbow* are each normally distributed. For each variable, report the statistical test, the statistical reference and interpret whether that means the variable is normal or not. [Suggested word count: 75]

Q5 [8% of total mark]

Use the appropriate statistical test to evaluate whether Rainbow and LFB's answers on the HONOUR test are significantly different from each other. Report the results. In your answer, don't worry about including descriptive statistics but do report which statistical test you used, the appropriate stats reference, the interpretation of this finding, and a measure of effect size and what it means. [Suggested word count: 100]

Q6 [5% of total mark]

Everyone is sitting around thinking about the results of the test when Hugo tentatively says, "I'm uneasy about something. How do we know the HONOUR scale actually captures who is honourable? What if it is measuring something else? How would we know?"

Little Blue answers, "Good question. Hmm. Well, I can definitely tell you that I've given it to the same people at different times, or with different people giving the test, and gotten similar scores."

(a) Is Hugo's question more about operationalization or measurement? Explain why with reference to the entities involved. (b) Does Little Blue's answer address Hugo's concern? If your answer is yes, explain why, making a clear link between the two, and give an example of something else that would address his concern equally well (note that the something can be hypothetical, it need not actually exist). If your answer is no, explain why not with clear reference to what Little Blue is talking about and what Hugo means, and give an example of something that would do a better job at easing Hugo's worries (note that the something can be hypothetical, it need not actually exist). [Suggested word count: 150]

* * *

The Others confer a bit and realise that regardless of the results of the HONOUR scale, over the course of working and talking with LFB and Foxy they have realized that the two are at least reasonably trustworthy. Following a long, whispered conference amongst each other, Rainbow the unicorn steps forward and unties them.

"Sorry for our suspicion. We've been having food problems ourselves," she confides quietly. "We haven't known what to do about it, and are pretty worried."

"Maybe we could help?" LFB offers. "I mean, I don't know much, but perhaps if we compare problems we'll be able to figure out what's going on. We can tell you what we know about our situation too."

Foxy nods and shares the survey data we saw in previous weeks. The Others share their food data that you went over in the tutorials, and everyone agrees that there is a problem.

"The thing is," Super Size observes (everyone is now very companionable and speaking frankly), "I fear that this is having a lot of bad indirect effects on everything else. People are more irritable and fighting more, they're sick more often, and things like that."

"Do you have any data about that?" LFB asks, curious.

There is a long silence, and then Sissily volunteers: "Well, we could look at health records."

"What do you mean?" asks Kevin. "I thought those were private."

"They are," Rainbow agrees. "But we have deidentified data that we can look at in the aggregate. For instance, ten years ago the government set a wellness standard – which I know they achieved

then – and we could see if we're still achieving it now. We could see if the number of people having food-related health problems is below that standard."

To be precise, we can consider three categories of health problems:

Low: if the health problem arose because of not enough food (e.g., starvation)
Nutrition: if the health problem arose because of enough food but a poor diet (e.g., malnutrition)
Nonfood: if the health problem arose because of something unrelated to food

When the government set the standard ten years ago, the aim was for 10% of issues to be related to not enough food, 10% to be due to malnutrition, and the other 80% to be due to something else. This standard was achieved then. The question is: is it being achieved now, or are there more problems now due to either not enough food or poor nutrition?

Everyone is enthusiastic about exploring this more, and the next day — after they find some data, have a long sleep, and share a companionable dinner with their new friends — they all gather around. The data is in the tibble called `dp`, which has already been loaded for you. It contains the following columns:

id: a code indicating a single person at a single doctor visit
problem: the problem being dealt with at that doctor visit (low, nutrition, or nonfood)
improved: TRUE if the patient got better, FALSE if they didn't

Q7 [8% of total mark]

Use the appropriate statistical test to evaluate whether the distribution of health problems is significantly different from the standard set by the government. Report on the results. In your answer, include descriptive statistics, a report on which statistical test you used, the appropriate stats reference, and the interpretation of this result. Do not worry about calculating or reporting effect size. [Suggested word count: 130]

Q8 [8% of total mark]

Suppose that out of 150 doctor visits, 134 of them were for reasons unrelated to food and 16 were for food-related reasons. Considering only these two possible categories of outcomes (food-related and non-food-related), calculate the probability of seeing 134 or more non-food visits assuming that the underlying true proportion of non-food-related problems in the population is 0.8. There are two separate ways to calculate this, with two different functions; a full credit answer will calculate it in both ways. Report the probability in the blank space provided on the answer sheet. Explain what each of your calculations is doing (as if you were teaching someone else about them). [Suggested word count: 125]

* * *

"Not to be a pain," says Foxy after a while, "but even though it's nice to have this data, it doesn't tell us anything about *why* we're seeing these patterns."

"It's very hard to infer causation from most data," says Kevin chidingly.

Foxy is too polite of a person to roll her eyes, but LFB can tell she wants to. "I know," she says instead. "We can't infer it for sure but if we could look at patterns of change over time, and see which kinds of measures change and which don't, that can indicate something."

Rainbow nods in support. "Yeah. Like, if people's health was getting worse over the same time the amount of food went down, it at least *suggests* that those things might be related."

"What if there was some other variable causing both?" asks Hugo. "Like maybe people's health is getting worse *and* there is less food because people are getting poorer and so can't afford it."

"Or maybe there's some disease causing health to drop, which makes people not feel well enough to harvest crops, and that is why the food is going down," chimes in Super Size.

"We can't tell for sure," LFB repeats again. "But these hypotheses all imply different patterns and relationships, and at least we can look to see what patterns there *are*."

Everyone nods again, but the mood is down. The task seems impossibly hard.

This time Kevin breaks the silence. "Little Blue, do you have data looking at health over time?"

Little Blue thinks, and then nods finally. "It's not as big of a dataset as some of the others, but myself and a bunch of my friends have been using an app that track different measures about our life. We have data from the last three years. We could look at that."

She brings it out and everybody clusters around and looks at it. "There is a sentient string in Otherland?" LFB asks incredulously.

Kevin looks up, miffed. "That's my best friend, Kevin Clark," he says. "What, do you think a string can't be intelligent? Or a guitar?"

"No, no, just curious," LFB backpedals hastily. "All good."

Rainbow whispers to her, "We don't understand it either. Just go with it."

Super Size clears his enormous throat. "Ahem. So now you have a sense of our dataset. That's reasonably representative of Otherland, I would say."

Sissily nods. "Yes. Mostly birds, bears, and bunnies, with a bunch of other things too."

"This is super fascinating," Foxy interrupts, "but let's have a close look!"

The data is in the tibble called dd, which has been loaded for you. It has the following columns:

name: the name of each person
species: the species of the person
size: the size of the person (small, medium, large, enormous)
time: when the data was collected. There are three time points separated by a year each (t1, t2, and t3). Each person contributes three rows to the dataset, one for each time point. The most recent time period, t3, occurred a few months ago.
health: that person's overall health rating on a scale of 0-100 where higher equals better
income: that person's income during that time period (higher equals better)

Q9 [7% of total mark]

The tibble dd_sum2, which has been loaded for you, was created using the group_by() and summarise() functions to calculate the mean, median, and standard deviation for *health* for each size at each time point. Using the same functions, make your own tibble called called dd_sum that looks exactly like dd_sum2. Ensure that dd_sum appears in your knitted document.

Q10 [10% of total mark]

Create a bar plot with the following specifications. There should be four panels/facets, each corresponding to one size of animal. Each panel should contain three bars, one for each time point (on the x axis) with the y axis showing health. Each bar should be outlined in black and have error bars corresponding to standard deviation. The colour of each bar should be semi-transparent and be different for each time. Individual data points should follow the same colour scheme as the corresponding bar and both bar and individual data points should be visible. Title and label the axes appropriately (there is no need for a subtitle). Remove the legend if it is redundant. What does this figure suggest about how health is changing over time for people of different sizes? (We know what you observe may not be significant; this question is just about describing the trends, which other tests can tell us if they are significant or not). [Suggested word count: 25]
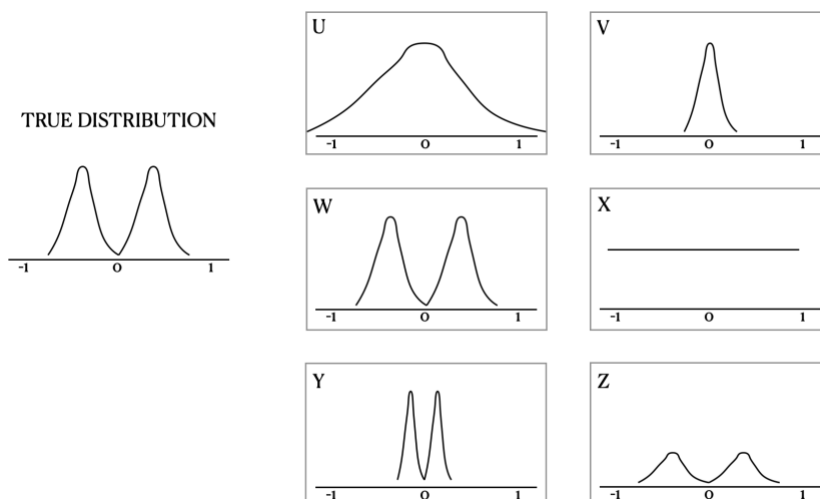
Q11 [8% of total mark]

Make a figure of your own using any of the datasets, with the goal of learning something new about the data that hasn't been shown by the previous plots. Requirements: (a) it needs to involve a geom other than one of the ones you used before; (b) it needs an informative title and axis label; (c) it should involve more than one facet; (d) it should be clear, with aesthetic choices that add to its clarity rather than detract from it; (e) you should explain what the graph suggests about the data. In your explanation be sure to describe the variables on each axis as well as what the pattern is and what it suggests about what is going on for our friends. Feel free to go beyond these requirements if you like (e.g., you can use more than one geom, subtitles, etc) but it is not necessary to get full marks. [Suggested word count: 160]

Q12 [7% of total mark]

Suppose our friends manage to measure some variable that whose true population distribution is shown in the diagram below on the left. Because of a wealthy benefactor, they are able to run 5000 experiments; in each of these, they sample independently from the true population distribution. In each of these experiments, they measure the mean of their sample.

Consider the six panels U through Z. (a) Suppose that each experiment has a sample size of 200. Give the letter corresponding to the panel that most accurately captures what you would expect the sampling distribution of the mean to look like. (b) Suppose instead that each experiment has a sample size of 1. Give the letter corresponding to the panel that most accurately captures what you would expect the sampling distribution of the mean to look like. (c) Explain each of your answers in (a) and (b), making reference to the central limit theorem and the definition of standard error. [Suggested word count: 175]

Q13 [8% of total mark]

Foxy ran a t-test that yielded a particular *t* statistic and p-value. (a) Suppose Foxy ran the same test on different data, this time with twice the sample size, and got the same *t* statistic. All else being equal, would the *p*-value be higher, lower, or the same as it was the first time? (b) Suppose Foxy ran the same test on different data, with the same sample size, and got the same *t* statistic except that it was negative instead of positive. All else being equal, would the *p*-value be higher, lower, or the same as it was the first time? For both (a) and (b), explain your answer, each time making reference to the *t* distribution, degrees of freedom, and the p-value.
[Suggested word count: 190]

Q14 [5% of total mark]

"I've been confused about this for a while but I think I've got it!" Super Size suddenly says with satisfaction. "A p-value of 0.4 means that the probability of the null being true given your data is 40%, and the probability of the alternative being true is 60%. We try to minimize Type I error by setting alpha equal to *p* only when *p* is less than 0.05." He looks around anxiously. "Is that right?"

Can you answer Super Size? In your answer, state clearly whether he is correct or not and explain why. (Be sure to address all of the parts of his statement) In your answer, describe clearly how the p-value relates to the null and alternative hypotheses, as well as alpha. [Suggested word count: 170]

Q15 [5% of total mark]

"I don't like statistical tests," Kevin says grumpily. "I think we should just always have very large samples; that way we can ensure that Type 1 and Type 2 error are zero, no matter what phenomena we are studying and no matter what the true effect size is."

Imagine for now that we live in a world with lots of resources and it is possible to always have very high sample size (though not infinite, nor would we be sampling the entire population). Would this indeed permit both kinds of error to be zero or extremely close to zero? Why or why not? A full credit answer will discuss the relationship between alpha, beta, effect size, and sample size; there is no need to include any equations or calculations, but your answer should give an intuitive sense of how these four factors are related and why. [Suggested word count: 200]

Q16 [2% of total mark]

This one is a freebie - any answer is fine as long as you answer it. What would you like to see happen in the Bunnyland story? (This can be anything, from a huge plot point to a tiny character development to a neat scene or anything in between). Use however many words you like; this doesn't contribute to your word count!