Quantitative Understanding in Biology I
2017 Midterm Exam

November 2nd, 2017

Instructions

After the exam, these pages will be separated by question for grading purposes. To ensure that your complete response to each question is considered when grading, please be sure to do the following:

- Hand in all pages that you were given, in order! Even the blank ones at the end, and even if you don't use them (this way we'll know that you didn't use them, and that we didn't lose them).
- Write your name at the top of each side of each page (the exams will be scanned before grading, so we need each page to have your name on it).
- If you do use any of the extra pages, please only respond to one question on a given page, and indicate which question you're addressing at the top of that page.

Question 1

The mean height of 17 year-old females is well established at 163.2 cm, with SD = 9.8 cm. You are planning a study to ascertain if the heights of female high school seniors from the remote Iron Islands of Westeros differs substantially (by more than 2 cm) from this mean.

Roughly how many female students' heights would you plan to measure if you wanted your scientific expedition to have a 50% chance of detecting a difference in the mean of more than 2 cm at 95% confidence? Explain your reasoning.

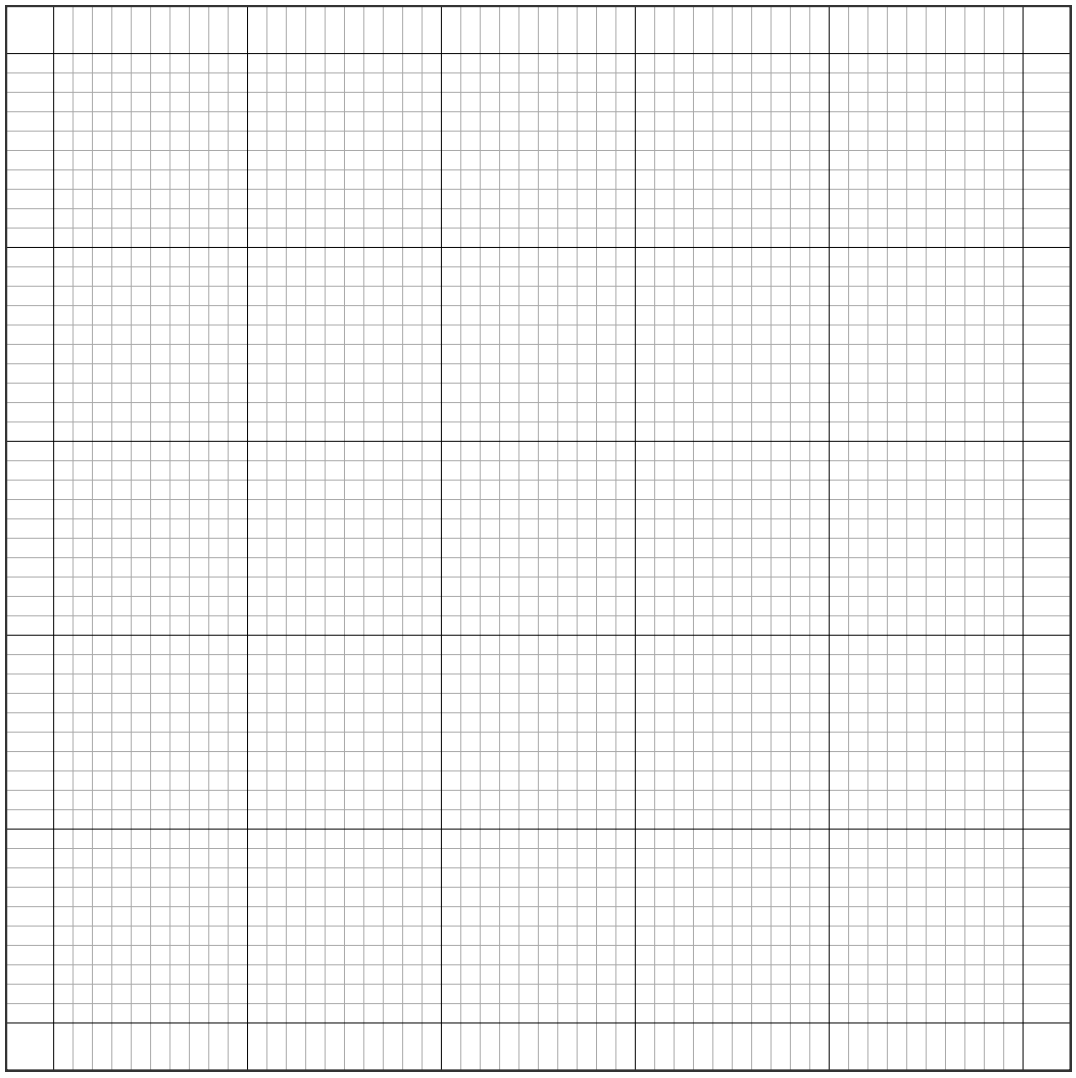
Question 2

You are working with a new mutagenic compound that is hypothesized to cause point mutations to DNA sequences at random loci on a chromasome (i.e., mutations are expected to be uniformly distributed across the genome). In a preliminary high-throughput sequencing study, mutations due to this compound are found at the following locations on Chromosome 1 (which is 249,250,621 base pairs long) in HeLa cells.

```
##  [1]  49783223 121580771 125604024 246158851 123316932 125067377
##  [7] 131290846 128617702  34451437 120655588  48473720 140255422
## [13] 122920300 120818148  25055135   7145635 122341358 117438941
## [19] 232201595 120266012 164628560 124080941  86214440 174587959
## [25]  19917322
```

Sketch an appropriate QQ plot that can be used to test the hypothesis that mutations are uniformly distributed across Chromosome 1.

```
##  [1]  49783223 121580771 125604024 246158851 123316932 125067377
##  [7] 131290846 128617702  34451437 120655588  48473720 140255422
## [13] 122920300 120818148  25055135   7145635 122341358 117438941
## [19] 232201595 120266012 164628560 124080941  86214440 174587959
## [25]  19917322
```



Based on your sketch, do you think the hypothesis that mutations are uniformly distributed across Chromosome 1 is plausible? Explain your reasoning.

What formal statistical test could you use to further evaluate this hypothesis?

Question 3

You are rotating in a lab that is searching for compounds that inhibit the motility of infectious bacteria. A high-throughput screen identified three candidate componds, and a followup assay was performed that ultimately uses a t-test (with a pre-determined $\alpha = 0.05$) to determine statistical significance. The p-values reported by the `t.test()` function in R for each of the three compounds are:

	Compound A	Compound B	Compound C
p-value	0.032	0.306	0.611

A postdoc presenting these results in lab meeting concluded:

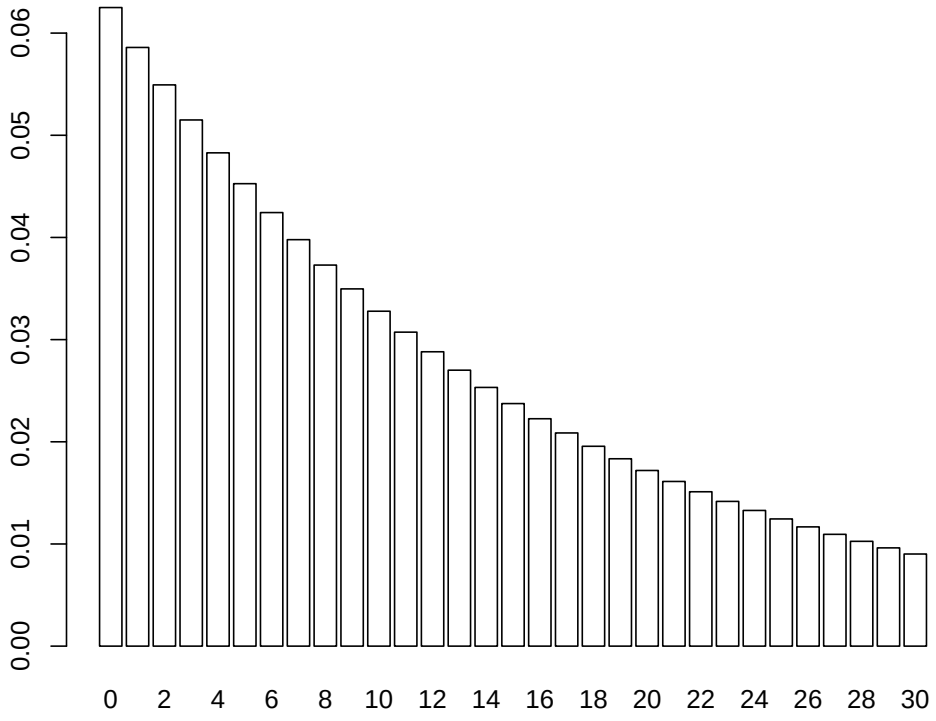
- Compound A has been proven to significantly reduce motility
- While the results for Compound B are not statistically significant, it is likely to have twice the efficacy of Compound C because the p-value for Compound C is twice that of Compound B. The postdoc plans to collect data from additional replicates for Compound B in order to formally establish significance for a manuscript that he is preparing to submit.

Do you agree with these conclusions? Why or why not? What advice would you offer, and what followup questions would you ask?

Question 4

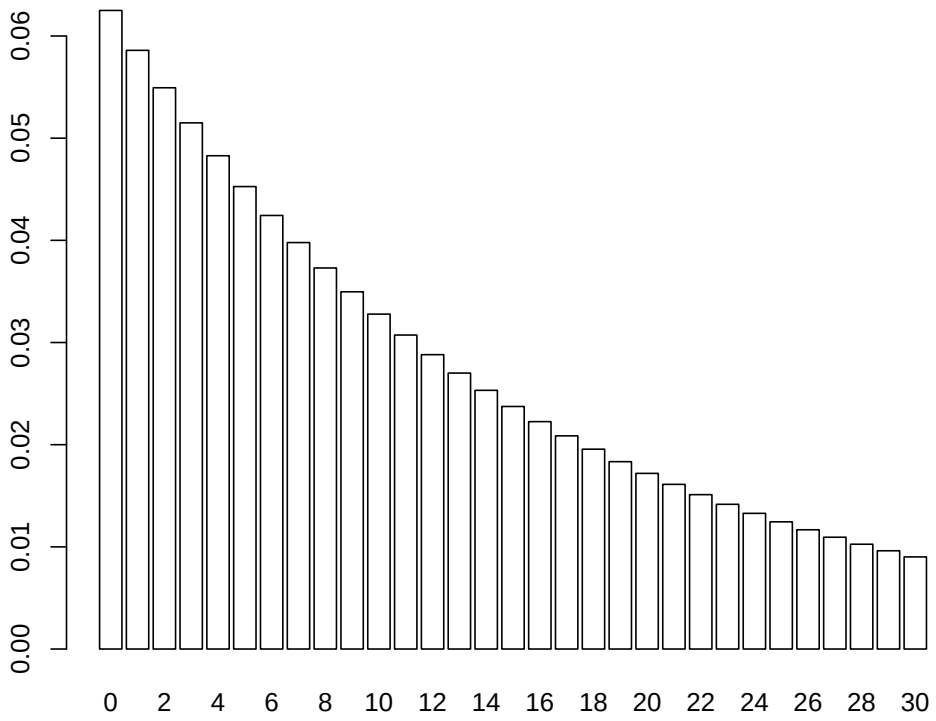
The plot below shows the probability density function for the geometric distribution where the underlying probability of success is $0.25 \cdot 0.25 = 0.0625$. It may be used as a model for the number of animals that need to be bred that simultaneously express two recessive traits when parents are heterozygotic in both genes (and classical Mendelian inheritance is assumed).

```
t <- 0:30
barplot(dgeom(t, prob = 0.25 * 0.25), names.arg = t, col = 'white')
```



Estimate of the value of the expression `pgeom(6, prob = 0.25 * 0.25)`, and annotate the plot above to show graphically how you obtained your answer.

Estimate the value of the expression `qgeom(0.5, prob = 0.25 * 0.25)`, and annotate the plot below to show graphically how you obtained your answer.
The plot is replicated below.



[extra space if you need it]

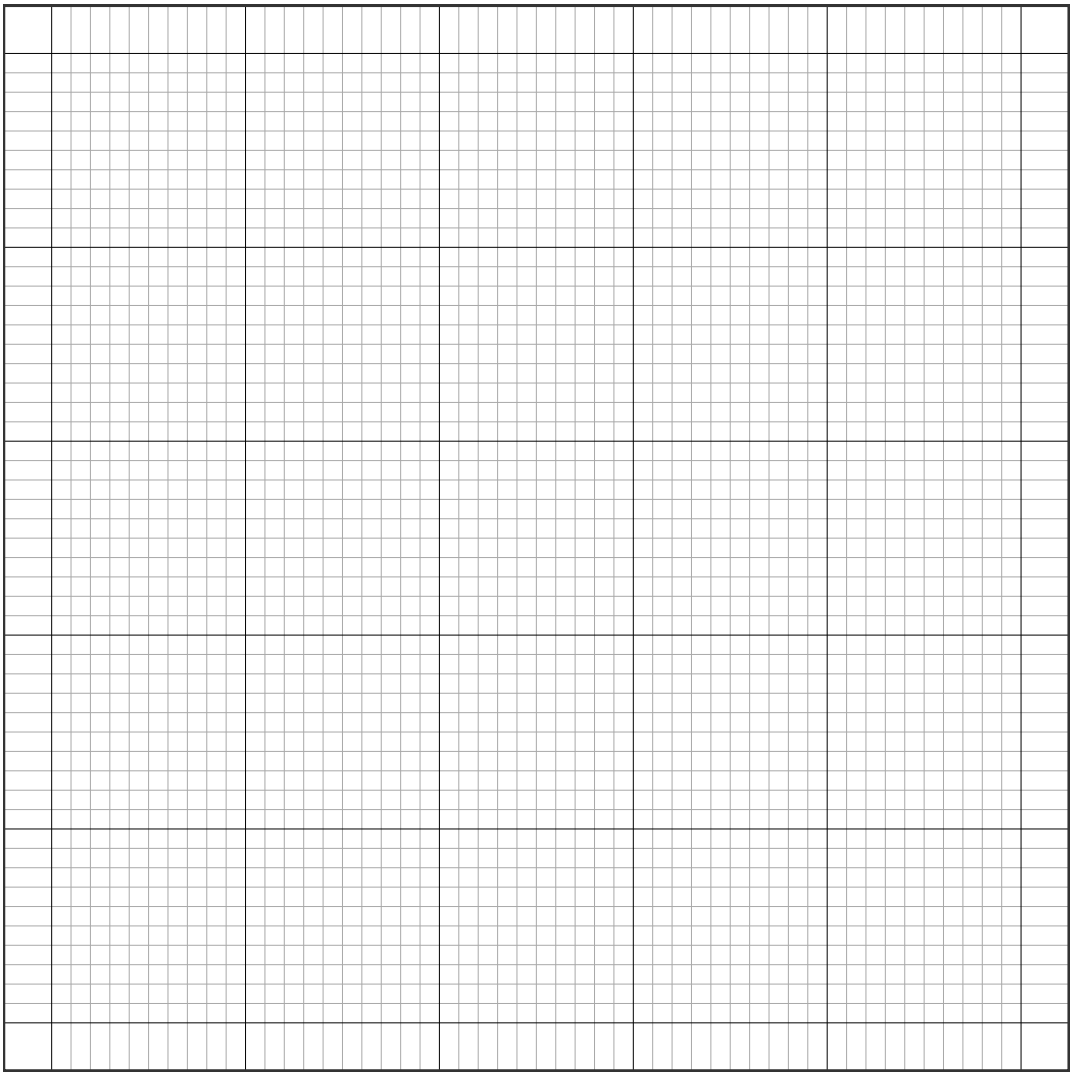
[extra space if you need it]

[extra space if you need it]

Name:

[extra space if you need it]

```
##  [1]  49783223 121580771 125604024 246158851 123316932 125067377
##  [7] 131290846 128617702  34451437 120655588  48473720 140255422
## [13] 122920300 120818148  25055135   7145635 122341358 117438941
## [19] 232201595 120266012 164628560 124080941  86214440 174587959
## [25]  19917322
```



[extra space if you need it]