Quantitative Understanding in Biology I
2018 Midterm Exam

November 1st, 2018

Instructions

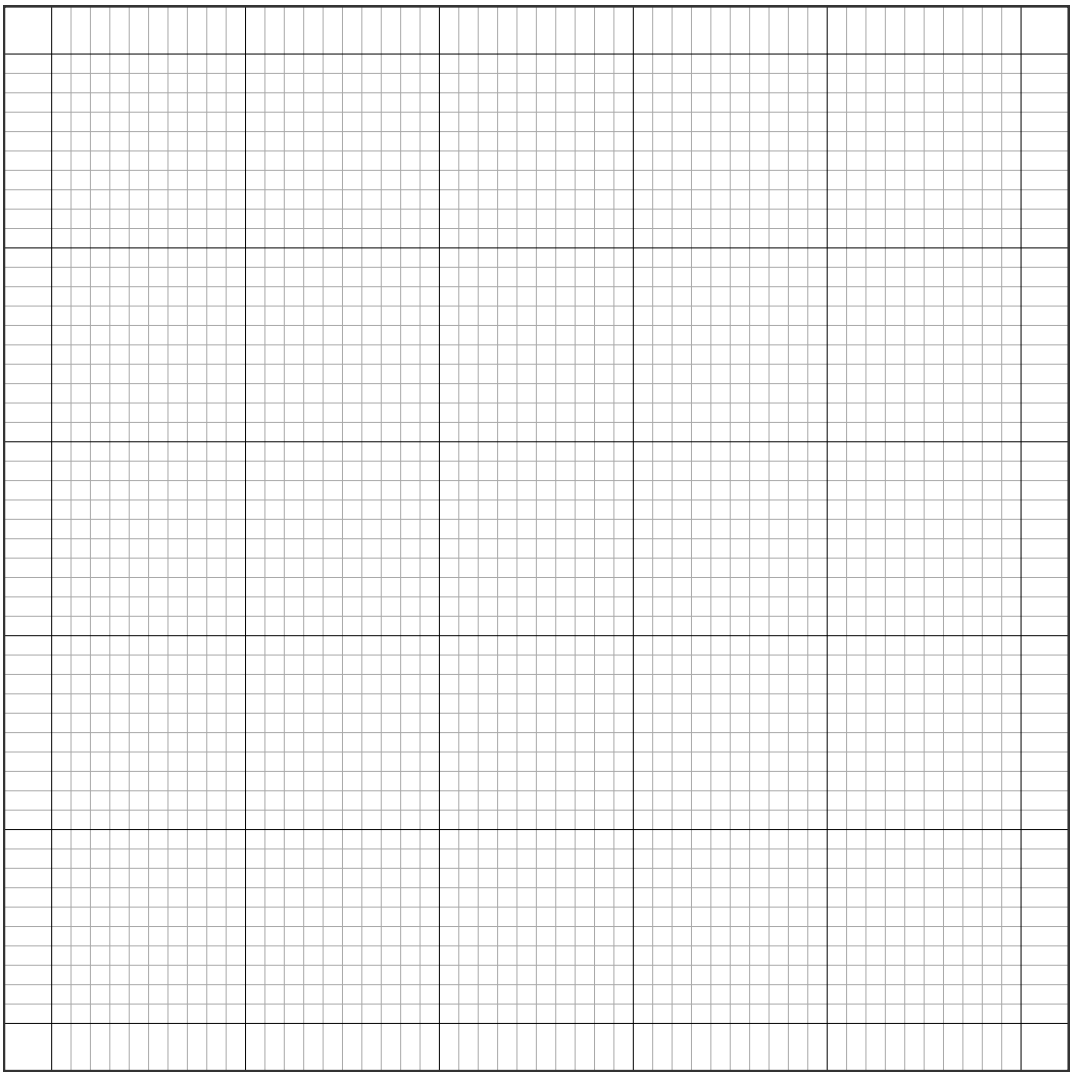
After the exam, these pages will be separated by question for grading purposes. To ensure that your complete response to each question is considered when grading, please be sure to do the following:

- 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.

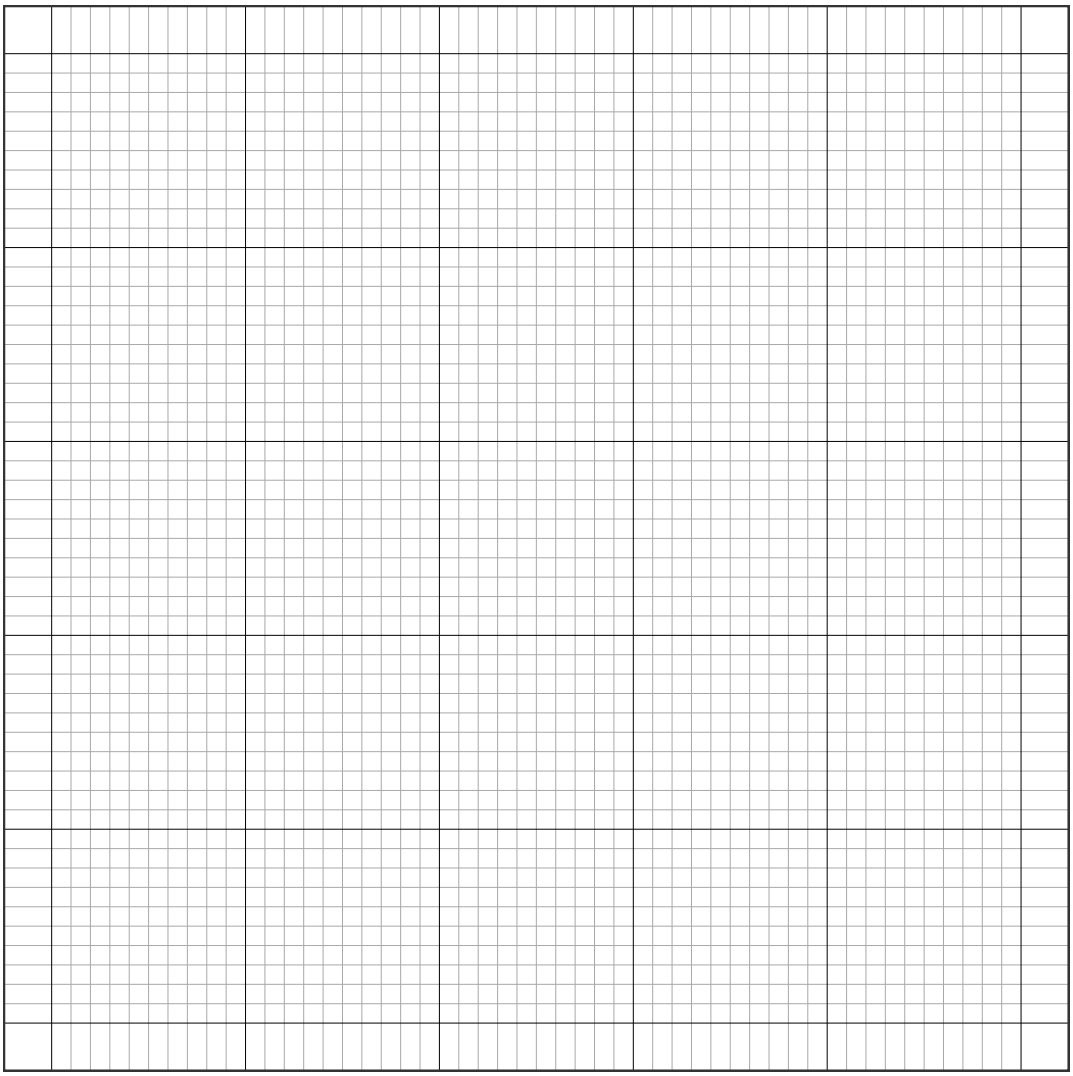
Question 1

For each of the distributions listed below, (i) state whether the distribution is discrete or continuous, (ii) state whether the distribution is symmetric, and (iii) sketch the distribution, being as quantitative as you can; i.e., indicate numerical values where you can (hint: working with fractions may be easier than their decimal equivalents).

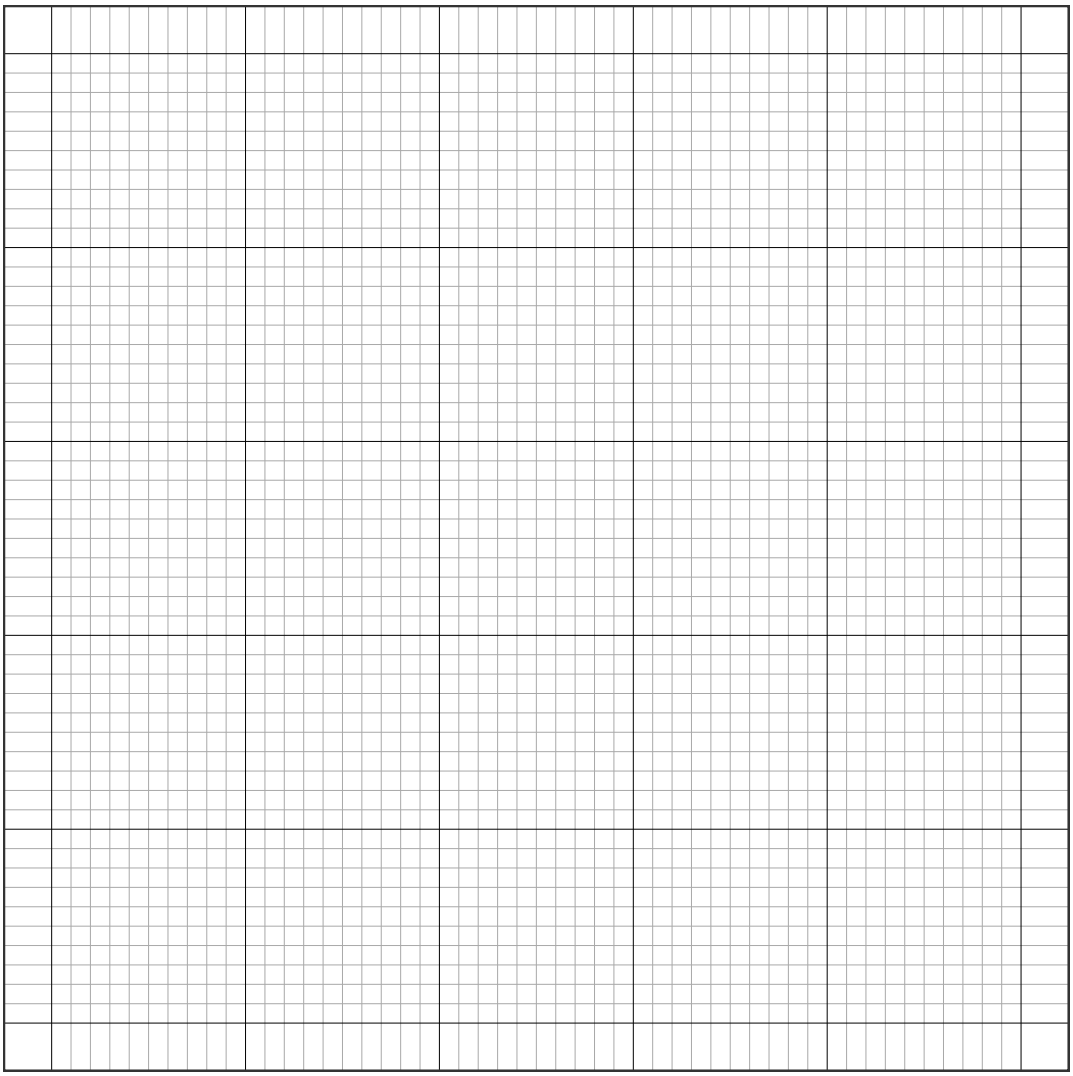
(a) **A uniform distribution with a range from 0 to 10.**



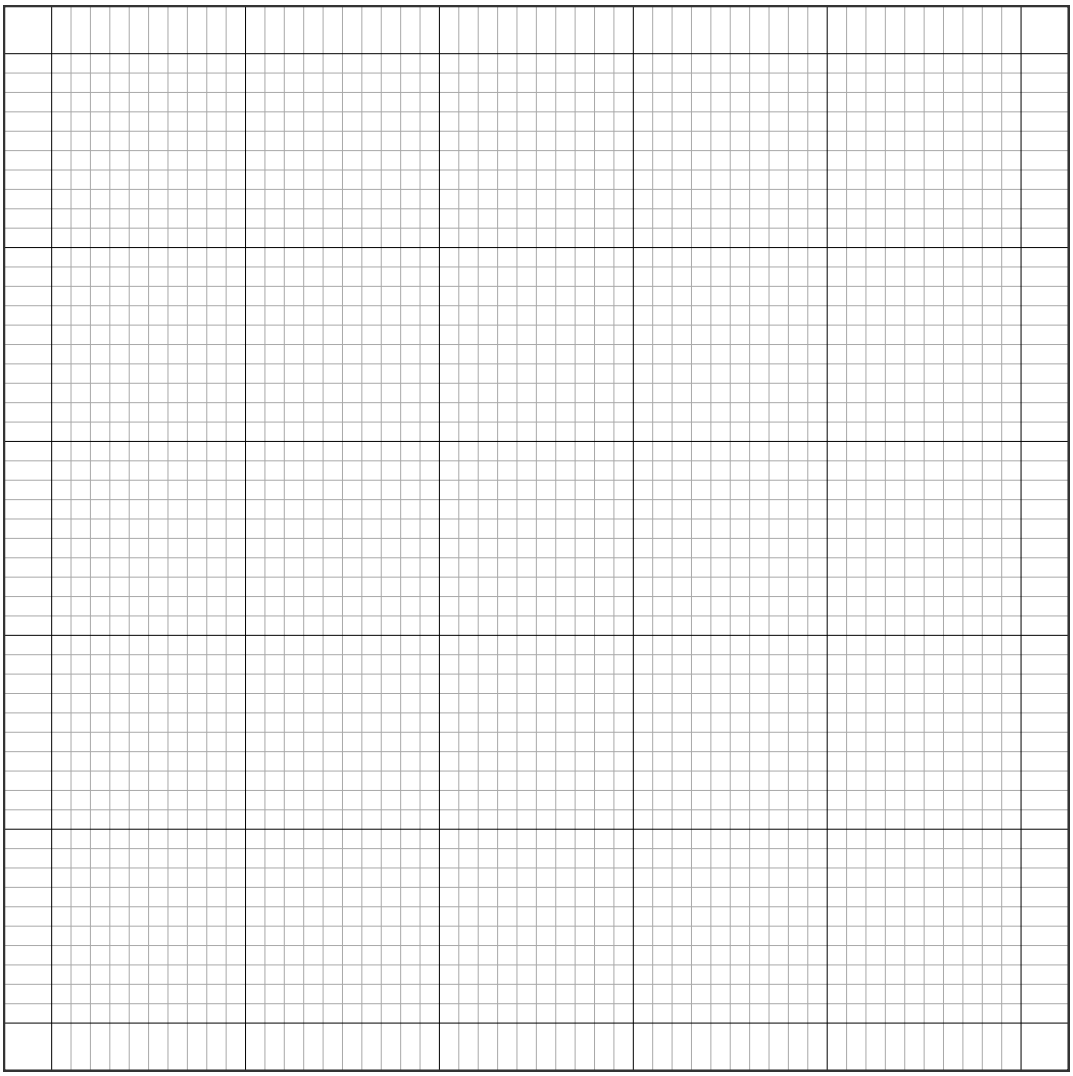
(b) **A Poisson distribution with a lambda parameter of four.**



(c) **A binomial distribution with n = 2 and p = 0.2.**



(d) **A geometric distribution with p = 0.5.**



Question 2

In your last lab meeting, a high-school intern has presented data that he claims demonstrates that the expression level of synaptotagmin in C. elegans cells is altered when the worms are repeatedly subjected to a compilation of Beethoven's sonatas. He established this by using 64 worms in each of the control group (which was exposed to white noise of an equal volume and duration) and the exposed group, and found a p-value of 0.035, which was lower than the pre-established alpha value of 0.05 that the lab traditionally uses.

(a) **One of your lab mates claims that this is not all that interesting because the relatively high p-value indicates that the effect size is small. How would you respond to this line of reasoning?**

(b) **When you ask if the intern could share the 95% confidence interval for the change in synaptotagmin expression, he says that he did compute it, but because the CI included zero, he didn't use that statistical method in his final analysis, and instead stuck with p-values instead. How would you respond to this statement?**

Question 3

You are being asked to investigate a disease outbreak at a remote, isolated campus. Some students suspect that proximity to the campus cafeteria is a factor, while the administration are more skeptical. The distance from each victim's dorm to the cafeteria (in meters) is given in the list below.

```
##  [1] 1830 1874  572 1661 1283 1038 1473 1567 1828 1853 1729 1860 1967
## [14] 1628 1731 1970 1989 1559 1737 1780
```

(a) **Sketch an appropriate QQ plot that can be used to test the hypothesis that incidents vary uniformly with distance from the victim's dorm to the cafeteria.**

(b) **What conclusions can you draw from your sketch? Explain your reasoning.**

A similar outbreak occurred several years ago, and similar distance data (this time, sorted, mercifully) for that outbreak is given here:

```
##  [1]  165  277  776  780  894 1028 1370 1475 1502 1504 1519 1519 1586
## [14] 1604 1622 1631 1667 1672 1673 1690 1699 1716 1718 1757 1806 1808
## [27] 1809 1811 1820 1839 1874 1892 1893 1916 1944 1953 1978 1979 1985
## [40] 1987
```

(c) **Sketch an appropriate QQ plot that can be used to test the hypothesis that distances from the cafeteria follow the same distribution in both outbreaks.**

(d) **What conclusions can you draw from this sketch? Again, explain your reasoning.**

Question 4

Your lab is screening for novel compounds that slow down the cell cycle in yeast cells. The lab initially targeted a subset of the library comprising 25,000 compounds. The post-doc who designed the experiment explains that, in order to ensure that only high quality hits are generated, she chose a p-value cutoff of 0.01. While she got 227 hits from the screening library after computing an independent p-value for each compound, when she showed the results to the lab head, she was told that she needed to learn something about multiple hypothesis corrections, and to seek your help since you're taking the qBio class.

The post-doc says that she's looked this up, and tried a Bonferroni correction, because it is the most conservative, and she wanted only good hits; however, when applying that correction, she found no significant results, so she's now asking if you could help with an FDR (False Discovery Rate) calculation.

(a) **How many hits do you expect to get if you analyze the same data using the Benjamini-Hochberg correction with a target FDR of 1%. Explain your reasoning.**

(b) **Your lab mate says that, based on her initial results, she thinks that it is very likely that there are over 200 active compounds in the library, because if each one was tested separately, that is the result she would have gotten. Do you agree with this line of reasoning? Can you say anything about the number of active compounds in the library based on the information you have?**

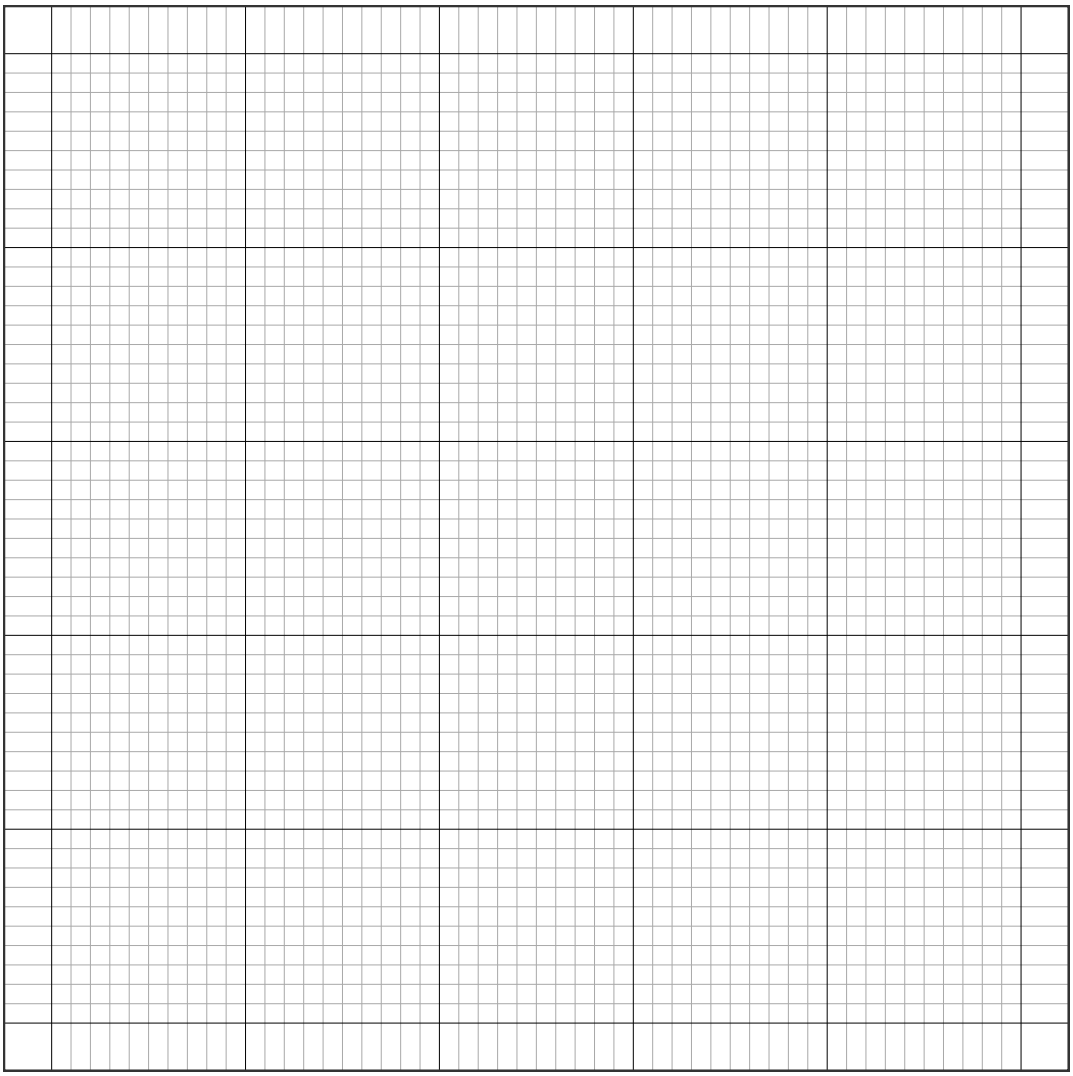
(c) **What recommendations would you make to further the analysis of the collected data, and for planning future screens?**

[extra space if you need it]

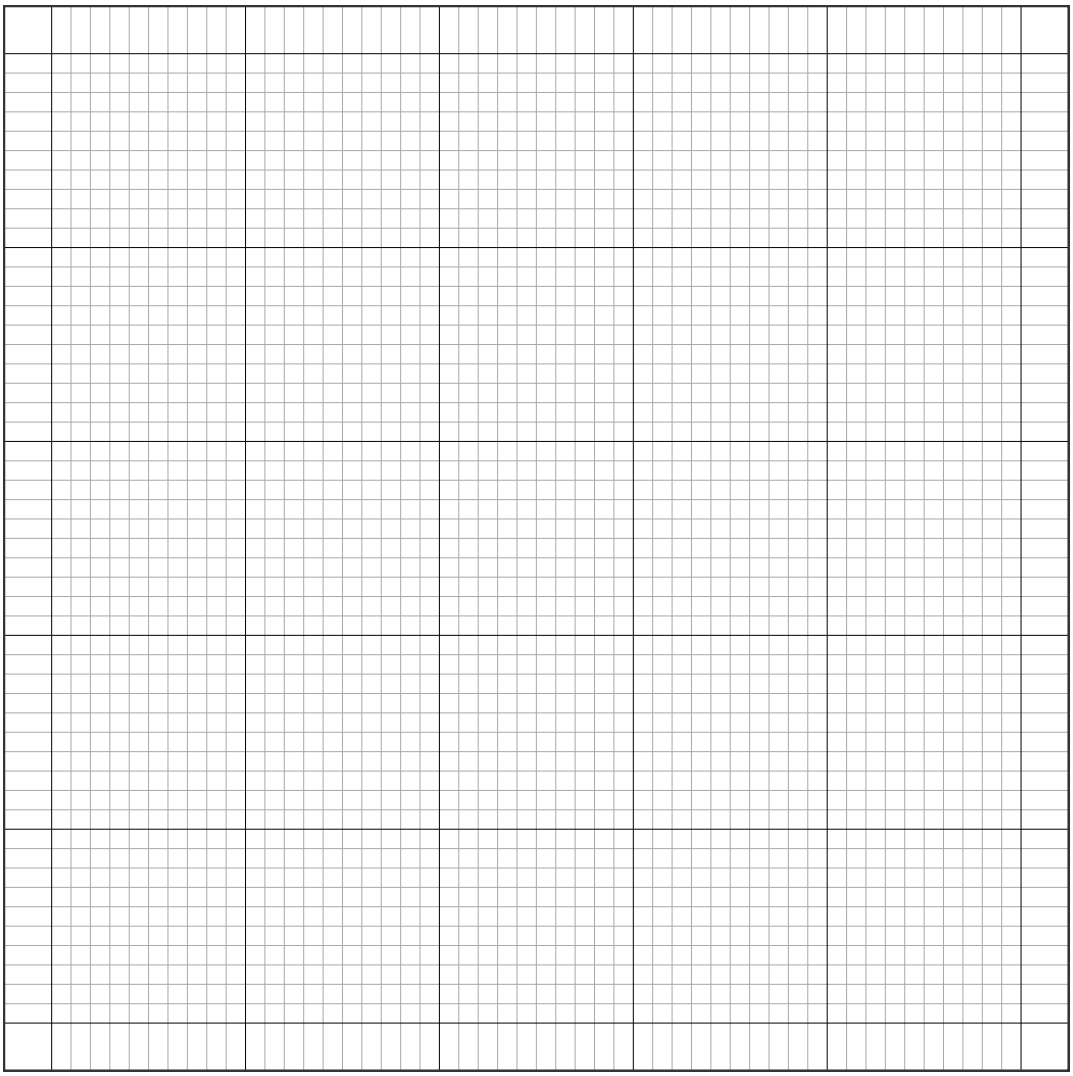
[extra space if you need it]

[extra space if you need it]

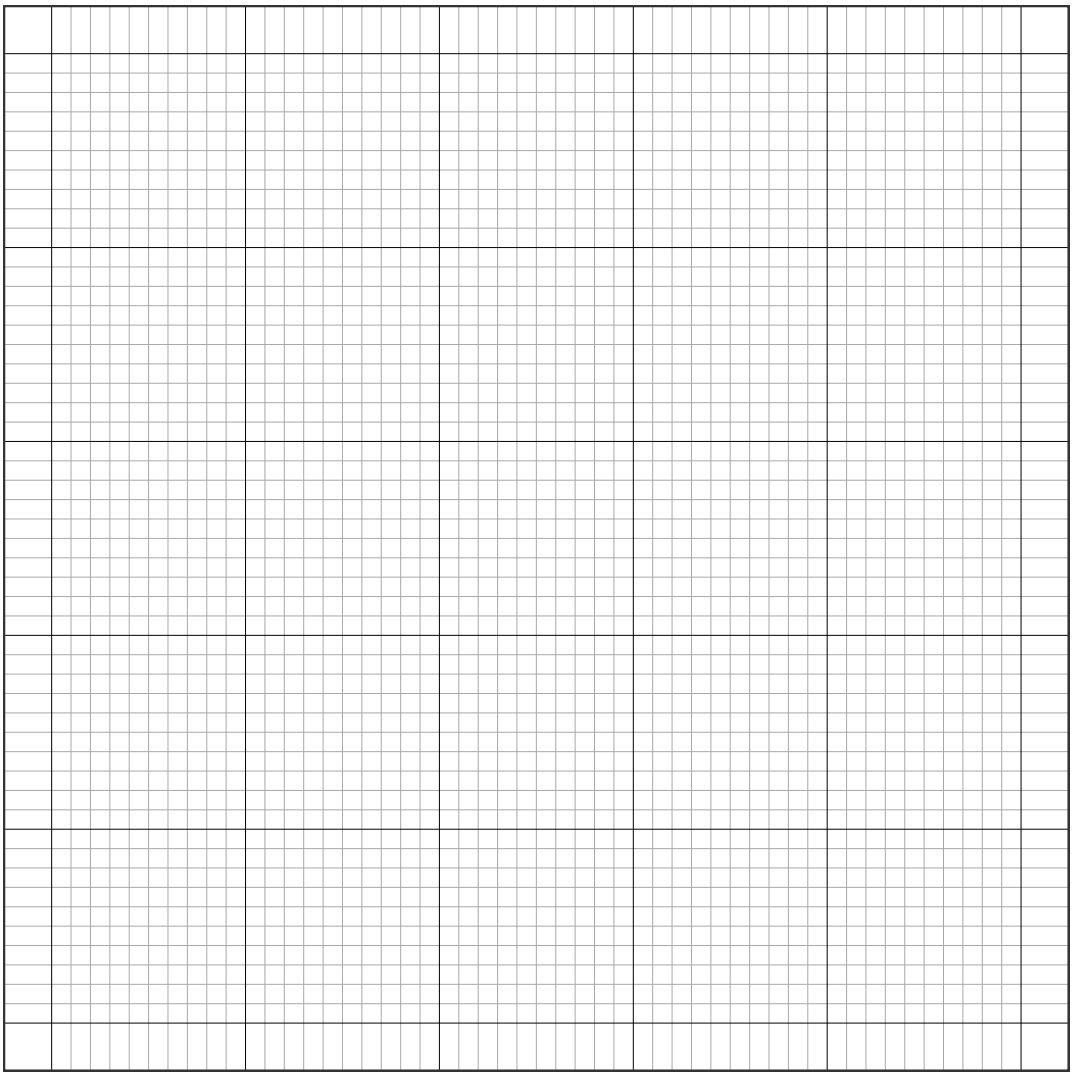
[extra space if you need it]



[extra space if you need it]



[extra space if you need it]



[extra space if you need it]