Here is a brief review of the material to be covered on our first exam. The exam covers chapters 1,2,3,4 and 5 in the Kitchen's book. At the end of each chapter is a section titled *Summary and Review*. This is a good place to look to practice questions for the exam. Please note that many of the odd-numbered questions are answered briefly in the back of the book.

Chapter 1 The material in chapter 1 is mostly definitions and concepts. The key things to be familiar with are the terms: variables, datasets, populations, census, surveys, experimental units, sampling from a population.

Chapter 2 This chapter studies univariate data. That is one variable at a time. First, it comes in 2 flavors: categorical and numeric. Furthermore, numeric data can be discrete or continuous. (Of course by lumping you can turn numeric data into categorical.) This chapter is concerned with summarizing such data with summary statistics and graphs. In particular you should know how to make all of the following for a data set: bargraph, boxplot, histogram, piechart, stem-and-leaf diagram, find the mean, median, trimmed mean, midrange, standard deviation, IQR, table of frequencies etc. As well, we talked about skewed (left or right), long or short tailed and modes of a histogram.

Chapter 3 This chapter studies bivariate data. It shows how to summarize graphically and numerically. The tools depend on the type of data: categorical vs. categorical or cat. vs. numeric or numeric vs. numeric. The tools include: n-way contingency tables, side-by-side barplots, scatterplots. For the latter, *if there is a linear relationship* we summarized with a Pearson correlation coefficient and a regression line which can be used for prediction. Know what the predictor and response variables are. For monotone relationships we can use the Spearman correlation coefficient. You should know what large or small values of R^2 mean and what the regression line does and how it predicts values.

Chapter 4 We learned some elementary rules of probability. The most important concepts are independence and the notion of a distribution. But we did more. We learned how to do some relatively easy calculations, we learned about the mean and standard deviation of a distribution, we learned about the two special models: the binomial which counts the number of success in n Bernoulli trials, and we learned about the normal. For sure, you should know how to compute using the normal distribution.

Chapter 5 We start with identical samples from some parent population, $X_1, X_2, \ldots X_n$. (What does that mean) and then form some statistic such as $\bar{y}$ but others are also available (the median, trimmed mean...). Since the X's are random, the statistics are too. How so? The description involves describing the sampling distribution of the statistic. We use words to describe the shape (bell-shaped ...) and the summaries of the mean and standard deviation to describe the shape with numbers.

We central limit theorem is the most important example. It says that $\bar{y}$ has a simple distribution: For large n it is approximately normal with mean μ and standard deviation σ_X/n. In the lab we discussed how the spread of other statistics varies with the input data (the X's). For long-tailed distributions we saw that the spread can be less for the median than $\bar{y}$.

For this material, you should be able to do a simulation with the computer, and use the central limit theorem to compute if the average is too large.