1 Chapter 1

2 Chapter 2

3 Chapter 3. The Histogram

1. In a histogram, the areas of blocks represent percentages, and the total area under histogram should be 100
2. To figure out height of of block over class interval, divide percentage by length of interval.
3. Standard convention for histograms is X-Axis : Property in some units, Y-Axis: Percent per Unit.
4. The height of histogram represents crowding in that particular interval.
5. A variable is a characteristic of the subjects in study. It can be either qualitative or quantitative(discrete or continuous).
6. A co-founding factor is sometimes controlled for by cross-tabulation.

4 Chapter 4. Average and Standard Deviation

1. The average of a list of numbers is equal to sum divided of how many there are.
2. The median is a positional entity, with half of observations falling greater and rest half of observation falling lower than itself.
3. RMS of a list is root of the mean of the square of items in the list. i.e. root(mean(sqr(items)))
4. SD of a list is root of the mean of the square of deviation of items from original avg. i.e. root(mean(sqr(deviations)))
5. Roughly 68% of entries on a list are within one SD of average, Roughly 95% are within 2 SDs of avg. (This assumption only holds for data that can be approximated by a normal curve)

5 Chapter 5. The Normal Approximation for Data

1. A value is converted to Standard Units (S.U) by seeing how many SDs it is above or below the average. The avg lies at 0 S.U and differences from avg are divided by SDs to give SU.
2. When you convert X-Axis to S.U, you should convert Y-Axis to percent per S.U. which will scaled (Y Axis = original value x S.D)
3. If a histogram does not follow a normal curve, then mean and S.D are poor summar statistics.

4. **Percentiles** All histograms cann be summarized using percentiles.

5. Percentiles can be calculated by cummulative addition of percentages of each classes. When we say X_{th} percentile is k, we mean $X\%$ of population has value less than or equal to k.

6. Interquartile range = 75th percentile - 25th percentile

7. Adding same number k to every element in list, makes avg = k + oldAvg, but S.D does not change. (think of all points shifting to the right, mean is change, but spread is same)

8. Multiplying same number k to every element in list, makes avg = k * oldAvg, S.D = k * oldSD (scaling argument)

6 Chapter 6. Measurement Error

1. No matter how carefully it was made, a measurement could have come out a bit differently. If the measurement is repeated, it will come out a bit different. By how much? The best way to answer this question is to replicate the measurement.

2. the S.D. of a series of repeated experiments estimates the likely size of chance error in a single measurement. (dont forget the single measurement part)

3. Bias affects all the measurements the same way, pushing them in the same direction. Chance errors change from measurement to measurement, sometimes up and sometimes down.

4. individual measurement = exact value + bias + chance error.

7 Chapter 8. Correlation

1. Scatter plots are important.

2. **Point of averages** The point whose x-cordinate is average of all x values, and y-coordinate is average of all y values.

3. Correlations are always between -1 and 1, but can take any value in between. A positive correlation means clouds slope up; as one variable increases so does the other. A negative correlation means the clouds slope down, as one variable increases, the other decreases.

4. **The SD Line** The SD Line is a line that passes through (μ_x, μ_y) and has slope $(\frac{SD_y}{SD_x})$

5. **Computing correlation co-efficient** Convert each variable to Standard units, The average of the products gives the correlation coefficient.

$$r = \frac{1}{n} \sum_i \left(\frac{x_i - \mu_x}{SD_x} \right) \left(\frac{y_i - \mu_y}{SD_y} \right) \tag{1}$$