Review for Exam 1

Chapter 1 (Introduction)

1. Be able to determine if a study is an anecdotal, observational or experimental. (question to be brought in)

2. Be able to identify which sampling type best describes the way that the data is collected: simple random sampling, random cluster sampling or stratified random sampling.

3. Be able to identify the sources of sampling bias and how to possibly rectify the situation. (question to be brought in)

Chapter 2 (Descriptive Statistics)

4. Be able to identify the variable in a situation and determine whether it is categorical (ordinal vs. nominal) or numerical (continuous vs. discrete)

5. Given a set of data,
   a) be able to draw a histogram (using frequency and relative frequency).
   b) be able to calculate the five-number summary minimum, Q_1, median, Q_3, maximum.
   c) Be able to determine if a point is an outlier: $y < Q_1 - 1.5 * IQR$ or $y > Q_3 + 1.5 * IQR$
   d) be able to draw a boxplot (modified) given the five-number summary

6. Be able to describe the distribution
   a) via shape – unimodal or bimodal, symmetrical, right or left skewed
   b) by the center
      i) by calculating the median
      $ii)\ calculating\ the\ mean, \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
   c) by the dispersion
      i) sample range = maximum – minimum
      ii) interquartile range (IQR) = $Q_3 - Q_1$
      $iii)\ sample\ standard\ deviation, s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - n\overline{y}^2}{n-1}}$
      $iv)\ sample\ variance, s^2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - n\overline{y}^2}{n-1}$
      $v)\ coefficient\ of\ variation\ (c.v.) = \frac{s}{\overline{y}}$

7. Describe the advantages and disadvantages of the methods describe in 6b and 6c.

8. State which descriptive measures display resistance and which don't.

9. Be able to transform variables, if the new variable is y' = cy + a, then
   a) $\overline{y}' = c\overline{y} + a$
   b) s' = |c| s
   c) $s'^2 = c^2 s^2$

10. Be able to determine a possible population from a given sample and vice versa.

11. Know the symbols for sample characteristics and population characteristics

characteristic	sample	population
proportion	$\hat{p}$	p
mean	$\bar{y}$	μ
SD	s	σ

and which is an estimation of which.

Chapter 3 (Probability)

12. Be able to determine a probability using the
   a) frequency interpretation: $\Pr(E) = \frac{\#\ of\ times\ E\ occurs}{\#\ of\ times\ chance\ operation\ is\ repeated}$
   b) equally likely assumption
   c) Probability tree

13. Be able to define and use: sensitivity, specificity, true positive, false positive, true negative, false negative.

14. State the probability rules and be able to use them:
   a) the probability of an event is between 0 and 1
   b) the sum of the probabilities of all events is 1
   c) $\Pr(E^c) = 1 - \Pr(E)$

15. Determine if two events are disjoint (by 'inspection')

16. Be able to calculate the 'or' (union) of two events
   a) Pr(E or F) = Pr(E) + Pr(F) – if E and F are disjoint
   b) Pr(E or F) = Pr(E) + Pr(F) – Pr(E and F) if E and F are not disjoint

17. Be able to calculate a conditional probability

$$P(A|B) = \frac{P(A\ and\ B)}{P(B)}$$

18. Be able to determine if two events are independent
   P(A|B) = P(A) if A and B are independent.

19. Be able to calculate the 'and' (intersection) of two events
   a) P(E and F) = P(E) x P(F) – if E and F are independent (probability tree)
   b) P(E and F) = P(E) x Pr(F|E) – if E and F are NOT independent

20. Be able to calculate probabilities given the areas under a density curve.

21. Be able to state the definition of a random variable.

22. For discrete random variables, be able to calculate:
   a) mean: $E(Y) = \mu_Y = \sum y_i \Pr(Y = y_i)$
   b) variance: $Var(Y) = \sigma_Y^2 = \sum (y_i - \mu_Y)^2 \Pr(Y = y_i)$
   c) standard deviation $\sigma_Y = \sqrt{Var(Y)}$

23. Be able to use the following formulas for the sum/difference of the mean, variance and standard deviation of two random variables
   a) E(X ± Y) = E(X) ± E(Y)
   b) Var(X ± Y) = Var(X) + Var(Y), if X and Y are independent
   c) $\sigma_{X \pm Y} = \sqrt{Var(X) + Var(Y)}$

24. Binomial distribution
   a) Be able to determine if a given situation follows a binomial distribution using BInS: Binary outcomes, Independent trials, n is fixed, Pr(S) = p is constant
   b) Be able to calculate the probability of y successes among n trials:

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, where \binom{n}{y} = {}_nC_y = \frac{n!}{y!\,(n-y)!}$$

   c) If Y~ Binomial(n,p), then be able to calculate the mean, variance and standard deviation
   E(Y) = np, Var(Y) = np(1-p), $\sigma_Y = \sqrt{np(1-p)}$

Chapter 4 (Normal Distribution)

25. Normal Distribution
   a) Be able to state what types of random variables can have normal distributions (Continuous) and why we are so interested in this type of distribution (it describes a lot of data very well and is useful in statistics)
   b) Be able to convert a normal distribution into a standard normal distribution and vice versa:
   If Y ~ N(μ,σ) and Z = $\frac{Y-\mu}{\sigma}$ (y = μ + zσ), then Z ~ N(0,1)
   c) Be able to use the Z-table to calculate probabilities and percentiles

26. Assessment of shape of data distribution
   a) Be able to make an educated guess if a data set is normal by using the 68%-95%->99% rule.
   b) Be able to make an educated guess if the distribution of the data set is normal, right skewed, left skewed, long tailed or short tailed looking at the QQPlot.

Chapter 5 (Sampling Distributions)

27. Determine if a situation describes a sampling distribution (mean or average) or the distribution of the data.

28. For a quantitative variable, be able to calculate the new mean and standard deviation for a sampling distribution.

$$\bar{Y} \sim Normal\left(\mu, {}^{\sigma}/_{\sqrt{n}}\right) \; where \; Y{\sim}Normal(\mu, \sigma)$$

29. For a quantitative variable, be able to determine if $\bar{Y}$ can be approximated by a normal distribution (because of the CLT, if n is 'large' enough, then $\bar{Y}$ can be approximated by a normal distribution) and then calculate probabilities if the sampling distribution is normal.

30. Be able to determine when the normal approximation to the binomial distribution can be used.
   np ≥ 5 and n(1 - p) ≥ 5.

31. Be able to use the normal approximation to the binomial distribution using the continuity correction.
   If Y ~ Binomial(n,p), then Y can be approximately by a normal distribution with μ = np and $\sigma = \sqrt{np(1-p)}$