Review for Exam 1

Chapter 1

1. Be able to identify the variable in a situation and determine whether it is categorical or numeric (continuous vs. discrete).

2. Be able to identify data as being univariate, bivariate or multivariate.

3. Be able to determine if a situation uses probability or statistical inference.

4. Determine if the best sampling method is a simple random sample or a stratified sampling.

5. Determine if a particular sampling scheme uses convenience sampling and explain why this is not a good sampling method.

6. Given a set of data (note: interpret means describe the shape and determine if there are outliers – see 7 below),
   a) be able to draw a stem-and-leaf plot, refined stem-and-leaf display, comparison stem-and-leaf display and use it to calculate Q_1, median, Q_3
   b) be able to interpret a dotplot.
   c) be able to interpret a histogram (using frequency and/or relative frequency) for both discrete, continuous and categorical data.
   d) be able to draw and interpret a boxplot (with outliers) and comparative boxplot (with outliers) given the minimum, Q_1, median, Q_3, maximum.

7. Be able to describe the shape of the distribution
   a) number of peaks: unimodal, bimodal, multimodal
   b) symmetry: symmetric, positively (right) skewed, negatively (left) skewed

8. Be able to calculate the location of the center
   $a)\ mean, \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
   b) median, $\tilde{x}$
   c) Q_1, Q_3
   d) trimmed mean

9. Be able to calculate the variability
   $$a)\ sample\ variance, s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{S_{xx}}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}{n-1}$$
   $b)\ sample\ standard\ deviation, s = \sqrt{s^2}$
   c) interquartile range (IQR) = $Q_3 - Q_1$

10. Be able to state which locations of center and types of variable are resistant to outliers, i.e., do not change that much when there are outliers.

11. Outlier
   a) Be able to determine if a point is a mild outlier: $y < Q_1 - 1.5 * IQR$ or $y > Q_3 + 1.5 * IQR$
   b) Be able to determine if a point is an extreme outlier: $y < Q_1 - 3 * IQR$ or $y > Q_3 + 3 * IQR$

Chapter 2

12. Be able to determine the sample space of a random experiment.

13. Be able to use the definitions of set theory
    a) union: A U B
    b) intersection: A ∩ B
    c) Compliment: A^c or A'

14. Be able to determine if two sets (events) are disjoint.

15. Be able to use Venn diagrams to calculate probabilities.

16. Be able to use the following axioms/properties of probability
    a) For any event E, P(A) ≥ 0
    b) P($\mathcal{S}$) = 1
    c) If A_1, A_2, A_3, … is a countably infinite (finite) collection of mutually exclusive (disjoint) events, then $P(A_1 \cup A_2 \cup A_3 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i)$
    d) P(∅) = 0
    e) For any event A, P(A) = 1 – P(A')
    f) For any event A, P(A) ≤ 1
    g) General Addition Rule: for any two events A and B,
    P(A U B) = P(A) + P(B) – P(A ∩ B)

17. Be able to determine a probability using equally likely outcomes
$$\mathrm{P}(A) = \frac{N(A)}{N}$$

18. Be able to determine what method(s) is(are) appropriate to calculate probabilities (Note: In these questions, the work includes the formula for the permutation and/or combination if used)
    a) product rule (or use a tree diagram)
    $b)\ permutations$: $P_{k,n} = n(n-1)\cdots(n-k+1) = \dfrac{n!}{(n-k)!}$
    $c)\ combinations$: $C_{k,n} = \binom{n}{k} = \dfrac{P_{k,n}}{k!} = \dfrac{n!}{k!\,(n-k)!}$

19. Be able to calculate a conditional probability (or use a tree diagram)
$$P(A|B) = \frac{P(A\ \cap\ B)}{P(B)}$$

20. Be able to use the general multiplication rule for 2 or more events: P(A ∩ B) = P(A|B) · P(B)

21. Be able to use Bayes' Theorem (or use a tree diagram)
$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

22. Be able to determine if two events are independent
    P(A|B) = P(A) if A and B are independent.
    P(A ∩ B) = P(A) · P(B) if A and B are independent (Electrical Circuits)

Chapter 3

23. For a particular situation, be able to determine the r.v.

24. Be able to determine if a r.v. is discrete or continuous.

25. Be able to use a pmf to calculate probabilities.

26. Be able to calculate the cdf from the pmf and vice versa.

27. Be able to use the cdf to calculate probabilities (discrete).

28. For discrete random variables, be able to calculate the mean:

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

29. Be able to use the properties of the mean (sample and population though only population are listed).

   $a)\ E[h(X)] = \sum_{D} h(x) \cdot p(x)$

   b) $E(aX + b) = aE(X) + b]$

   c) For r.v. $X_1, \ldots, X_n$, $E(a_1X_1 + \ldots + a_nX_n) = a_1E(X_1) + \ldots + a_nE(X_n)$

30. For discrete random variables, be able to calculate the variance and standard deviation

$$Var(X) = \sigma_X^2 = \sum_{D} \left(x - E(X)\right)^2 \cdot p(x) = E\left(X - E(X)\right)^2 = E(X^2) - [E(X)]^2$$

$$\sigma_X = \sqrt{Var(X)}$$

31. Be able to use the properties of the variance (sample and population though only population is listed).

   a) $Var(aX + b) = a^2Var(X)$

   b) $\sigma_X = |a|\sigma_X$

   $c)\ Var[h(X)] = \sum_{D} \{h(x) - E[h(X)]\}^2 \cdot p(x) = E(h^2(x)) - [E\left(h(x)\right)]^2$

32. Binomial distribution (Note: work does NOT include the formula for the combination). Remember Objective 16.

   a) Be able to determine if a given situation follows the conditions for a binomial experiment (BInS). (Multiple Choice Only)

   b) Be able to calculate the probability of x successes among n trials (pmf):

   $$b(x; n, p) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \ldots, n$$

   c) Calculate the mean, variance and standard deviation of a binomial distribution

   $E(X) = np$, $Var(X) = np(1-p)$, $\sigma_X = \sqrt{np(1-p)}$

   d) Be able to determine when the hypergeometric distribution can be approximated by the binomial distribution

   $$\frac{n}{N} \le 0.05$$

33. Hypergeometric distribution (Note: work does NOT include the formula for the combination). Remember Objective 16.

a) Be able to determine if a given situation follows the conditions for a hypergeometric distribution (Multiple Choice only)

b) Be able to calculate the probability of x successes among n trials with M total successes in a population of N (pmf):

$$h(x; n, M, N) = P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

c) Calculate the mean, variance and standard deviation of a hypergeometric distribution

$$E(X) = n \cdot \frac{M}{N}, Var(X) = \left(\frac{N-n}{N-1}\right) \cdot np(1-p)$$