Review for Exam 1

Tables provided: Z-table (Table III from the book).

Chapter 1 (except 1.3)

1. Be able to briefly define the three branches of statistics and be able to determine which one that a specific occurrence belongs to: Data Collection, Descriptive Statistics, Inferential Statistics.

2. Be able to define a population and a sample and know whether to indicate them with a Greek or Latin letter.

3. Be able to determine if a listing of objects refers to the population or the sample.

4. Be able to determine if a specific situation exemplifies probability or inferential statistics.

Chapter 2

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

6. Be able to identify the variable in a situation and determine whether it is categorical or quantitative.

7. Be able to determine if a variable is discrete or continuous.

8. Given a set of data, be able to interpret histograms
   a) interpret means describe the shape and determine if there are outliers

9. Be able to describe the shape of the distribution
   a) Number of peaks: unimodal, bimodal, multimodal
   b) Symmetry: symmetric, right skewed or left skewed.

Chapter 3

10. Be able to calculate the location of the center for discrete data.
    $a)\ mean, \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
    b) median, $\tilde{x}$. Work required is the sorted data with some sort of indication of where the median is located
    c) mode: no work required.

11. Be able to calculate the sample variance and standard deviation
    $a)\ sample\ variance, s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{1}{n-1}\left[\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2\right]$
    Note: the summations will be provided if required.
    $b)\ sample\ standard\ deviation, s = \sqrt{s^2}$

12. Be able to calculate Q_1, Q_3 and IQR = $Q_3 - Q_1$. Work required is the sorted data with the calculations of $d_1 = n/4$ and $d_3 = 3n/4$

13. Be able to determine if the sample set contains outliers (it is not necessary to differentiate between mild and extreme outliers)
    a) $IF_L < Q_1 - 1.5 * IQR$ or $IF_H > Q_3 + 1.5 * IQR$

14. Be able to write down the five-number summary: minimum, Q_1, median, Q_3, maximum

15. Given a set of data, be able to draw and interpret a modified boxplot (with outliers) using the five-number summary
    a) interpret means describe the shape and determine if there are outliers
    b) Be able to interpret side-by-side boxplots.

16. Be able to state which of the measures of location and spread should be used in a specific example (whether there are outliers or not).

Chapter 4

17. Be able to write down the sample space for a specific experiment.

18. Be able to determine if two events are disjoint.

19. Be able to state the frequentist interpretation of probability and when this is valid.
    $\lim_{n\to\infty} \frac{n(E)}{n} \approx P(E)$

20. Be able to recognize and use the properties of probability.
    a) $0 \leq P(A) \leq 1$
    b) $P(A) = \sum_i \omega_i$
    c) $P(S) = 1$
    d) $P(\{\}) = 0$

21. Be able to calculate a probability via
    $a)\ P(A) = \frac{number\ of\ times\ that\ A\ occurs}{total\ number\ of\ times\ that\ you\ performed\ the\ experiment}$
    (equally likely or theoretical)
    b) a table of the probability distributions of a discrete random variable (empirical)

22. Be able to use Venn diagrams to calculate probabilities

23. Be able to use the Probability Rules
    a) $P(A') = 1 - P(A)$ (Complement)
    b) $P(A \cup B) = P(A) + P(B) - P(A \text{ and } B)$ (General Addition)
    c) $P(A \text{ or } B) = P(A) + P(B)$.(if A and B are disjoint)

24. Be able to calculate a conditional probability
    $P(A|B) = \frac{P(A\ and\ B)}{P(B)}$

25. Be able to use the General Multiplication rule.
    a) $P(A \text{ and } B) = P(A)\ P(B|A) = P(B)\ P(A|B)$
    b) $P(A \text{ and } B \text{ and } C) = P(A)\ P(B|A)\ P(C|A \text{ and } B)$

26. Be able to use tree diagrams and Bayes's Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

27. Be able to determine if two events are independent. A and B are independent if
    a) P(A|B) = P(A)
    b) P(B|A) = P(B)
    c) P(A and B) = P(A) · P(B) (also valid for more than 2 events) – Multiplication Rule

28. Be able to differentiate between independent events and disjoint events.

Chapter 5

29. Be able to recognize and use the properties of a valid probability distribution.
    a) $0 \le p(x) \le 1$
    b) $\sum_i p(x) = 1$

30. Be able to calculate probabilities using a pmf (discrete distribution).

31. Be able to calculate the mean of a discrete random variable.

$$\mathrm{E}(X) = \mu_X = \sum x \cdot p(x)$$

32. Be able to use the rules for the means
    1) E(a + bX) = a + bE(X)
    2) E(X ± Y) = E(X) ± E(Y)
    3) $\mathrm{E}(g(X)) = \sum g(x) \cdot p(x)$

33. Be able to calculate the variance and standard deviation:
    a) $\mathrm{Var}(X) = \sigma_X^2 = \mathrm{E}[(X - \mu_X)^2] = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$
    b) $\sigma_X = \sqrt{\sigma_X^2} = \sqrt{Var(X)}$

34. Be able to use the rules of the variance (standard deviation)
    1) $\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$
    2) If X and Y are independent, Var(X ± Y) = Var(X) + Var(Y)

35. Binomial distribution
    a) Be able to determine if a distribution is Binomial or not (BInS)
    b) Be able to determine the parameters, n and p, from the context of the question.
    c) Be able to calculate the probabilities:

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

    d) Be able to calculate P(X > x), P(X ≥ x), P(X < x), P(X ≤ x)
    e) Be able to determine the skewedness of the distribution from the value of p
       p < 0.5: right skewed, p = 0.5: symmetric, p > 0.5 left skewed
    e) Be able to calculate mean and standard deviation
       E(X) = np
       $\sigma = \sqrt{np(1-p)}$

36. Be able to write down what is meant by the cumulative density function (CDF)
    F(x) = P(X ≤ x)

37. Poisson distribution
    a) Be able to state when a Poisson distribution is appropriate
    b) Be able to determine the parameter, λ
    c) Be able to calculate the probability of x successes

$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, x = 0,1,2, \ldots$

    d) Be able to calculate P(X > x), P(X ≥ x), P(X < x), P(X ≤ x)
    e) Be able to calculate mean and standard deviation
        E(X) = λ
        $\sigma = \sqrt{\lambda}$

Chapter 6

38. Be able to calculate a probability, median (percentile) and mean from a density function (continuous random variable)
    a) For any two numbers a < b

$proportion\ of\ values\ between\ a\ and\ b = \int_{a}^{b} f(x)dx$

$b) percentile: \int_{0}^{y} f(x)\,dx = p$

$c)\ median \int_{0}^{\tilde{\mu}} f(x)\,dx = 0.5$

$d)\ E(X) = \mu_x = \int_{-\infty}^{\infty} xf(x)dx$

39. Be able to use the rules for the means for continuous random variables
    c) $\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$

40. For the normal distribution: be able to use the z-table to calculate probabilities and percentiles.

41. For the normal approximate to the binomial:
    a) Be able to determine if the normal approximation to the binomial is appropriate
        np ≥ 10 and n(1 – p) ≥ 10
    b) Be able to use the normal approximation to the binomial (with or without the continuity correction will be stated in the question)
    c) Be able to state why the continuity correction is useful

42. For a given QQ plot, be able to determine if it is plausible if the distribution is normal or not or determine if it is skewed, or symmetric, but not normal.

43. Uniform distribution:
    a) Be able to calculate probabilities

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x < b \\ 0 & otherwise \end{cases}$$

    b) Be able to calculate mean and standard deviation

$$E(X) = \frac{a+b}{2}$$

$$\sigma = \sqrt{\frac{(b-a)^2}{12}}$$

44. Exponential distribution:
    a) Be able to calculate probabilities

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & otherwise \end{cases}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases}$$

    b) Be able to calculate and/or use the mean and standard deviation

$$E(X) = \frac{1}{\lambda}$$

$$\sigma = \frac{1}{\lambda}$$

Chapter 7

45. Be able to state what a parameter and statistic are in terms of population and sample.

46. Be able to determine if a problem is using a sampling distribution and state why it is so important

47. Be able to calculate the mean (μ) and standard deviation $\left(\frac{\sigma}{\sqrt{n}}\right)$ of a sampling distribution for the sample mean ($\bar{x}$).

48. Be able to calculate a probability from a sampling distribution where the population distribution is normal.

49. Using the Central Limit Theorem (CLT), be able to determine if the average of sample total from an unknown underlying distribution can be considered normal with appropriate mean and standard deviation.
    a) Be able to calculate probabilities using this information.