1 Chapter 1: Descriptive Statistics

Sample Average Population Average
$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$ $$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

To calculate the $p$’th percentile $x_{[p]}$:
1. Let $x_{(i)}$ refer to our data set in ascending order.
2. Let $i_p = np/100$.
3. Find the first index $i$ such that $i > i_p$.
4. The $p$’th percentile is then:
   $$x_{[p]} = \begin{cases} \frac{x_{(i-1)} + x_{(i)}}{2} & \text{if } i - 1 = i_p \\ x_{(i)} & \text{otherwise} \end{cases}$$

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = \frac{1}{n-1} \left( \sum x_i^2 - \frac{\left( \sum x_i \right)^2}{n} \right)$$

Chebychev’s Rule: The proportion of observations that are within $k$ standard deviations $(p_k)$ of the mean is at least:
$$p_k = 1 - \frac{1}{k^2}$$

2 Chapter 2: Probability

Multiplication rule Permutation Combination
$$n_1 \times n_2 \times n_2 \ldots \times n_k$$ $$P_{k,n} = \frac{n!}{(n-k)!}$$ $$(n)_k = \frac{n!}{k!}$$

For any two events $A$ and $B$:
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

The conditional probability of $A$ given that $B$ occurred $(P(B) > 0)$:
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Two events $A$ and $B$ are independent if
$$P(A|B) = P(A)$$

If $A$ and $B$ are independent then
$$P(A \cap B) = P(A)P(B)$$

If $A, B, C, D, \ldots$ are mutually independent then
$$P(A \cap B \cap C \cap D \ldots) = P(A)P(B)P(C)P(D) \ldots$$
3 Chapter 3: Discrete PDF’s

\[ E[X] = \mu = \sum_{X \in S} x \cdot p(x) \]
\[ E[h(X)] = \mu_{h(x)} = \sum_{X \in S} h(X) \cdot p(x) \]
\[ E(aX + b) = aE(X) + b \]
\[ V(X) = \sigma^2 = E[(x - \mu)^2] = \sum_{X \in S}(x - \mu)^2 \cdot p(x) = E[X^2] - E[X]^2 \]
\[ V(aX + b) = a^2V(X) = a^2\sigma^2 \]

3.1 Binomial Distribution

For \( X \sim \text{binomial}(n, p) \)
\( n \) = fixed number of trials
\( p \) = probability of succes (S)
\( x \) = number of successes (S)
\[ P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad x = 0, 1, 2, \ldots, n \]
\[ \mu = E[X] = np \]
\[ \sigma^2 = V[X] = E[(x - \mu)^2] = np(1 - p) \]

3.2 Multinomial Distribution

For \( X \sim \text{multinomial}(n, p_1, \ldots, p_r) \)
\( n \) = Number of trials.
\( r \) = Number of possible outcomes.
\( p_i \) = \( P(\text{Outcome } i \text{ on any particular trial}) \).
\( x_i \) = Number of trials resulting in outcome i.
\[ p(x_1, x_2, \ldots, x_r) = \frac{n!}{x_1!x_2! \ldots x_r!} p_1^{x_1} p_2^{x_2} \ldots p_r^{x_r} \]
\( x_i = 0, 1, 2, \ldots \quad x_1 + x_2 + \ldots x_r = r \)
3.3 Hypergeometric Distribution

For $X \sim \text{hypergeometric}(n, M, N)$

- $n$ = sample size
- $M$ = number of $S$ in the population
- $N$ = population size
- $x$ = number of $S$ in the sample

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

where $x : \max(0, n - N + M) \leq x \leq \min(n, M)$.

- $E[X] = \mu = n \frac{M}{N} = np$ where $p = \frac{M}{N}$
- $V[X] = \sigma^2 = (\frac{N-n}{N-1}) n \frac{M}{N}(1 - \frac{M}{N})$
  $$= (\frac{N-n}{N-1}) np(1 - p)$$

3.4 Binomial Approximation to the Hypergeometric

If we sample with replacement of if $n$ is small relative to $N$ and $M$, we can approximate the Hypergeometric distribution by using the binomial distribution with $p = \frac{M}{N}$:

$$X \sim \text{hypergeometric}(n, M, N) \rightarrow X \sim \text{binomial}(n, p = \frac{M}{N})$$

3.5 Hypergeometric Distribution for $k$ Cells

$N$ items are partitioned into $k$ cells $A_1, A_2, \ldots, A_k$ with $a_1, a_2, \ldots, a_k$ elements respectively. Then the probability distribution of the random variables $X_1, X_2, \ldots, X_k$ representing the number of elements selected from $A_1, A_2, \ldots, A_k$ in a random sample of size $n$ is:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{\binom{a_1}{x_1} \binom{a_2}{x_2} \cdots \binom{a_k}{x_k}}{\binom{N}{n}}$$

where

$$\sum_{i=1}^{k} x_i = n \quad \sum_{i=1}^{k} a_i = N$$

For the case $k = 2$:

- $A_1$ = $S$ (success)
- $A_2$ = $F$ (failure)
- $a_1$ = $M$ (number of $S$ in $A_1$)
- $a_2$ = $N - M$ (number of $F$ in $A_2$)
- $n$ = sample size
- $N$ = population size
3.6 Negative Binomial Probability Distribution

For $X \sim \text{negative binomial}(r, p)$

- $r$ = number of $S$
- $p$ = probability of $S$
- $x$ = the number of failures preceding the $r$’th success

\[
P(X = x) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x, \quad x = 0, 1, 2, \ldots
\]

\[
E[X] = \mu = \frac{r(1-p)}{p}
\]

\[
V[X] = \sigma^2 = \frac{r(1-p)}{p^2}
\]

If $r = 1$ we have a Geometric distribution.

\[
P(X = x) = p(1-p)^x, \quad x = 0, 1, 2, \ldots
\]

3.7 Poisson Distribution

For $X \sim \text{poisson}(\lambda)$

- $\lambda$ = the rate per unit time or rate per unit area.
- $x$ = the number of successes occurring during a given time interval or in a specified region

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

\[
E[X] = \mu = \lambda
\]

\[
V[X] = \sigma^2 = \lambda
\]

3.8 Poisson Approximation to the Binomial Distribution

Let $X$ be a binomial random variable with probability distribution $X \sim \text{binomial}(n, p)$. When $n \to \infty$ and $p \to 0$ and $\lambda = np$ remains fixed at $\lambda > 0$, then

\[
X \sim \text{binomial}(n, p) \to X \sim \text{poisson}(\lambda = np)
\]

As a rule of thumb, this approximation can be safely applied if:

\[
n \geq 100 \quad p \leq .01 \quad np \leq 20
\]
Chapter 4: Continuous PDF’s

\[ P(a \leq X \leq b) = \int_a^b f(x) \, dx \]

\[ F(x) = P(X \leq x) = \int_{-\infty}^x f(y) \, dy \]

\[ P(a \leq X \leq b) = F(b) - F(a) \]

\[ F'(X) = f(x) \]

For \(0 \leq p \leq 1\) the \((100p)\)’th percentile of a continuous distribution you must solve \(p = F(x)\) for \(x\) where \(x\) is the \((100p)\)’th percentile.

\[ E[h(x)] = \int_{-\infty}^{\infty} h(x) \cdot f(x) \, dx \]

\[ E[X] = \mu = \int_{-\infty}^{\infty} x \cdot f(x) \, dx \]

\[ E[(X - \mu)^2] = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) \, dx \]

Remember: \(E[(X - \mu)^2] = E[X^2] - E[X]^2 = \sigma^2\)

### 4.1 The Uniform Distribution

The family of uniform distributions has the following PDF:

\[ f(x, A, B) = \begin{cases} \frac{1}{B - A} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases} \]

### 4.1.1 The Exponential Distribution

The family of exponential distributions has the following PDF:

\[ f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \lambda > 0 \\ 0 & \text{otherwise} \end{cases} \]