Statistics Crib Sheet

Introduction

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_1 - \bar{X})^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n}\left(x_1^2 - n\bar{X}^2\right) \\ s &= \sqrt{s^2} \\ &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_1^2 - n\bar{X}^2)} \\ &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_1^2 - n\bar{X}^2)} \end{aligned}$$

$$\begin{aligned} \bar{Y} &= a + b\bar{X} \\ S_Y^2 &= b^2 s_X^2 \\ S_Y &= |b| s_x \end{aligned}$$

$$p\text{th percentile} = \frac{p}{100}(n+1)$$

- A histogram is perfectly symmetric if its right half is a mirror image of its left half.
- A histogram with a long right-hand tail is called skewed to the right or positively skewed.
- A histogram with a long left-hand tail is called skewed to the left or negatively skewed.
- A histogram is unimodal if it has only one peak, or mode. A histogram is bimodal if it has two peaks, or mode. A bimodal histogram, in some cases, indicates that the sample can be divides into two subsamples that differ from each to other. If there are more than two peaks in a histogram, then it is said to be multimodal.
- Steps in construction of a boxplot

1. Computer the median and first and third quartiles of the sample. Indicate these with horizontal lines.
2. Find largest sample value that no more than $1.5IQR$ above the third and quartile, and smallest value less than the first quartile.
3. Plot points. (first quartile) $1.5IQR < x < 1.5IQR$ (third quartile)

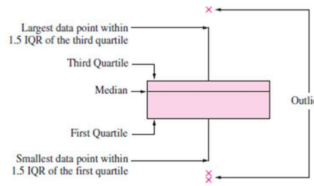

FIGURE 1.13 Anatomy of a boxplot.

Probability

1. Let S be a sample space. Then $P(S) = 1$
2. For any event, $0 \leq P(A) \leq 1$
3. If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$.

Complement Rule

Complement Rule $P(A^C) = 1 - P(A)$

Addition Rule $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- The probability with equally likely outcomes has a probability $P(A) = \frac{K}{N}$
- The number of permutation of k objects chosen from a group of n objects is

$$P_{n,k} = \frac{n!}{(n-k)!}$$

- The number of combinations of k objects chosen from a group of n objects

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- The number of ways of dividing a group of n objects into groups of k_1, k_2, ..., k_n

$$\frac{n!}{k_1!k_2!\ldots k_r!}$$

- A probability that is based on a part of a sample space called a conditional probability.

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

- If A and B are two events with $P(B) \neq 0$, then

$$\begin{aligned} P(A \cup B) &= P(B)P(A|B) \quad \forall B \in \mathbb{R}, B \neq 0 \\ &= P(A)P(B|A) \quad \forall A \in \mathbb{R}, A \neq 0 \end{aligned}$$

- Let A and B be events with $P(A) \neq 0$, $P(A^C) \neq 0$, and $P(B) \neq 0$.

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

Generally,

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

- The cumulative distribution function of r.v X is defined as:

$$F(X) = P(X \leq x) = \sum P(X \leq t) = \sum P(t)$$

- Let X be a discrete random variable with probability mass function $p(x) = P(X \leq x/$ The **Mean** is given by

$$\begin{aligned} \mu_x &= \sum x P(X = x) \\ &= \int_{-\infty}^{\infty} x\, f(x)\, dx \end{aligned}$$

- The **variance** of X is given by

$$\begin{aligned} \sigma_X^2 &= \sum (x - \mu_x)^2 P(X = x) = \sum x^2 P(X = x) - \mu_x^2 \\ &= \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2\, f(x)\, dx - \mu_x^2 \end{aligned}$$

- The median of X is the point x_m that solves the equation

$$F(x_m) = P(X \leq x_m) = \int_{-\infty}^{x_m} f(x)\, dx = 0.5$$

- The pth percentile is

$$F(x_m) = P(X \leq x_m) = \int_{-\infty}^{x_m} f(x)\, dx = \frac{p}{100}$$

- If a random variable X is multiplied by a constant a and then added to another constant b, then we have a new random variable Y, where

$$\begin{aligned} Y &= a * X + b \\ E(Y) &= \mu_y = a * E(X) + b \\ Var(Y) &= \sigma_y^2 = a^2 * var(X) \\ \sigma_Y &= |a|\sigma_x \end{aligned}$$

- If $X_1, \ldots, X_n$ is a simple random sample from a population with mean μ and variance σ^2, then the sample mean $\bar{X}$ is a random variable with

$$\begin{aligned} \mu_{\bar{X}} &= \mu \\ \sigma_{\bar{X}}^2 &= \frac{\sigma^2}{n} \end{aligned}$$

The standard deviation of $\bar{X}$ is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Chebyshev's Inequality Let X be a random variable with mean u_x and standard deviation σ_x. Then

$$P(|X - \mu_x| \geq k\sigma_x) \leq \frac{1}{k^2}$$