Ch 6 Review

- random variable : numerical measurement of random phenomenon
- probability distribution : assigns probabiilties to each value of a random variable and have the following properties.
  1. for each value x that X can take, we have $0 \leq P\{X = x\} \leq 1$
  2. $\sum_x P\{X = x\} = 1$
- there are two types of random variables
  1. discrete
  2. continuous
- summary of center (population mean)

$$\mu = \sum_x xP\{X = x\}$$

  also called a weighted average.
- summary of spread : population standard deviation denoted σ.
- note : for the distribution of categorical variables we can use a certain class of discrete random variables called the binomial (discussed in more detail later).

The Normal Distribution

- characterized by parameters mean (μ) and standard deviation (σ)
- cummulative probabilities found on table that was handed out in class. to find cummulative probaiblities, we note that $\Phi(a) = P\{Z \leq a\}$ where $Z \sim N(0, 1)$, called the standard normal random variable.
- because the normal distribution is symmetric about the mean, we have the following properties

$$P\{-a \leq Z \leq a\} = 2\Phi(a) - 1 \qquad P\{Z \leq -a\} = 1 - \Phi(a)$$
$$P\{Z \geq a\} = 1 - \Phi(a) \qquad P\{a \leq Z \leq b\} = \Phi(b) - \Phi(a)$$

- empirical rule is derived from the cummulative probabilities of the normal distribution
- we can use z scores to compare variables with different scales (e.g. SAT and ACT)

Binomial Distribution

- count of independent identically distributed binary r.v.
- has two paramters n and p
  - $n = \#$ of independent trials
  - $p =$ probability of success on each trial
- conditions that a binomial r.v. must meet
  1. each trial is independent of all others
  2. probability of success is the same for each trial
  3. each trial has only two outcomes usually called success and failure with values 1 (for success) and 0 (for failure).
- if $X \sim \text{binom}(n, p)$ then

$$P\{X = k\} = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

- for the binomial distribution,

$$\mu = np \qquad\qquad \sigma = \sqrt{np(1-p)}$$

- since the binomial is symmetric, it can be approximated with the normal distribution. this means that

$$\text{binom}(n, p) \approx N(np, \sqrt{np(1-p)})$$

- ex : it is expected that Joe will have 200 customers on any given Saturday. The probaiblitiy that any customer will purchase something is 0.1.
  1. what are the parameters for the binomial random variable of number of purchases on a Saturday?
  2. what is the probabilty that more than 25 people will purchase something?

Sampling Distribution

- sampling distribution : the probability distribution of a statistic.
- two main types for us
  1. sample proportion. if $X \sim \text{binom}(n, p)$ then X/n has properties $\mu = p$ and standard error $se = \sqrt{p(1-p)/n}$.
  2. sample mean. if X is a quantatative variable, then $\bar{X}$ has properties $\mu = \bar{x}$ and $se = \sigma/\sqrt{n}$.
- the Central Limit Theorem stats that for large n, we have
  1. $X/n \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$
  2. $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$.