Reference Sheet for CH 6 and 7

Variance, and Standard Deviation of a Probability distribution

Variance

[Standard Deviation = the square root of the variance]

Means: $\sigma^2 = \sum (x-\mu)^2 \cdot p(x)$

Proportions: $\sigma^2 = \sum [x^2 \cdot p(x)] - \mu^2$

Binomial Distribution

There are many experiments and situations with results that can be described as yes/no (binary)

Requirements:

1. Binary – there can only be 2 ways the event can turn out
2. Independent – events must be independent for each trial
3. Fixed – there must be a set number of trials
4. Probability of success is constant for all trials and the we are counting successes

Extra Rule – the sample size must be less than 10% of the entire population size.

Formula to find the probability of any value in a binomial distribution:

$$P(x) = \frac{n!}{(n-x)!x!} \cdot p^x \cdot q^{n-x} \quad for\ x = 0,1,2,3,\ldots,n$$

n = # of trials

x = # of successes among n trials

p = probability of success in any one trial

q = probability of failure in any one trial (q = 1 – p)

This formula can be written as $\boldsymbol{P(x=k)} = \boldsymbol{\binom{n}{k} p^k q^{n-k}}$ where the parenthesis is called n choose k. This is the combination formula for choosing k items from n total items. In combinations, order does NOT matter.

Binomial Distribution Formulas: $\mu = np$ $\sigma = \sqrt{np(1-p)}$

The Geometric Distribution

While in a binomial distribution the random variable was the number of successes in a fixed number of trials, in a **geometric distribution** the random variable is the number of trials it takes to achieve a success.

P(X = n) = $(1\text{-}p)^{n\text{-}1}$p where p = probability of success and n is the trial of first success.

Requirements:

1) Each trial in the experiment must have only two possible outcomes (success or failure)
2) The probability of success, p, doesn't change from trial to trial
3) The trials in the experiment are independent
4) The variable of interest is the number of trials required to reach the first success.

The mean, μ, of a geometric distribution (the average number of times we can expect to repeat the trials before a success occurs) is simply 1/p where p is the probability of success.

Geometric Distribution Formulas: $\mu = \frac{1}{p}$ $\sigma = \sqrt{\frac{1-p}{p^2}}$

Parameter: A number that describes an aspect of a population

Statistics: A number that is computed from sample data; often used to estimate an unknown parameter.

Notation	Parameter	Statistic
Proportion	p	$\hat{p}$
Mean	□	$\bar{x}$

Sampling distributions and Sampling Variability

If we take repeated samples from the NWHS senior population and measure the proportion of seniors from those samples that got into college early and their weights, we will see different statistics for the different samples. This is referred to as sample variability.

We can create a distribution of the proportions of all the samples we took and draw a histogram. The histograms below shows the results of a simulation.

Sampling Distribution:

A Sampling Distribution is shape, center, and spread for the collection of **all** possible samples of the same size from the population

Properties:

- The overall shape of the distribution is symmetric and approximately normal. The larger the sample size the closer the shape is to a normal distribution.
- There are no outliers or other important deviations from the main pattern
- The mean (center) of the distribution is equal to the true population parameter
- The variability (spread) of the sampling distribution depends on the sample size. The larger the sample-size the smaller the variability of the sampling distribution.
- Averages are less variable than individual observations. [because σ is divided by $\sqrt{n}$]
- Averages are more normal than individual observations.

Sampling Distribution of a Sample Mean $\bar{x}$ from a Normal Population

Draw an SRS of size *n* from a population that has a normal distribution with mean □ and standard deviation □. The expected value of the sampling distribution of $\bar{x}$ is $\mu_{\bar{x}} = \mu$

Then the sample mean $\bar{x}$ has a normal distribution with mean □ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Since the mean of the sampling distribution is equal to □, this makes $\bar{x}$ an unbiased estimator of □.

✪ ✪ ✪ The Central Limit Theorem ✪ ✪ ✪

Sampling Distribution of a Sample Mean $\bar{x}$ from a Non-Normal Population

Draw an SRS of size n from a population where the shape is unknown or if the shape is skewed that has a normal distribution with mean □ and standard deviation □. The expected value of the sampling distribution of $\bar{x}$ is $\mu_{\bar{x}} = \mu$

The distribution of the sample means will *still* be close to normal as long as the sample size is large enough. The rule of thumb: **we expect a normal distribution if n ≥ 30**.

This idea is called the Central Limit Theorem.

Draw an SRS of size *n* from any population whatsoever with mean □ and finite standard deviation □. When *n* is large the sampling distribution of the sample mean $\bar{x}$ is close to the normal distribution with mean □ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Note: in order to be able to use the standard deviation formula, the sample size must be less than 10% of the population size. So N must be at least 10 times larger than n. so 10n ≤ N