AP Statistics Name___________________________

Important Facts about $\bar{X}$ and $\hat{p}$

Suppose you have a population of *N* people or objects. On each object a **quantitative** variable *X* is measured.

- You want to know about the **mean** value of *X* for all members of the population, denoted by μ. Also, the standard deviation of *X* for all members of the population is denoted by σ.

- Now, to obtain μ would entail examining **every** member of the population. This is too hard, expensive, and time consuming! So what do we do?

- We take a **simple random sample** of size *n* from this population and examine the sample. In particular, we look at the **mean** of the sample, denoted by $\bar{x}$. This should be a good estimate of the quantity we are really interested in, the mean of the population, μ.

The question then becomes: **What do we know about $\bar{x}$?**

- Well, there are literally millions of $\bar{x}$'s … there is an $\bar{x}$ for each possible sample we could select from the population. The distribution of all possible $\bar{x}$'s is known as the **sampling distribution of $\bar{x}$.**

What do we know about the sampling distribution of $\bar{x}$?

(1) The mean of all possible values of the sample mean ($\bar{x}$) equals the population mean μ.

$$\mu_{\bar{x}} = \mu$$

Thus, we say $\bar{x}$ is an **unbiased estimator** for μ.

(2) The standard deviation of all possible values of the sample mean ($\bar{x}$) equals the standard deviation of the population divided by the square root of the sample size.

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

(3) What about the shape of the *s***ampling distribution of $\bar{x}$** ? Take this in two parts…

(A) If the **population** is normally distributed…then the sampling distribution of $\bar{x}$ is normally distributed with mean and standard deviation from (1) and (2) above.

(B) If the **population** is **not** normally distributed…then the sampling distribution of $\bar{x}$ becomes more normal as the sample size *n* increases. This is the **Central Limit Theorem**. Note that even when the population is **not** normally distributed, the facts given in (1) and (2) are **still true**!

Suppose you have a population of N people or objects. On each object a **categorical** variable is measured with each member of the population put into one of two categories, "success" or "failure".

- You want to know about the **proportion** of members of the population that fall into the success category, denoted by p.
- Now, to obtain p would entail examining **every** member of the population. This is too hard, expensive, and time consuming! So what do we do?
- We take a **simple random sample** of size n from this population and examine the sample. In particular, we look at the **proportion** of the sample, denoted by $\hat{p}$, that fall into the success category. This should be a good estimate of the quantity we are really interested in, the proportion of the population, p, which falls into the success category.

The question then becomes: **What do we know about $\hat{p}$?**

- Well, there are literally millions of $\hat{p}$'s … there is a $\hat{p}$ for each possible sample we could select from the population. The distribution of all possible $\hat{p}$'s is known as the **sampling distribution of $\hat{p}$.**

What do we know about the sampling distribution of $\hat{p}$?

(1) The mean of all possible values of the sample proportion ($\hat{p}$) equals the population proportion, p.

$$\mu_{\hat{p}} = p$$

Thus, we say $\hat{p}$ is an **unbiased estimator** for p.

(2) The standard deviation of all possible values of the sample proportion ($\hat{p}$) depends on the sample size and is given by the following formula

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

(3) What about the shape of the **sampling distribution of $\hat{p}$** ?

- If $np \geq 10$ and $n(1-p) \geq 10$ …then the sampling distribution of $\hat{p}$ can be well approximated by a normal distribution with mean and standard deviation from (1) and (2) above.

NOTE: For each of the two situations given (quantitative and categorical), the size of the population should be at least 10 times the sample size.