# Lecture 13

## Binomial distribution

*When do we deal with a binomial trial or distribution?* An experiment is said to be a **binomial experiment** if: 1) The experiment is performed a fixed number of times, usually denoted by $n$. Each repetition is called a trial.
2) The trials are independent (the outcome of one does not depend on the other)
3) For each trial, there are 2 mutually exclusive outcomes: success or failure.
4) The probability of success is fixed for each trial of the experiment.The probability of success is $p$ while of failure is $1 - p$
5) We say that a r.v. is binomially distributed if $X$ counts the number of successes in $n$ independent trials of the experiment. So the possible values for $X$ are 0, 1, 2, ..., $n$.

Mathematicians showed that the probability of obtaining $x$ successes in $n$ independent trials of a binomial experiment where the probability of success is $p$ is given by

$$P(X = x) = {}_nC_x p^x (1-p)^{n-x}, \;\; x = 0, 1, 2, ..., n$$

Also they showed that such a binomial random variable will have the mean given by

$$\mu_X = E(X) = np$$

and the standard deviation given by the formula:

$$\sigma_X = \sqrt{np(1-p)}$$

## Confidence Intervals about a Population Proportion

Suppose a simple random sample of size $n$ is obtained from a population in which each individual either has or has not a certain characteristic. The best point estimate for $p$, the proportion of the population with the characteristic, is given by

$$\hat{p} = \frac{X}{n},$$

where $X$ is the number of individuals in the sample with the characteristic.
*Remark* 1. 1) Distinguish between $X$ and $\hat{p}$
2) $X$ has a binomial distribution while $\hat{p}$ does **not** have a binomial distribution.

## Sampling Distribution of $\hat{p}$

For a simple random sample of size $n$ such that $n \leq 0.05 \times N$ (sample size no more than 5% of population size), the sample distribution of $\hat{p}$ is approximately normal, with mean:

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

and standard deviation:

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

provided that:

$$np(1-p) \geq 10.$$

The formula for $\sigma_{\hat{p}}$ is exactly correct in the binomial setting. It is approximately correct for an SRS from a large population as described above.

*Remark* 2. The fact that the mean of $\hat{p}$ is $p$ states in statistical language that the sample proportion $\hat{p}$ is an **unbiased estimator** for the population proportion $p$.

## Normal Approximation for counts and proportions

For an SRS of size $n$ from a large population having population proportion $p$ of success such that $np \geq 10$ and $n(1-p) \geq 10$, when $n$ is large the sampling distributions of the statistics $X$ and $\hat{p}$ are given by:

$$X \text{ is approximately } N(np, \sqrt{np(1-p)})$$

and

$$\hat{p} \text{ is approximately } N(p, \sqrt{\frac{p(1-p)}{n}})$$