Lecture 9

Binomial distribution

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

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

Sampling distribution of sample mean

Let $\overline{x}$ be the mean of a SRS of size n from a population with mean μ and SD σ. The mean and SD of $\overline{x}$ are

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

$$\sigma_{\overline{x}} \frac{\sigma}{\sqrt{n}}$$

Moreover, if the population is normally distributed then $\overline{x}$ is normally distributed $N(\mu, \frac{\sigma}{\sqrt{n}})$ and if the population is not necessarily normally distributed but the sample size is large then $\overline{x}$ is **approximately** normally distributed, $N(\mu, \frac{\sigma}{\sqrt{n}})$.