Continuing 5.2, 5.4

Recall 1 (CLT). Let $Y_1, \ldots, Y_n$ be independent copies of $Y$. Then

- 1) $\mu_{\bar{Y}} = \mu_Y$
- 2) $\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}$
- 3a) If $n$ is large ($n \geq 30$), then regardless of the distribution of $Y$, $\bar{Y} \sim N(\mu_Y, \frac{\sigma_Y}{\sqrt{n}})$ approximately.
- 3b) If $Y$ is already $N(\mu_Y, \sigma_Y)$, then for any $n$, $\bar{Y}$ is exactly $N(\mu_Y, \frac{\sigma_Y}{\sqrt{n}})$

Note 1. This forms the basis for a large part of statistics.

Example 1. Recall the $\bar{Y}$ for the Mg example with $n = 100$. ($\mu_Y = 24.32, \sigma_Y = .67$) A computer calculates a 12.24% probability that $\bar{Y} < 24.25$, or $P(\bar{Y} < 24.25) = .1224$. Use the CLT to calculate this probability.

Solution. By the CLT,

$$\bar{Y} \sim N(\mu = 24.32, \frac{\sigma_Y}{\sqrt{n}} = \frac{.67}{\sqrt{100}} = .067).$$

Then,

$$P(\bar{Y} < 24.25) = P \left( \frac{\bar{Y} - \mu_Y}{\sigma_Y/\sqrt{n}} < \frac{24.25 - \mu_Y}{\sigma_Y/\sqrt{n}} \right) = P(Z < -1.045) = .1492$$

Example 2. Same setup as the previous example. The computer calculates that

$$P(\bar{Y} > 24.36) = .2409.$$
Solution.

\[ P(\bar{Y} > 24.36) = P(Z > \frac{24.36 - 24.32}{.067}) \]
\[ = P(Z > \frac{24.36 - 24.32}{.067}) \]
\[ = P(Z > .597) \]
\[ = 1 - P(Z \leq .597) \]
\[ = 1 - .7247 \]
\[ = .275 \]

\[ \square \]

Rule of Thumb: In most situations, the CLT can be used if \( n \geq 30 \). This is not a theoretical result but rather practical advice. Certainly there are cases where \( n \) must be much larger than 30 in order to trust the CLT.

Example 3. Suppose the number of minutes it takes you to bike to campus follows a normal distribution with \( \mu = 10 \) min and \( \sigma = 1.1 \) min. Find the probability that the total time spent biking to campus over a 5 day period exceeds 60 min.

Solution. Let \( Y_1, ..., Y_5 \) be the random variables giving the length of the ride for each of the 5 days. We need to find

\[ P(Y_1 + ... + Y_5 > 60) = P\left( \frac{Y_1 + ... + Y_5}{5} > \frac{60}{5} \right) \]
\[ = P(\bar{Y} > 12) \]
\[ = P\left( \frac{\bar{Y} - 10}{1.1/\sqrt{5}} > \frac{12 - 10}{1.1/\sqrt{5}} \right) \]
\[ = P(Z > 4.06) \]
\[ = 0 \]

\[ \square \]

Example 4. Find the probability that it takes more than 12 minutes to ride to campus on a given day.
Solution. Consider just day 1, or $Y_1$.

$$P(Y_1 > 12) = P\left(\frac{Y - 10}{1.1} > \frac{12 - 10}{1.1}\right)$$

$$= P\left(\frac{Y - 10}{1.1} > 1.8\right)$$

$$= P\left(Z > 1.8\right)$$

$$= P\left(Z < -1.8\right)$$

$$= .035$$

\[\square\]

Note 2. Averages are more “protected” than single observations against extreme behavior.

5.4: Normal Approximation to the Binomial

Setup: A binomial random variable $Y$ takes the values $0, ..., n$ with probability of success $p$. $Y$ is the number of successes in $n$ trials.

Consider $\hat{p} = Y/n$. This new random variable can be viewed as the proportion of successes in $n$ trials. We could think of

$$\hat{p} = \frac{W_1 + ... + W_n}{n}$$

if $W_i$ were the random variable corresponding to success on the $i$th binomial trial, or following a binomial distribution with 1 trial and success probability $p$. In this way, $\hat{p}$ is an average (think $\bar{W}$) and thus subject to the Central Limit Theorem.

We have the following two approximations

- $\hat{p} \sim N(p, \frac{\sqrt{p(1-p)}}{\sqrt{n}})$
- $Y \sim N(np, \sqrt{np(1-p)})$

Note 3 (Rule of Thumb). The normal approximation to the binomial is relatively trustworthy if

$$np \geq 5$$

and

$$n(1-p) \geq 5.$$  

Example 5. Let $Y$ be a binomial distribution with $n = 50$ and $p = .4$. Calculate $P(Y \leq 10)$ both according to the binomial probability distribution and the CLT approximation.
Solution. The exact binomial solution:

\[
P(Y \leq 10) = P(Y = 0) + P(Y = 1) + \ldots + P(Y = 10) \]
\[
= \binom{50}{0} p^0 (1 - p)^{50} + \binom{50}{1} p^1 (1 - p)^{49} + \ldots + \binom{50}{10} p^{10} (1 - p)^{40} = .002
\]

The CLT approximation:

\[
P(Y \leq 10) = P \left( \frac{Y - np}{\sqrt{np(1 - p)}} \leq \frac{10 - np}{\sqrt{np(1 - p)}} \right)
\]
\[
= P(Z \leq -2.88)
\]
\[
= .0019
\]