Recitation Assignment due 2-9-10 in recitation.
Lecture 2-8-10 will be based on this assignment as will part of 2-10-10.
Part of lecture 2-10-10 will transition to chapter 19.

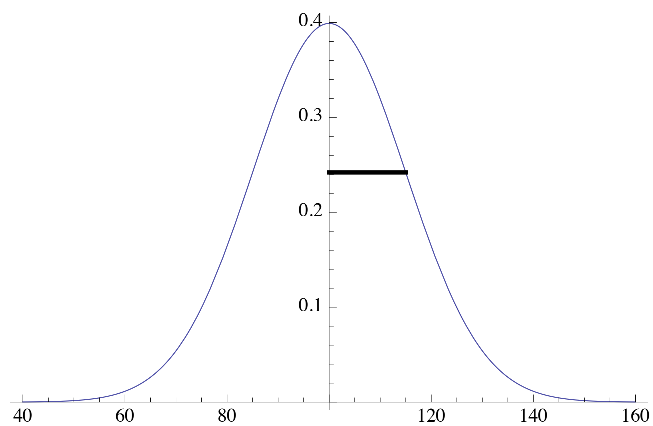
Central Limit theorems. The ***Central Limit Theorem*** of probability and statistics comes in a simple version as well as more general and subtle versions. Each of these theorems gives conditions under which the distribution of a sum (or an average) of random variables is approximately normal (bell curve).

What are normal distributions? Normal distributions are not mere bell shaped forms but are mathematically defined ideals towards which the distributions of sums of randomness progress if appropriate conditions are met.

Definition of a normal distribution having mean "mu" (written μ) and standard deviation "sigma" (written σ):

(A) $$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ for all values $-\infty < x < \infty$.

Here is the plot of p(x) for an IQ distribution having μ = 100 and σ = 15. I've plotted it for x between 100-60 and 100+60, which is 4σ either side of μ = 100:



1. In the above plot of the idealized normal distribution for IQ identify μ and σ and label them with their Greek symbols as well as their numerical values.

2. Regardless of the mean μ and sd σ, the area under (A) is one. If random variable X has the normal distribution with some mean μ and sd $\sigma > 0$ then the random variable $Z = \frac{X-\mu}{\sigma}$ (called the "standard score" of X) has the normal distribution with mean 0 and standard deviation 1. So "all normal distributions are alike in standard deviation units from the mean." The main relations are:

$$z = \frac{x - \mu}{\sigma} \quad \text{and} \quad x = \mu + \sigma\, z.$$

A way to think is "z-score of x is the number of σ units (pos or neg) by which x is removed from μ."

a. If IQ is normal distributed with mean 100 and sd 15 what is the standard score z of a person whose IQ is x = 115?

b. If IQ is normal distributed with mean 100 and sd 15 what is the standard score z of a person whose IQ is x = 100?

c. If IQ is normal distributed with mean 100 and sd 15 what is the standard score z of a person whose IQ is x = 130?

d. If IQ is normal distributed with mean 100 and sd 15 what is the IQ score x of a person whose standard score of IQ is z = 1?

e. If IQ is normal distributed with mean 100 and sd 15 what is the IQ score x of a person whose standard score of IQ is z = -2?

3. Properties expressed in #2 have the consequence that if random variable X has a normal distribution with mean μ and sd $\sigma > 0$ then for every interval [a, b] we have:

P(X falls in [a, b]) = P(Z defined by $\frac{X - \mu}{\sigma}$ falls in $[\frac{a - \mu}{\sigma}, \frac{b - \mu}{\sigma}]$).

For example,

P(IQ falls in [100, 115]) = P(Z falls in [0, 1]) = 0.34.

We soon will learn to use the "z-table" of areas captured by the standard normal Z. For now just concentrate on the interplay between X, Z and the "rule of thumb" of 68% and 95%.

a. P(IQ > 130) = P(Z >   ) ~

b. P(85 < IQ < 115) = P(< Z <   ) ~

c. P($|$ IQ - 100 $|$ < 15) = P($|$ Z $|$ <   ) ~

d. P(IQ <) = P(Z < 0) =

e. P(< IQ <) = P(-2 < Z < 1) =

f. We will find in the "z-table" that P(Z > 3.09) ~ 0.001. So
P(IQ >) = P(Z > 3.09) ~

g. P(| IQ - 100 | <) = P(| Z | < 3.09) ~

h. P(Z = 1.57889) = 0 since it cannot exceed the area under the normal curve for any interval [1.57889 - δ, 1.57889 + δ] with $\delta > 0$. These areas shrink to zero as δ is reduced towards zero. Likewise for every probability distribution on the real line whose probabilities are described by areas under a continuous curve we have the unusual consequence that every possible value x has zero probability. However, intervals having positive area between them and the curve will have positive probability. Sketch what is going on.

4. In chapter 18 there is a discussion of the normal approximation of the Binomial:

X Binomial n, p ~ Normal with μ = np and $\sigma = \sqrt{\text{npq}}$

provided both of np and nq are at least 10.

Keep in mind that X counts the random "number of successes" in independent trials (such as the random number X of Democrat voters found in a random sample of n = 400 voters). In the voting setup, as with many others, we are really interested in the sample fraction $\hat{p} = \frac{X}{n}$. For instance, if a total of x = 180 vote Democrat in a random sample of n = 400 then the **sample fraction** of Democrat voters is calculated $\hat{p} = \frac{X}{n} = \frac{180}{400} = 0.45$. By the properties of E and sd dividing by the constant n in changing from X to $\hat{p}$ produces a normal distribution for $\hat{p}$ with:

$$E\ \hat{p} = \frac{E\,X}{n} = \frac{n\,p}{n} = p \qquad \text{and} \qquad \text{sd}\ \hat{p} = \frac{\text{sd}\,X}{n} = \frac{\sqrt{\text{npq}}}{n} = \sqrt{\frac{\text{pq}}{n}}\ .$$

Let's play "what if." Let us suppose that a population of voters might have fraction p = 0.53 who favor a ballot proposal. If so, and we sample n = 400 voters at random (with-replacement), what is the approximate distribution of $\hat{p}$? It will be normal with mean μ = p = 0.53 and sd $\sigma = \sqrt{\frac{0.53 \times 0.47}{400}}$.

a. Sketch the distribution just above identifying the resulting mean and sd as recognizable elements of your sketch. Verify that the "at least 10" conditions are met.

b. Give the 95% interval for $\hat{p}$ in (a).

c. $P(\hat{p} > 0.607111) \sim P\left(Z > \frac{0.607111 - \mu}{\sigma}\right)$

$$= P\left(Z > \frac{0.607111 - p}{\sqrt{\frac{pq}{n}}}\right) = P\left(Z > \frac{0.607111 - 0.53}{\sqrt{\frac{0.53\ 0.47}{400}}}\right) =$$

5. Similar to #4.

a. Sketch the approximate normal distribution of $\hat{p}$ = the sample fraction of Democrat voters in a random (with replacement) sample of 30 voters from a population in which the fraction p = 0.55 are Democrat. Identify the mean and sd as recognizable elements of your sketch. Verify that the "at least 10" conditions are met.

b. Give the 95% interval for $\hat{p}$ in (a).

c. $P(\hat{p} > 0.731659) \sim P(Z > \qquad) =$

d. Probability we call the election wrong when p = 0.55 and n = 30 is

$P(\hat{p} < 0.5) \sim P(Z < \qquad)$

(requires z-table to evaluate, we will do it later on)