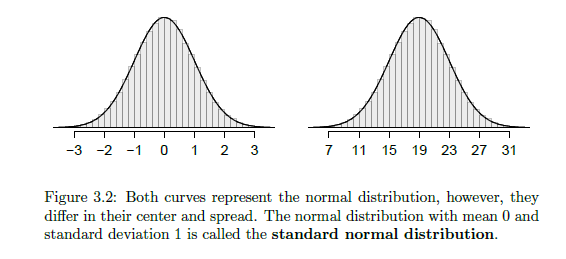
Coursera Stats Ch 3 3.1, 3.2, and 3.4

3.1 – The Normal Distribution

Looks like a bell curve, adjustable by 2 parameters, the mean and the standard deviation. The mean shifts the center to the left or right, the standard deviation stretches or shrinks the curvature of the distribution



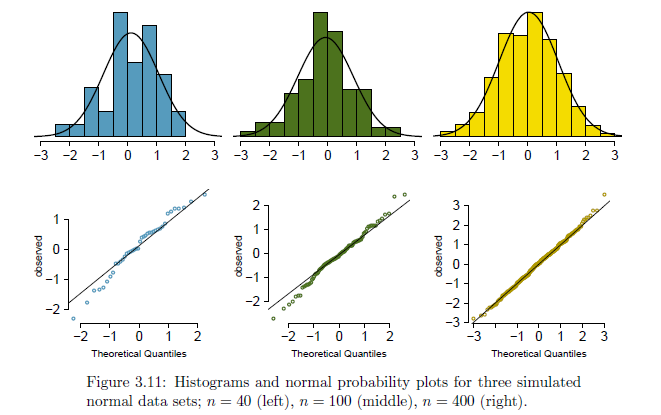
Formally described as N(µ, σ), or the normal distribution with mean µ and standard deviation σ.

**Z-score:**  the z-score of an observation is defined as the number of standard deviations it falls above or below the mean. For example, If the observation is one standard deviation above the mean, it has a Z-score of 1. An observation *x’*s Z score can be found in general by:  
 Z(x) =

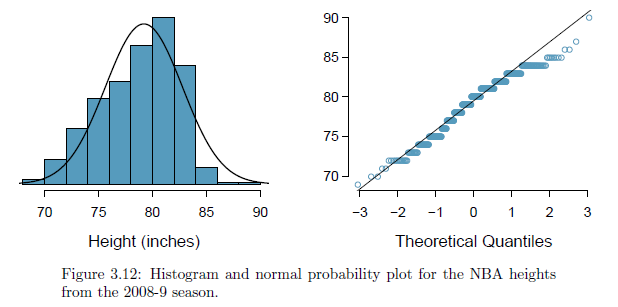
**Evaluating the Normal Curve**

Two visiual methods of measuring a set of data to see if the normal distribution is a good fit:

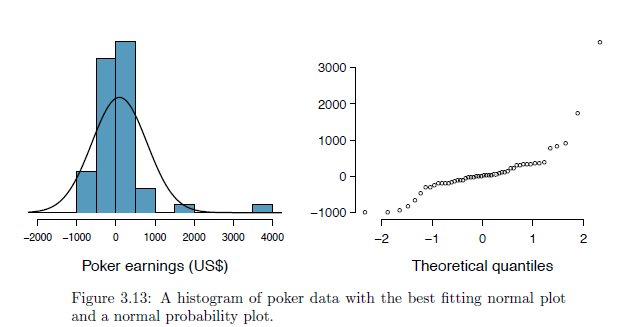
* Sort data into a histogram and overlay a normal distribution plot
  + The sample mean and the standard deviation are used as parameters for the best fitting normal plot
* Examine the normal probability plot (aka the quantile-quantile plot). If it trends towards the straight line , then the normal distribution fits well



If a dataset is left skewed instead of normal, the normal probability plot appears to have an S wave resembling the population distribution



If a dataset is right skewed instead of normal, the normal probability plot appears to have a skew to the upper right of the normal distribution plot



**3.4 – Binomial Distribution**

**Binomial Distribution**: describes the probability of having exactly *k* successes in *n* independent trials with a probability of success *p*.

Formal: Suppose the probability of a single trial being a success is *p*. Then the probability of observing exactly *k* successes in *n* independent trials is given by

Where the mean, variance, and standard deviation of the number of observed successes are:

Conditions for Binomial Distribution

1. The trials are independent
2. The number of trials *n* are fixed
3. Each trial outcome can be classified as a success or failure
4. The probability of success *p* is the same for each trial

**Normal Approximation to Binomial Distribution**

The binomial distribution with probability of success *p* is nearly normal when the sample size *n* is sufficiently large that *np* and *n*(1-*p*) are both at least 10. The approximate normal distribution has parameters corresponding to the mean and standard deviation of the binomial distribution

**Note: the normal approximation breaks down on small intervals:** The normal approximation to the binomial distribution tends to perform poorly when estimating the probability of a small range of counts, even when conditions are met. E.g. If the general population has about 20% smokers, observing 69, 70, or 71 smokers in a population of 400 people will not give a normal approximation close to the true binomial value.

**Improving the normal approximation over small intervals:** The normal approximation can be improved if the ovserved range is expanded to include a range 0.5 below the request and 0.5 above the request. E.g. If the general population has about 20% smokers and you want to observe the probability that 69, 70, or 71 smokers are observed in a population of 400, review the normal distribution for 68.5 to 71.5 instead of 69-70. This adjustment does not work over larger intervals or tail area approximations.