Distributions

LECTURE 6

Objectives

- Define terms.
- Describe and distinguish between Poisson, binomial, and normal distributions.
- Describe in qualitative and quantitative terms a normal distribution.
- Describe the different types of deviations from normality (skew and kurtosis).
- Calculate proportions of a normal curve using Z-scores.

Objectives

- Explain the central limit theorem.
- Interpret the standard error of a sample and calculate it.

Frequency distributions

- Empirical distributions
- Conceptual distributions

Binomial distribution

- Discrete probability distribution – counts
- There a number (n) of observations
  - a priori
- The observations are independent.
- An observation is classified as a “success” or a “failure.”
- The probability of a success, p, is the same for each observation.

Binomial distribution

- A binomial distribution has two parameters:
  - n – number of observations
  - p – probability of success

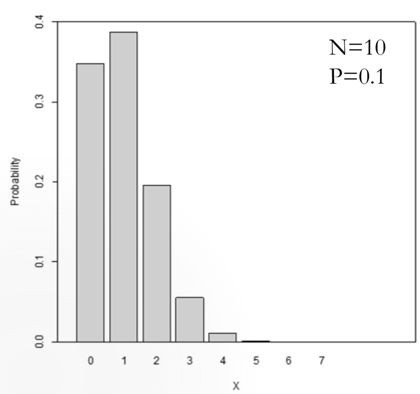
Binomial distribution

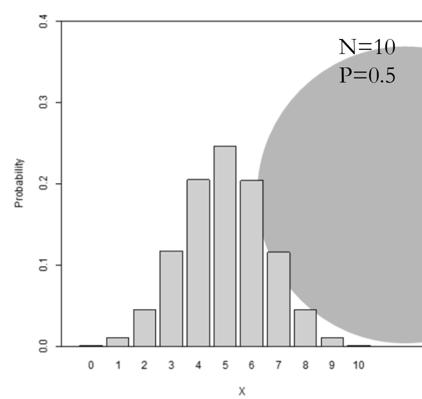
- Binomial coefficient

$$\binom{n}{X} = \frac{n!}{X!\,(n-X)!}$$

- Binomial probability

$$P(X) = \binom{n}{X} p^X (1-p)^{n-X}$$





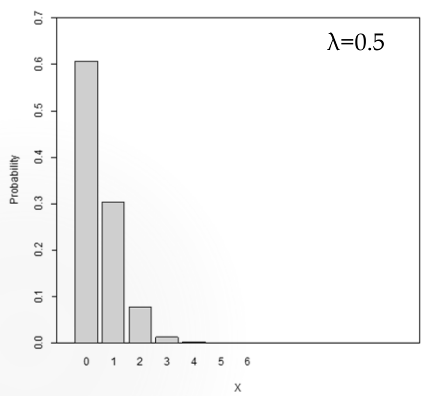
Poisson distribution

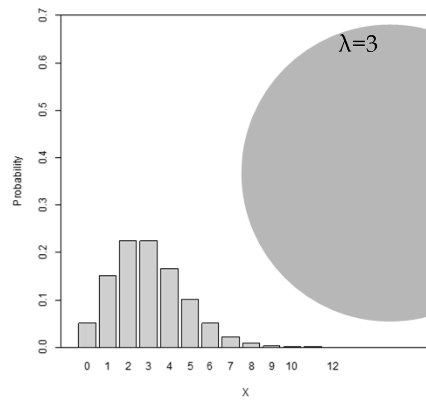
- Discrete probability distribution – counts
- Describes the count of X occurrences of a defined event in fixed, finite intervals of time or space when:
  - Occurrences are independent.
  - The probability of an occurrence is the same over all intervals.

Poisson distribution

- Poisson Probability
  - μ -- Mean number of occurrences per interval
  - X – Number of occurrences in an interval

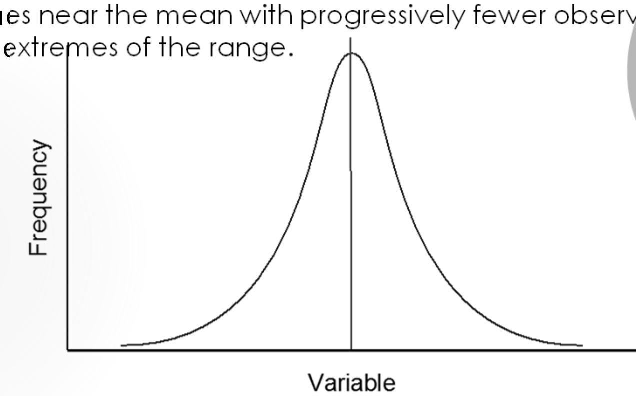
$$P(X) = \frac{e^{-\mu}\mu^{X}}{X!}$$





Bell-shaped curves

- A distribution of interval or ratio scale data observed to have most of its values near the mean with progressively fewer observations near the extremes of the range.



Normal distribution

- Not all bell-shaped curves are normal distributions
- Continuous probability distribution.
- A normal distribution is one in which the curve at X_i is expressed by the relation :

$$Y_i = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(X_i-\mu)^2}{2\sigma^2}}$$

Normal distribution

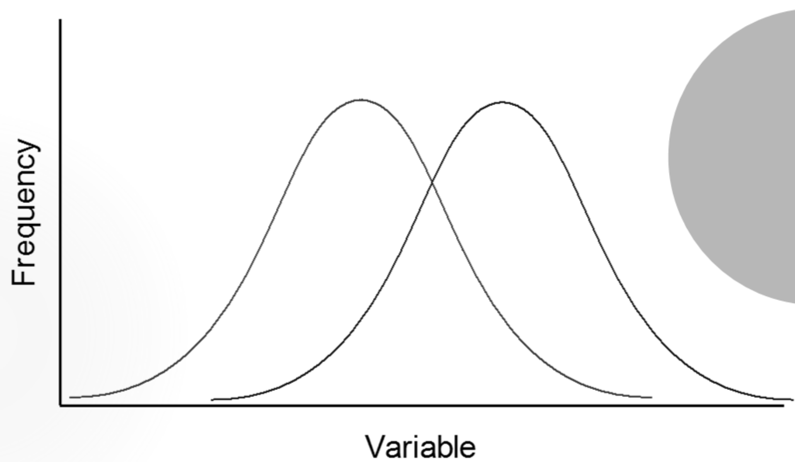
$$Y_i = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(X_i-\mu)^2}{2\sigma^2}}$$

- Notice that the normal curve is completely described by two parameters: the mean and the standard deviation.

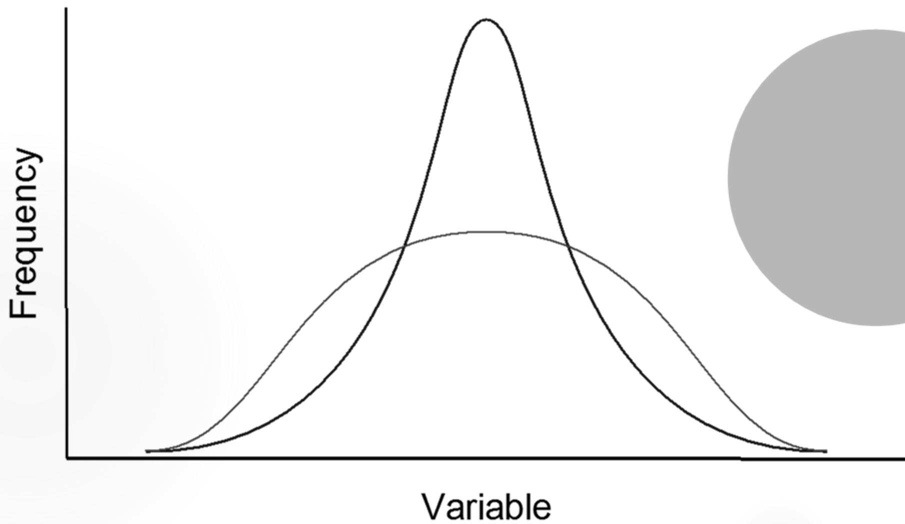
Normal distribution

- The position of the curve is determined by the mean.
- The shape, to a large degree, is determined by the standard deviation.

Normal distribution



Normal distribution



Normal distribution

- The “standard normal curve”
  - $\mu = 0$
  - $\sigma = 1$
- Equation for normal distributions simplifies to:

$$Y_i = \frac{1}{\sqrt{2\pi}} e^{\frac{-X_i^2}{2}}$$

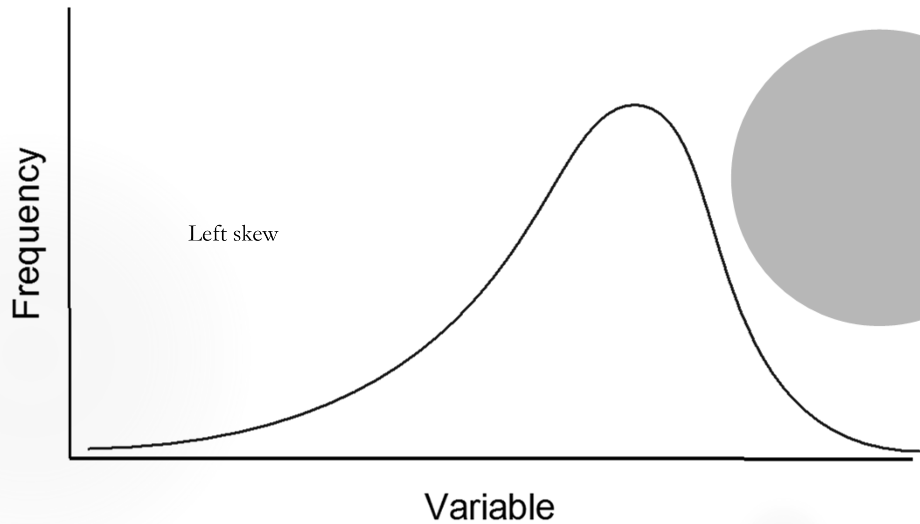
Normal distribution

- The normal curve is "THE standard" in statistics because many large sets of data rather closely approximate a normal curve.
  - Not all data is normally distributed.
  - Much biological data is not normally distributed.
    - Despite what many people say.

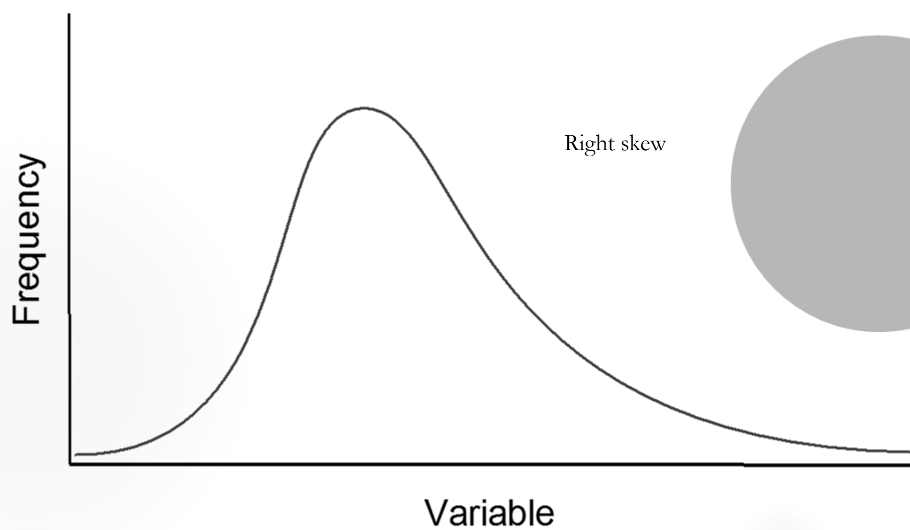
Symmetry and kurtosis

- Distributions can depart from normal in two important ways.
  - Symmetry – the distribution may not be symmetrical around the mean we call this type of distribution "skewed."
  - Kurtosis – "peakedness" – a distribution may be too "pointy" (leptokurtic) or may be too flat (platykurtic) to be a normal curve.

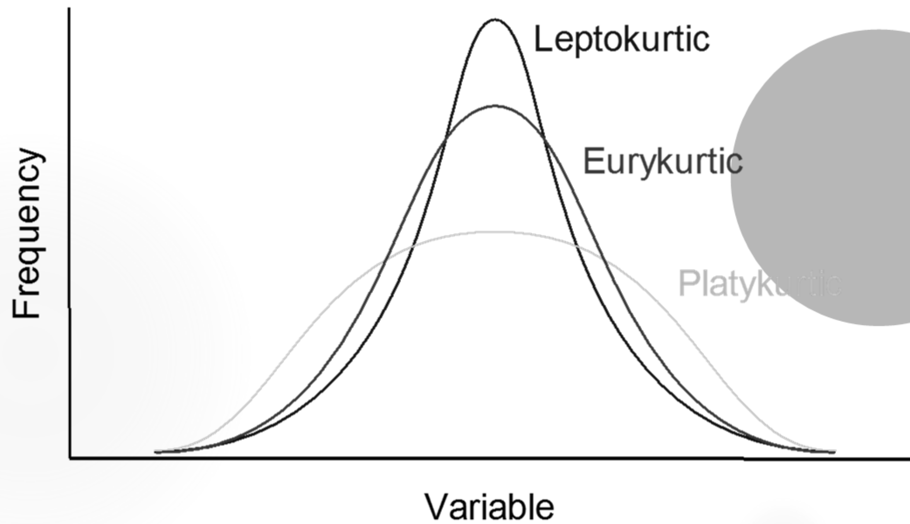
Symmetry



Symmetry



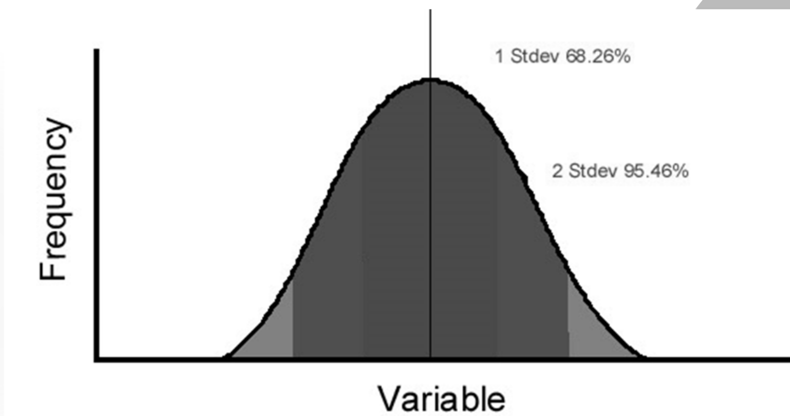
Kurtosis



Proportions of a normal distribution

- Since standard deviations are in the same units as the mean, the standard deviation can be represented as a distance from the mean ($\mu \pm \sigma$).
- Frequently normal curves are divided into different areas according to the standard deviations.

Proportions of a normal distribution



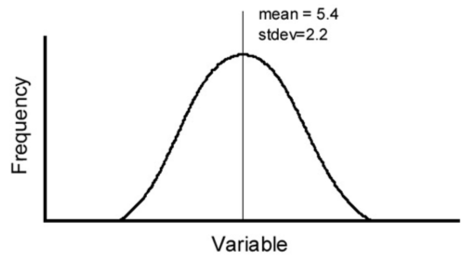
Z-score

- The number of standard deviations a particular measurement is above or below the mean (standard deviation units).

$$Z = \frac{X_i - \mu}{\sigma}$$

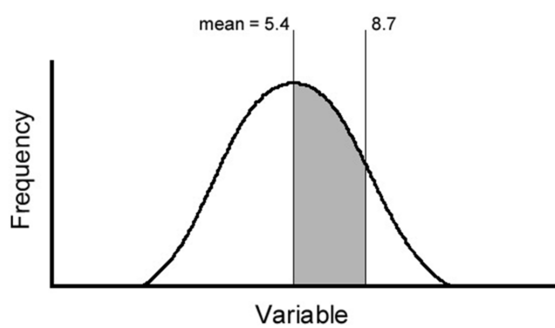
Example

- Data: Normal Distribution
  - $\mu = 5.4$
  - $\sigma = 2.2$



Example

- What proportion of the distribution is between the mean and 8.7?



Example

$$Z = \frac{X_i - \mu}{\sigma}$$

$$Z = \frac{X_i - \mu}{\sigma} = \frac{8.7 - 5.4}{2.2} = 1.50$$

- Go to Table A

Example

- What proportion of the distribution is between the mean and 8.7?
- 43.32% of the distribution lies between the mean and 8.7.

Distribution of means

- If you take random samples from a normal population, then calculate the mean for each sample the means of the samples will be normally distributed.

Distribution of means

- If you take random samples from a non-normal population, then calculate the mean for each sample – the means of the samples will not be normally distributed.

Central Limit Theorem

- However, as the size of samples increases the distribution of means tends toward normality.
- Furthermore as the size of samples increases the variance of the distribution of the means will tend to decrease.

Distribution of means

- The variance of the population of all possible means of samples of size n from a population with a variance σ^2 is

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

Distribution of means

- The standard deviation derived from this variance is called the standard error.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Distribution of means

- Samples work similarly.

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$