

PROBABILITY, RANDOM VARIABLES, AND SAMPLING DISTRIBUTIONS

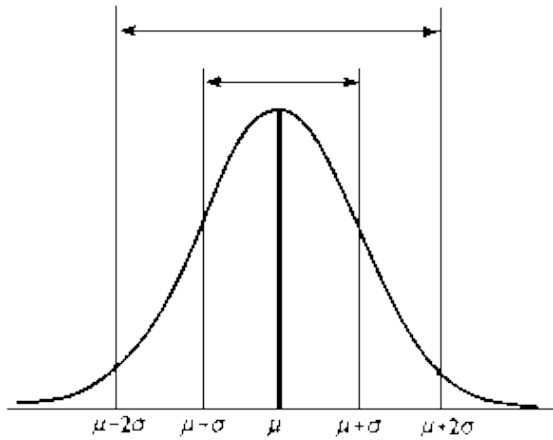
Chapter 6. THE NORMAL DISTRIBUTION

6.1 Introducing Normally Distributed Variables

The distributions of some variables including aptitude-tests scores, heights of women/men, have roughly the shape of a normal curve (bell shaped curve)

Normally Distributed Variable

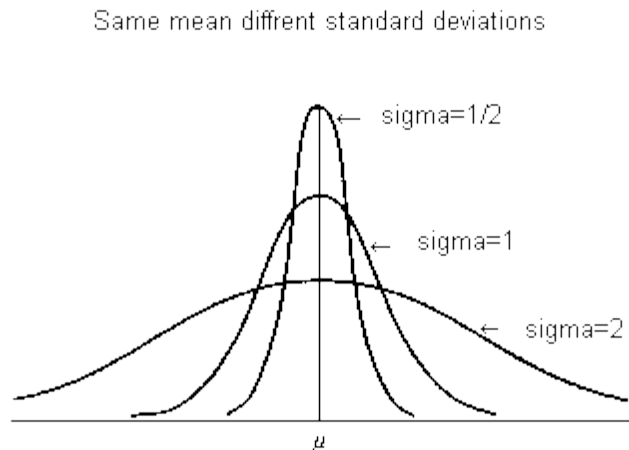
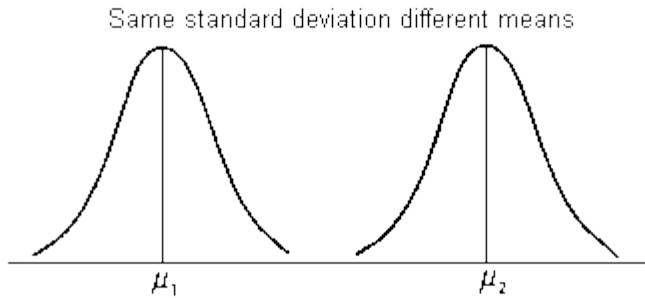
A variable is said to be **normally distributed** or to have a **normal distribution** if its distribution has the shape of a normal curve.



- Normal distribution (curve) completely determined by mean (μ) and standard deviation (σ).
- **Parameters** of Normal distribution = (μ, σ)

Characteristics of Normal distribution

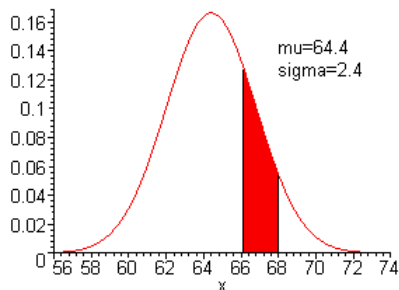
- Bell-shaped
- Symmetric around the mean μ
- Close to the horizontal axis outside the range from $\mu-3\sigma$ to $\mu+3\sigma$
- Spread depends on the standard deviation σ .
- Area under the curve is 1 for any (μ, σ) .



Normally distributed variables and normal-curve areas

For a normally distributed variable, the percentage of all possible observations that lie within any specified range equals the corresponding area under its associated normal curve expressed as a percentage. This holds true approximately for a variable that is approximately normally distributed.

Example (Heights of Female of College Students): A college has an enrollment of 3264 female students. Records show that the mean height of these students is 64.4 inches and the standard deviation is 2.4 inches. Since the shape of the relative histogram of this sample college students approximately normally distributed, we assume the total population distribution of the height of all the female college students follows the normal distribution with the same mean and the standard deviation. Now if you want to find out the percentage of students whose heights are between 66 and 68 inches, you have to evaluate the area under the normal curve from 66 to 68.



$$\text{Area} = \int_{66}^{68} \frac{1}{\sqrt{2\pi}(2.4)^2} e^{-\frac{(x-64.4)^2}{2(2.4)^2}} dx = 0.1846 \text{ (by TABLE)}$$

$$\text{Relative frequency} = 0.1100 + 0.0735 = 0.1835 \text{ (by relative frequency distribution)}$$

Standardizing a Normally Distributed Variable

Facts:

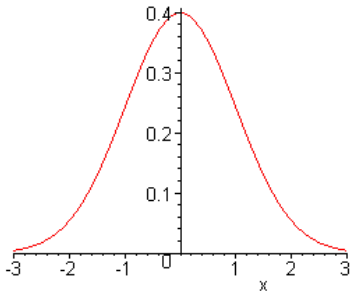
- 1) Once we know the mean and standard deviation of a normally distributed variable, we know its distribution and associated normal curve
- 2) Percentages for a normally distributed variable are equal to areas under its associated normal curve.

How do we find areas under a normal curve?

Integration? Or tables for each different μ and σ ? Or standardize your normal curve and use only one table with mean(μ)=0 and standard deviation(σ)=1?

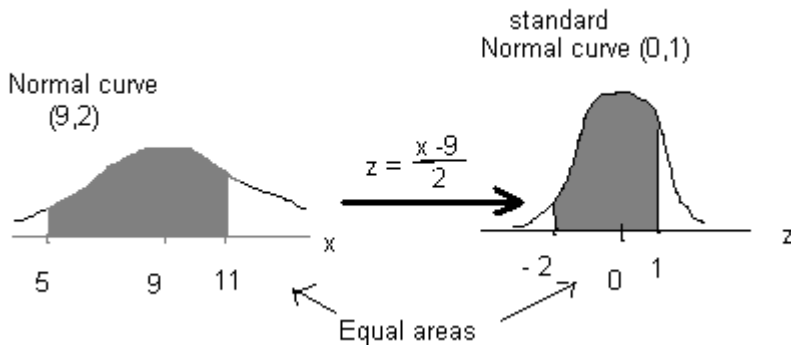
Standard Normal Distribution; Standard Normal Curve

A normally distributed variable having mean 0 and standard deviation 1 is said to have the **standard normal distribution**. Its associated normal curve is called the **standard normal curve**.



Standardized Normally Distributed Variable

The standardized version of a normally distributed variable x , $z = \frac{x - \mu}{\sigma}$ has the standard normal distribution.



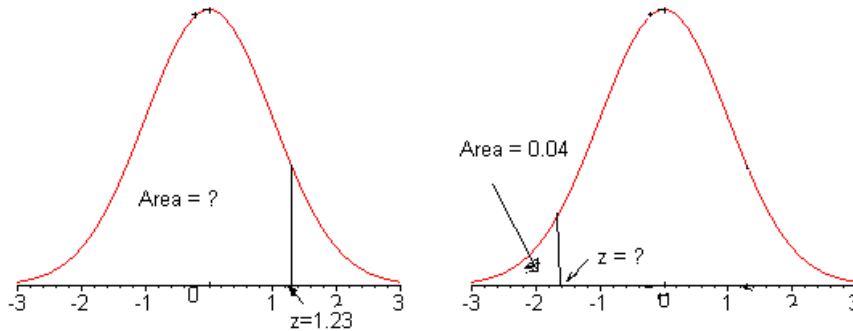
6.2 Areas under the Standard Normal Curve

Basic Properties of the Standard Normal Curve

1. The total area under the standard normal curve is equal to 1.
2. The standard normal curve extends indefinitely in both directions, approaching, but never touching, the horizontal axis as it does so.
3. The standard normal curve is symmetric about 0; i.e., the part of the curve to the left of 0 is the mirror image of the part of the curve to the right of 0.
4. Most of the area under the standard normal curve lies between -3 and 3 .

Using the Standard-Normal Table

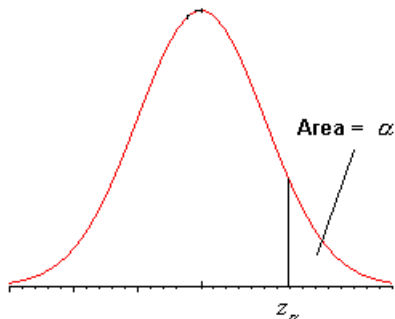
There are infinitely many normally distributed variables, however, if these variables can be standardized, then the standard normal tables can be used to find the areas under the curve.



- * Table set up to accumulate the area under the curve from $-\infty$ to and specified value.
- * The table starts at -3.9 and goes to 3.9 since outside this range of values the area is negligible.
- * The table can be used to find a z value given an area, or an area given a z value.

The z_α Notation

The symbol z_α is used to denote the z -score having area α (alpha) to its right under the standard normal curve. z_α - z sub alpha or simply $z \alpha$.



6.3 Working with Normally Distributed Variables

To Determine a Percentage or Probability for a normally Distributed Variable

1. Sketch the normal curve associated with the variable.
2. Shade the region of interest and mark the delimiting x-values.
3. Compute the z-scores for the delimiting x-values found in step 2.
4. Use Table II to obtain the area under the standard normal curve delimited by the z-scores found in step 3.

Example (contd.) Height of Female students: Normal distribution with $\mu = 64.4$, $\sigma = 2.4$.
Want to determine the probability between 66 and 68.

z-score for $x = 66$: $z = (66-64.4)/2.4 = 0.67$, $x=68$: $z = (68-64.4)/2.4 = 1.5$

area under standard normal curve: $z=1.5 \rightarrow 0.9332$, $z = 0.67 \rightarrow 0.7486$

resulting probability: $0.9332 - 0.7486 = 0.1846$

Visualizing a Normal Distribution

1. 68.26% of all possible observations lie within one standard deviation to either side of the mean, i.e., between $\mu - \sigma$ and $\mu + \sigma$.
2. 95.44% of all possible observations lie within two standard deviations to either side of the mean, i.e., between $\mu - 2\sigma$ and $\mu + 2\sigma$.
3. 99.74% of all possible observations lie within three standard deviations to either side of the mean, i.e., between $\mu - 3\sigma$ and $\mu + 3\sigma$.

To Determine the Observations Corresponding to a specified Percentage or Probability for a Normally Distributed Variable.

1. Sketch the normal curve associated with the variable.
2. Shade the region of interest.
3. Use Table II to obtain the z-scores delimiting the region in step 2.
4. Obtain the x-values having the z-scores found in step 3:
$$x = \mu + z(\sigma)$$

Example (contd.)

a. Obtain the Q_3 (75th percentile) of the height of female students.

The z-score corresponding to Q_3 is the one having an area of 0.75 to its left under the standard normal curve. From Table II, that z-score is 0.67, approximately.

So the x-value (height) corresponding to that z-score is $64.4 + (0.67)*2.4 = 66$ inches.

b. Obtain the 10th percentile.

z-score corresponding to P_{10} is the one having an area of 0.1 to its left under the standard normal curve. From Table II, that z-score is -1.28 , approximately. So the x-value (height) corresponding to that z-score is $64.4 + (-1.28)*2.4 = 61.32$.