# UCLA STAT XL 10

## Introduction to Statistical Reasoning


- **Instructor**: **Ivo Dinov**, Asst. Prof. in Statistics and Neurology

University of California, Los Angeles, Spring 2002

*http://www.stat.ucla.edu/~dinov/*




## Chapter 5

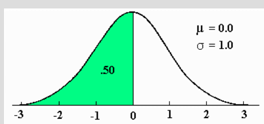
**Normal Distribution**



## Standard Normal Curve

- The standard normal curve is described by the equation:

$$y = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$




Where remember, the natural number e ~ 2.7182…

We say: X~Normal(μ, σ), or simply X~N(μ, σ)

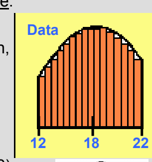
AdditionalInstructorAids/NormalCurveInteractive.html

AdditionalInstructorAids/QuincunxApplet.html

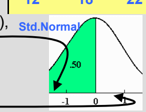


## Standard Normal Curve

- Many histograms are similar in shape to the standard normal curve, provided they are drawn in the same (*density*) scale.
- A value is converted to standard units by calculating how many standard deviations is it above or below the average.
- Example, assume we have observations, whose (partial) density-scale histogram is as shown, come from a process with mean value of 18 and standard deviation of 5. Compute the limit values (12 and 22) to standard units.



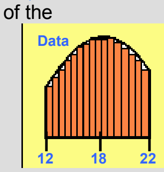

- 12 is (12-18)/5 = -6/5 = -1.2SD below the mean (18), and hence 12 orig.units → -1.2 std.units
- 22 is (22-18)/5 = 4/5 = 0.8SD above the mean (18), and hence 18 orig.units → +0.8 std.units



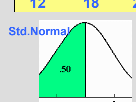



## Standard Normal Curve

- In general, the transformation X → (X-μ)/σ, standardizes the observed value X, where μ and σ are the average and the standard deviation of the distribution X is drawn from.



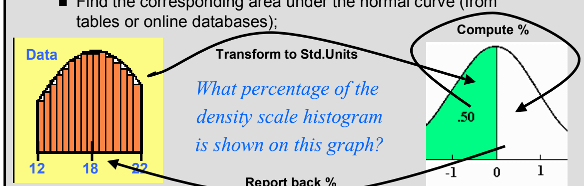

- 12 is (12-18)/5 = -6/5 = -1.2SD below the mean (18), and hence
  12 orig.units → -1.2 std.units
- 22 is (22-18)/5 = 4/5 = 0.8SD above the mean (18), and hence
  18 orig.units → +0.8 std.units





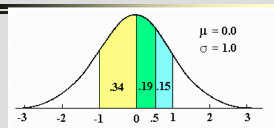

## Standard Normal Approximation

- The standard normal curve can be used to estimate the percentage of entries in an interval for any process. Here is the protocol for this approximation:
  - Convert the interval (we need the assess the percentage of entries in) to standard units. We saw the algorithm already.
  - Find the corresponding area under the normal curve (from tables or online databases);



# Areas under Standard Normal Curve

- Protocol for calculating any area under the standard normal curve:
  - Sketch the normal curve and shade the area of interest
  - Separate your area into individually computable sections
  - Check the Normal Table and extract the areas of every sub-section
  - Add/compute the areas of all sub-sections to get the total area.



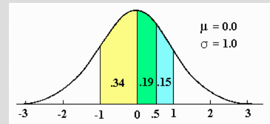

- Example: compute area on the interval [0.5 : 1.0]
  - Draw curve
  - Blue region is the area we need
  - See table values (p. A-105)
  - Compute final result

Area under the Normal curve on [-z : z]

| Z | Height | Area |
|---|---|---|
| 0.50 | 35.21 | 38.29 |
| 1.0 | 24.20 | 68.27 |



# Areas under Standard Normal Curve




- Example: compute area on the interval [0.5 : 1.0]
  - Draw curve
  - Blue region is the area we need
  - Table values (p. A-105)
  - Compute final result

Area under the Normal curve on [-z : z]

| Z | Height | Area |
|---|---|---|
| 0.50 | 35.21 | 38.29 |
| 1.0 | 24.20 | 68.27 |

- So far so good, but how do we really compute the area under [0.5 : 1.0] ?
  - Compute final result
    - [-0.5 : 0.5] → 38.29
    - ½ [-0.5 : 0.5] → 19.14
    - [-1.0 : 1.0] → 68.27
    - ½ [-1.0 : 1.0] → 34.13
    - ½ [-1.0 : 1.0] - ½ [-0.5 : 0.5] = 34.13 - 19.14 = 14.99 ~15

**Note there are more than one strategies to compute the correct area. Try to think of other area separations which compute the same area!**

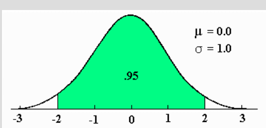


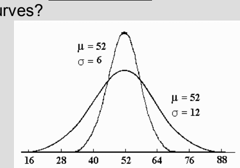
# General Normal Curve

- The general normal curve is defined by:
  - Where $\mu$ is the average of (the symmetric) normal curve, and $\sigma$ is the standard deviation (spread of the distribution).

$$y = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

  - Why worry about a standard and general normal curves?
  - How to convert between the two curves?



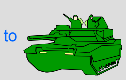






# Areas under Standard Normal Curve

- Many histograms are similar in shape to the standard normal curve. For example, persons height. The height of all incoming female army recruits is measured for custom training and assignment purposes (e.g., very tall people are inappropriate for constricted space positions, and very short people may be disadvantages in certain other situations). The mean height is computed to be 64 *in* and the standard deviation is 2 *in*. Only recruits shorter than 65.5 in will be trained for tank operation and recruits within ½ standard deviations of the mean will have no restrictions on duties.

  - What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?



  - About what percentage of the recruits will have no restrictions on training/duties?

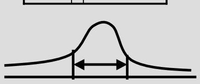


# Standard Normal Curve – Table differences

- There are different tables and computer packages for representing the area under the standard normal curve. But the results are always interchangeable.

Area under Normal curve on [-z : z]

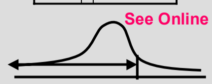
| Z | Area |
|---|---|
| 0.50 | 38.29 |
| 1.0 | 68.27 |




Area under Normal curve on [-infinity : z]

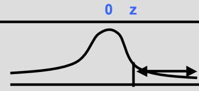
| Z | Area |
|---|---|
| 0.50 | 69.15 |
| 1.0 | 84.13 |

**See Online tables!**




Area under Normal curve on [z: infinity]

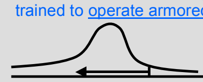
| Z | Area |
|---|---|
| 0.50 | 30.85 |
| 1.0 | 15.87 |





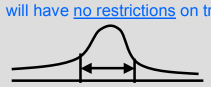

# Areas under Standard Normal Curve

- The mean height is 64 *in* and the standard deviation is 2 *in*.
  - Only recruits shorter than 65.5 in will be trained for tank operation. What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?




**X → (X-64)/2**
**65.5 → (65.5-64)/2 = ¾**
**Percentage is 77.34%**

  - Recruits within ½ standard deviations of the mean will have no restrictions on duties. About what percentage of the recruits will have no restrictions on training/duties?




**X → (X-64)/2**
**65 → (65-64)/2 = ½**
**63 → (63-64)/2 = -½**
**Percentage is 38.30%**

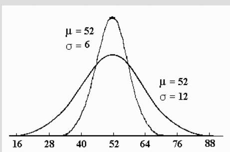
# Review

- Estimating sample mean from raw data and from the frequency table:

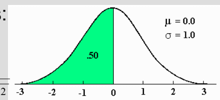
  SampleMean = (1/N)Sum(RawNumericObservations)

  SampleMean = (1/N)Sum(value x frequency)

Standard and general Normal curves:




$$y = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

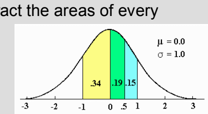



$$y = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$



# Areas under Standard Normal Curve – Normal Approximation

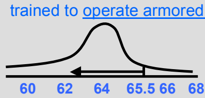
- Protocol:
  - Convert the interval (we need to assess the percentage of entries in) to Standard units. Actually convert the end points in Standard units.
    - In general, the transformation X → (X-μ)/σ, standardizes the observed value X, where μ and σ are the average and the standard deviation of the distribution X is drawn from.
  - Find the corresponding area under the normal curve (from tables or online databases);
    - Sketch the normal curve and shade the area of interest
    - Separate your area into individually computable sections
    - Check the Normal Table and extract the areas of every sub-section
    - Add/compute the areas of all sub-sections to get the total area.






# Areas under Standard Normal Curve – Normal Approximation

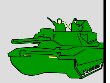
- The mean height is 64 *in* and the standard deviation is 2 *in*.
  - Only recruits shorter than 65.5 in will be trained for tank operation. What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?




**X → (X-64)/2**

**65.5 → (65.5-64)/2 = ¾**

**Percentage is 77.34%**

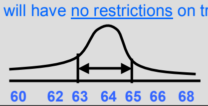


  - Recruits within ½ standard deviations of the mean will have no restrictions on duties. About what percentage of the recruits will have no restrictions on training/duties?

**X → (X-64)/2**

**65 → (65-64)/2 = ½**

**63 → (63-64)/2 = -½**



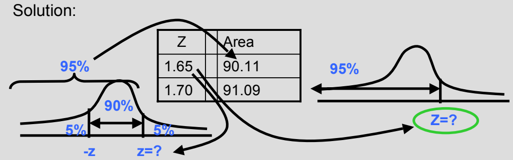

**Percentage is 38.30%**



# Percentiles for Standard Normal Curve

- When the histogram of the observed process follows the normal curve Normal Tables (of any type, as described before) may be used to estimate percentiles. The N-th percentile of a distribution is P is N% of the population observations are less than or equal to P.

- Example, suppose the Math-part SAT scores of newly admitted freshmen at UCLA averaged 535 (out of [200:800]) and the SD was 100. Estimate the 95 percentile for the score distribution.
- Solution:

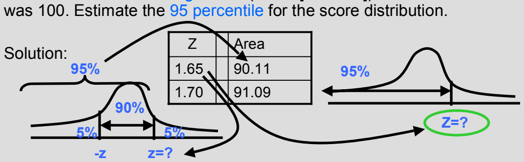



| Z | Area |
|---|---|
| 1.65 | 90.11 |
| 1.70 | 91.09 |



# Percentiles for Standard Normal Curve

- Example, suppose the Math-part SAT scores of newly admitted freshmen at UCLA averaged 535 (out of [200:800]) and the SD was 100. Estimate the 95 percentile for the score distribution.
- Solution:




| Z | Area |
|---|---|
| 1.65 | 90.11 |
| 1.70 | 91.09 |

- Z=1.65 (std. Units) → 700 (data units), since

  X → (X – μ)/σ, converts data to standard units and

  X → σ X + μ, converts standard to data units!

  σ = 100; μ =535, 100 x 1.65 + 535 = 700.



# Summary

1. The Standard Normal curve is symmetric w.r.t. the origin (0,0) and the total area under the curve is 100% (1 unit)
2. Std units indicate how many SD's is a value below (-)/above (+) the mean
3. Many histograms have roughly the shape of the normal curve (bell-shape)
4. If a list of numbers follows the normal curve the percentage of entries falling within each interval is estimated by: 1. Converting the interval to StdUnits and 2. Computing the corresponding area under the normal curve (Normal approximation)
5. A histogram which follows the normal curve may be reconstructed just from ($\mu,\sigma^2$), mean and variance=std_dev$^2$
6. Any histogram can be summarized using percentiles
7. E(aX+b)=aE(X)+b, Var(aX+b)=a$^2$Var(X), where E(Y) the the mean of Y and Var(Y) is the square of the StdDev(Y),

# Example – work out in your notebooks

1. Compute the chance a random observation from a distribution (symmetric, bell-shaped, unimodal) with m=75 and SD=12 falls within the range [53 : 71].
2. (53-75)/12 = -11/6=-1.83 Std unit
3. (71-75)/12=-0.333(3) Std units
4. Area[53:71] =
5. (SN_area[-1.83:1.83] –SN_area[-0.33:0.33])/2
6. = (93% - 25%)/2 = 34%
7. Compute the 90$^{th}$ percentile for the same data:
8. b+a+b=100% } a=80% → A=0.8
9. a+b=90% } b=10% Z=1.3 SU
10. 90% P = $\sigma$1.3 + $\mu$ =12x1.3+75=90.6

**Check Work**
**Should it be**
**<50% or >50%?**