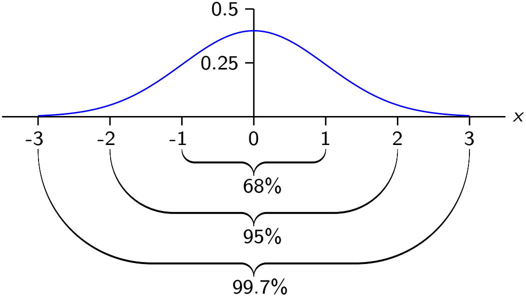
Math 207: Introduction to Statistics

The Normal Curve and Normal Approximation to Data



Dr. Ralph Wojtowicz

Mathematics Department

SHENANDOAH UNIVERSITY

1. The Normal Curve
   - The Normal Curve
   - Standard Units

2. Finding Areas Under the Normal Curve
   - Finding Areas Under the Normal Curve with R
   - Finding Areas Under the Normal Curve

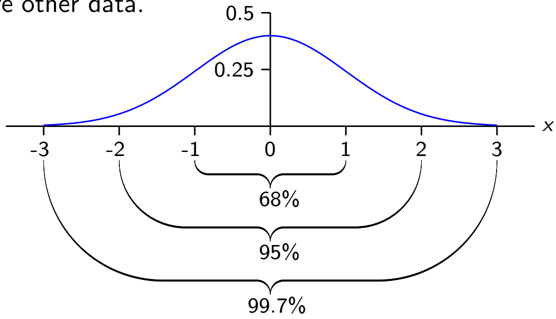
3. The Normal Approximation for Data
   - The Normal Approximation for Data
   - Non-Gaussian Data

4. Percentiles
   - Percentiles

5. Change of Scale
   - Change of Scale

The Normal Curve

- The **standard normal** (or Gaussian) curve is an ideal histogram to which we will compare other data.



- The total area under the curve is 1 (or 100%).
- The curve is symmetric about the line $x = 0$.
- It has mean $= 0$ and SD $= 1$.

Standard Units

- If $x_1, \ldots, x_n$ is a list of numbers, we convert the values in the list to **standard units** using the following formula:

$$z_i = \frac{x_i - \text{mean}}{\text{SD}}$$

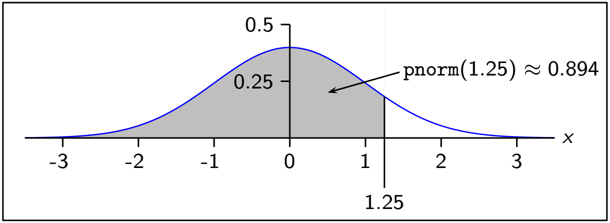
- z_i measures how far (in units of SD) x_i is from the mean (average) of the list
- Example: Consider the list 13, 9, 10, 11, 7.

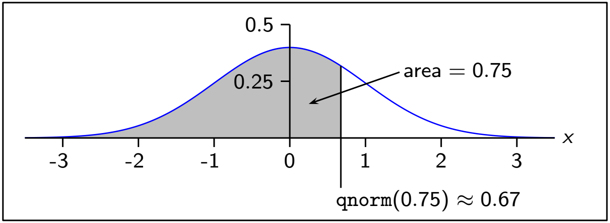
```
> x <- c(13, 9, 10, 11, 7)
> mean(x)
[1] 10
> sd(x) * sqrt(4/5)
[1] 2
```

  Now convert 13 to standard units:

```
> (13-10)/2
[1] 1.5
```

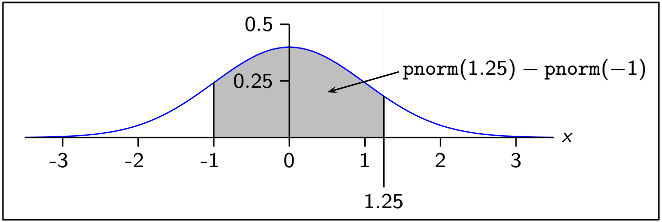
Finding Areas Under the Normal Curve with R





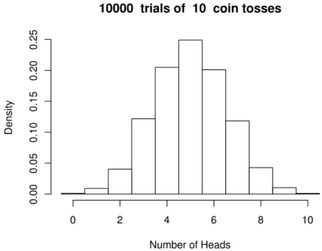
Finding Areas Under the Normal Curve

- Use one or more of the following to find areas under the normal curve:
  - Total area under the curve is 1 (that is, 100%)
  - The area is symmetric about vertical the line $x = 0$
  - (area to the left of x) $=$ (1 $-$ area to the right of x)
  - The 68%, 95%, 99.7% rules (see slide 2)
  - The `pnorm` or `qnorm` functions in `R`
  - Normal table in your text
- Example:



The Normal Approximation for Data

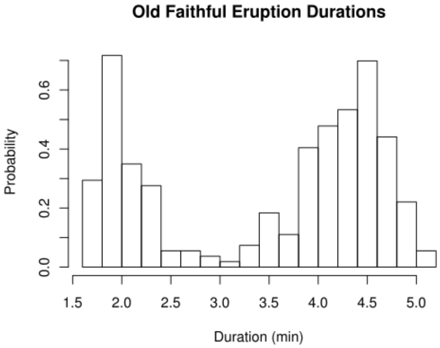
- For the normal curve,
  - 68% of the area is within 1 SD of the mean
  - 95% of the area is within 2 SDs of the mean
  - 99.7% of the area is within 3 SDs of the mean
- The same area vs. SD rule roughly holds for histograms generated by many other data sets.
- From another perspective, for many data sets, if we convert the data to standard units, the histogram will look a lot like the normal curve.
- Example:



Non-Gaussian Data

- Not all data follows the normal curve:

```
> hist(faithful$eruptions, breaks=20,
  main="Old Faithful Eruption Durations",
  prob=TRUE, xlab="Duration (min)", ylab="Probability")
```



Percentiles

- For data that does not follow the normal curve, we use percentiles, quartiles and related descriptive statistics to summarize the data.
- The 25^{th} percentile is a number x for which 25% of the data is less than x
- The 50^{th} percentile is a number x for which 50% of the data is less than x (this is the same as the median)
- The 75^{th} percentile is a number x for which 75% of the data is less than x
- Use the `quantile`, `summary` or `fivenum` functions in R to find quartiles.

```
> x <- c(3, 6, 7, 8, 8, 10, 13, 15, 16, 20, 22)
> quantile(x)

  0%  25%   50%   75%   100%
 3.0  7.5  10.0  15.5   22.0

> summary(x)
  Min  1st Qu. Median   Mean  3rd Qu.   Max
 3.00    7.50   10.00  11.64   15.50  22.00
```

Change of Scale

- If you add a number A to each element of a list, then
  - Mean of new list $= A +$ mean of old list
  - SD of new list $=$ SD of old list
- If you multiply each element of a list by a number A, then
  - Mean of the new list $= A \times$ mean of old list
  - SD of new list $= |A| \times$ SD of old list
- Example: The mean of the list 1, 3, 4, 5, 7 is 4 and the SD is 2.
  The mean of the list 101, 103, 104, 105, 107 is 104 and the SD is 2.
  The mean of the list -2, -6, -8, -10, -14 is -8 and the SD is 4.
- Example: For the 100 measurements on Table 1 on p. 99 of our text mean $= 0.000405$ grams and SD $= 0.000006$ grams. In ounces, the mean is 0.0000142 and the SD is 2×10^{-7}. (1 gm $= 0.0352739619$ oz)
- Example: The mean of a list of temperature data is 0°C and the SD is 5°C. This corresponds to a mean of 32°F and SD of 9°F. $\left(\mathrm{F} = \frac{9}{5}\,\mathrm{C} + 32\right)$