Areas under the Normal curve
Percehiles

Probability
- quantifying chance
- frequency definition of probability
- drawing with + without replacement
  - conditional probability

Multiplication rule
- independent events

counting.
Normal curve.

Mean.

SD = 1

Standard units.

to convert to
standard units

\[
\frac{\text{data} - \text{mean}}{\text{SD}}
\]

Symmetric interval.

68%.

-3  -2  -1  0  1  2  3.
non symmetric

tail area.

\[ \text{Area from } -1 \text{ to } 1 \text{ is } 68\% \]

\[ \Rightarrow \text{ area from } 0 \Rightarrow 1 \text{ is } \frac{1}{2} \times 68 = 34\% \]
\[ \frac{1}{2} \times 95\% + \frac{1}{2} \times 68\% = 82\% . \]
two tails together are 100 - A

one tail has area

\[ \frac{100 - A}{2} \]

\[ A \quad 84\% \]

\[ -1.4 \quad 1.4 \]

\[ \text{two tails together} \quad 100 - 84 \]

\[ \text{one tail} \quad \frac{1}{2} (100 - 84) \]

\[ = 8\% \]

**Normal Approximation for Data**

Group of men had mean height 69 in.
SD height 3 in.

What % had heights between 63 and 72 in?
Assuming the data follows the Normal Curve.

63 in standard units is \( \frac{63 - 69}{3} = -2 \).

72 in standard units is \( \frac{72 - 69}{3} = +1 \).

\[
\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 0.95 = 0.82\%
\]
## Tables

A NORMAL TABLE

<table>
<thead>
<tr>
<th>$z$</th>
<th>Height</th>
<th>Area</th>
<th>$z$</th>
<th>Height</th>
<th>Area</th>
<th>$z$</th>
<th>Height</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>39.89</td>
<td>0</td>
<td>1.50</td>
<td>12.95</td>
<td>86.64</td>
<td>3.00</td>
<td>0.443</td>
<td>99.730</td>
</tr>
<tr>
<td>0.05</td>
<td>39.84</td>
<td>3.99</td>
<td>1.55</td>
<td>12.00</td>
<td>87.89</td>
<td>3.05</td>
<td>0.381</td>
<td>99.771</td>
</tr>
<tr>
<td>0.10</td>
<td>39.69</td>
<td>7.97</td>
<td>1.60</td>
<td>11.09</td>
<td>89.04</td>
<td>3.10</td>
<td>0.327</td>
<td>99.806</td>
</tr>
<tr>
<td>0.15</td>
<td>39.45</td>
<td>11.92</td>
<td>1.65</td>
<td>10.23</td>
<td>90.11</td>
<td>3.15</td>
<td>0.279</td>
<td>99.837</td>
</tr>
<tr>
<td>0.20</td>
<td>39.10</td>
<td>15.85</td>
<td>1.70</td>
<td>9.40</td>
<td>91.09</td>
<td>3.20</td>
<td>0.238</td>
<td>99.863</td>
</tr>
<tr>
<td>0.25</td>
<td>38.67</td>
<td>19.74</td>
<td>1.75</td>
<td>8.63</td>
<td>91.99</td>
<td>3.25</td>
<td>0.203</td>
<td>99.885</td>
</tr>
<tr>
<td>0.30</td>
<td>38.14</td>
<td>23.58</td>
<td>1.80</td>
<td>7.90</td>
<td>92.81</td>
<td>3.30</td>
<td>0.172</td>
<td>99.903</td>
</tr>
<tr>
<td>0.35</td>
<td>37.52</td>
<td>27.37</td>
<td>1.85</td>
<td>7.21</td>
<td>93.57</td>
<td>3.35</td>
<td>0.146</td>
<td>99.919</td>
</tr>
<tr>
<td>0.40</td>
<td>36.83</td>
<td>31.08</td>
<td>1.90</td>
<td>6.56</td>
<td>94.26</td>
<td>3.40</td>
<td>0.123</td>
<td>99.933</td>
</tr>
<tr>
<td>0.45</td>
<td>36.05</td>
<td>34.73</td>
<td>1.95</td>
<td>5.96</td>
<td>94.88</td>
<td>3.45</td>
<td>0.104</td>
<td>99.944</td>
</tr>
<tr>
<td>0.50</td>
<td>35.21</td>
<td>38.29</td>
<td>2.00</td>
<td>5.40</td>
<td>95.45</td>
<td>3.50</td>
<td>0.087</td>
<td>99.953</td>
</tr>
<tr>
<td>0.55</td>
<td>34.29</td>
<td>41.77</td>
<td>2.05</td>
<td>4.88</td>
<td>95.96</td>
<td>3.55</td>
<td>0.073</td>
<td>99.961</td>
</tr>
<tr>
<td>0.60</td>
<td>33.32</td>
<td>45.15</td>
<td>2.10</td>
<td>4.40</td>
<td>96.43</td>
<td>3.60</td>
<td>0.061</td>
<td>99.968</td>
</tr>
<tr>
<td>0.65</td>
<td>32.30</td>
<td>48.43</td>
<td>2.15</td>
<td>3.96</td>
<td>96.84</td>
<td>3.65</td>
<td>0.051</td>
<td>99.974</td>
</tr>
<tr>
<td>0.70</td>
<td>31.23</td>
<td>51.61</td>
<td>2.20</td>
<td>3.55</td>
<td>97.22</td>
<td>3.70</td>
<td>0.042</td>
<td>99.978</td>
</tr>
<tr>
<td>0.75</td>
<td>30.11</td>
<td>54.67</td>
<td>2.25</td>
<td>3.17</td>
<td>97.56</td>
<td>3.75</td>
<td>0.035</td>
<td>99.982</td>
</tr>
<tr>
<td>0.80</td>
<td>28.97</td>
<td>57.63</td>
<td>2.30</td>
<td>2.83</td>
<td>97.86</td>
<td>3.80</td>
<td>0.029</td>
<td>99.986</td>
</tr>
<tr>
<td>0.85</td>
<td>27.80</td>
<td>60.47</td>
<td>2.35</td>
<td>2.52</td>
<td>98.12</td>
<td>3.85</td>
<td>0.024</td>
<td>99.988</td>
</tr>
<tr>
<td>0.90</td>
<td>26.61</td>
<td>63.19</td>
<td>2.40</td>
<td>2.24</td>
<td>98.36</td>
<td>3.90</td>
<td>0.020</td>
<td>99.990</td>
</tr>
<tr>
<td>0.95</td>
<td>25.41</td>
<td>65.79</td>
<td>2.45</td>
<td>1.98</td>
<td>98.57</td>
<td>3.95</td>
<td>0.016</td>
<td>99.992</td>
</tr>
<tr>
<td>1.00</td>
<td>24.20</td>
<td>68.27</td>
<td>2.50</td>
<td>1.75</td>
<td>98.76</td>
<td>4.00</td>
<td>0.013</td>
<td>99.9937</td>
</tr>
<tr>
<td>1.05</td>
<td>22.99</td>
<td>70.63</td>
<td>2.55</td>
<td>1.54</td>
<td>98.92</td>
<td>4.05</td>
<td>0.011</td>
<td>99.9949</td>
</tr>
<tr>
<td>1.10</td>
<td>21.79</td>
<td>72.87</td>
<td>2.60</td>
<td>1.36</td>
<td>99.07</td>
<td>4.10</td>
<td>0.009</td>
<td>99.9959</td>
</tr>
<tr>
<td>1.15</td>
<td>20.59</td>
<td>74.99</td>
<td>2.65</td>
<td>1.19</td>
<td>99.20</td>
<td>4.15</td>
<td>0.007</td>
<td>99.9967</td>
</tr>
<tr>
<td>1.20</td>
<td>19.42</td>
<td>76.99</td>
<td>2.70</td>
<td>1.04</td>
<td>99.31</td>
<td>4.20</td>
<td>0.006</td>
<td>99.9973</td>
</tr>
<tr>
<td>1.25</td>
<td>18.26</td>
<td>78.87</td>
<td>2.75</td>
<td>0.91</td>
<td>99.40</td>
<td>4.25</td>
<td>0.005</td>
<td>99.9979</td>
</tr>
<tr>
<td>1.30</td>
<td>17.14</td>
<td>80.64</td>
<td>2.80</td>
<td>0.79</td>
<td>99.49</td>
<td>4.30</td>
<td>0.004</td>
<td>99.9983</td>
</tr>
<tr>
<td>1.35</td>
<td>16.04</td>
<td>82.30</td>
<td>2.85</td>
<td>0.69</td>
<td>99.56</td>
<td>4.35</td>
<td>0.003</td>
<td>99.9986</td>
</tr>
<tr>
<td>1.40</td>
<td>14.97</td>
<td>83.85</td>
<td>2.90</td>
<td>0.60</td>
<td>99.63</td>
<td>4.40</td>
<td>0.002</td>
<td>99.9989</td>
</tr>
<tr>
<td>1.45</td>
<td>13.94</td>
<td>85.29</td>
<td>2.95</td>
<td>0.51</td>
<td>99.68</td>
<td>4.45</td>
<td>0.002</td>
<td>99.9991</td>
</tr>
</tbody>
</table>
Standard units

how many SDs the value is above or below the mean.

63 with mean 69
SD 3. \( \underline{\text{known}} \)

63 is 6 units below mean.
6 is 2 SD

63 is \( \underline{-2} \) in standard units.

\[
\frac{\text{value} - \text{mean}}{\text{SD}} \leq \text{converts to standard units.}
\]
Percentiles.

Some data sets are not well approximated by the Normal Curve.

→ choose a data value
the % of the data up to and including that value is the percentile.

→ choose a percentile
start counting up data items in order, from the lowest first, and stop when you've seen the chosen % of the data. the last data value is the one corresponding to the percentile.
By definition, 50th %ile is the median.

25th %ile = 1st quartile,
75th %ile = 3rd quartile.

Difference between 75th %ile and 25th %ile is called inter-quartile range.

Can be a better measure of spread for long-tailed distributions.
Introduction to Probability.

Quantifying chance.
- using probabilities
- mathematics to calculate quantities associated with uncertain events.

- what do we mean by chance?
- how do we interpret probabilities?

Rolling a die - if the die is balanced, when rolling it many times, each side should come up as often as any other.

eg the side with 2 spots should come up ¼ of the time.

Frequency definition of probability
the chance of an event is the percentage of times an event is expected to happen when the process is repeated over and over, independently and under the same conditions.

Properties

If something is impossible, it happens 0% of the time, its probability is zero.

Conversely, if something always happens, it happens 100% of the time, its probability is 1 (or 100%).

Probabilities are between 0 and 1.

Chance of something happening or it not happening is 1.
chance of an event is 100% = chance of opposite event.

or

if an event has chance $p$ of happening, the opposite event has chance $1-p$

chance of rolling a 6 is $\frac{1}{6}$

chance of not rolling a 6 is $\frac{5}{6}$ ($= 1 - \frac{1}{6}$)

Example

Box contains red and blue marbles.

One is drawn at random.

If it's red - win $\$1$

blue - win $\$0$

Box 1: 3 red, 2 blue.
Box 2: 30 red, 20 blue

Both boxes give equal chance of drawing red.
What's important is the ratio

\[
\frac{\text{# of outcomes we're interested in}}{\text{total # of outcomes}}.
\]

When all outcomes are equally likely,

\[
\begin{align*}
\frac{\text{# red}}{\text{total #}} &= \frac{3}{5} \quad \text{Box 1} \\
\frac{30}{50} &= \frac{3}{5} \\ 
\text{Box 2}.
\end{align*}
\]

List all the equally likely outcomes:

\[
\begin{align*}
R & \quad 3 \quad \text{fit the condition} \quad (3). \\
R & \\
R & \\
G & \\
G & \quad \text{Divide by the total number.} \quad \frac{3}{5} = 60% \\
\end{align*}
\]
Drawing with + without replacement.

with replacement:

- draw one, note the color
- put it back
- when I draw the second
- the contents are the same

without replacement:

- imagine 1st draw is blue
- when drawing the second,
- the box contains

So the chances of drawing each ball have changed.

To deal with this, we need conditional probability.
Conditional Probability

The occurrence of one event affects the chances of another event.

- If 1st ball is blue, the chance of the 2nd being green is different than if the 1st ball was green.

- What Deal 2 cards from a shuffled deck. What is chance 2nd card is Q♥?

- Shuffling puts the cards in a random order. The Q♥ is equally likely to be in any of the 52 spots. Chance of Q♥ being in the 2nd position is \( \frac{1}{52} \).
Turn over the 1st card - it's the 5♦.

What's the chance of the 2nd card being Q♥?

- There are now only 51 available spots for Q♥, so chance is $\frac{1}{51}$.

This is the conditional probability of 2nd card being Q♥ given that 1st card was 5♦.

↑ equivalent to "conditioned on"