We have been talking about histograms.

Now, we need to talk about a few features/properties of histograms that are not only useful, but they also set the stage to doing more advanced things later (tomorrow!).

Two variations on histograms are:

Relative freq. histograms: \( \text{Rel. freq.} = \frac{\text{freq.}}{\text{total sample size}} \).

Density scale histograms: \( \text{Rel. freq.} / \text{bin size} \).

Density histograms have a nice property:

Area = proportion of something = probability of something.

E.g. area for \( a < x < b \) = prob. of \( a < x < b \).

\[ \text{E.g.: } x \text{ discrete (no bin size)} \]

\[ \text{Rel. freq.} \Rightarrow \text{Then proportion of times} \]

\[ \text{that } 2 < x \leq 3 \text{ is } 80\%. \]
### Questionnaire Results

<table>
<thead>
<tr>
<th>Question</th>
<th>Excellent</th>
<th>Very Good</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
<th>Very Poor</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>The course as a whole:</td>
<td>27%</td>
<td>33%</td>
<td>22%</td>
<td>12%</td>
<td>3%</td>
<td>3%</td>
<td>3.80</td>
</tr>
<tr>
<td>Textbook overall:</td>
<td>33%</td>
<td>30%</td>
<td>27%</td>
<td>10%</td>
<td>0%</td>
<td>0%</td>
<td>3.94</td>
</tr>
<tr>
<td>Instructor overall:</td>
<td>50%</td>
<td>28%</td>
<td>10%</td>
<td>7%</td>
<td>2%</td>
<td>3%</td>
<td>4.50</td>
</tr>
<tr>
<td>Instructor's contribution:</td>
<td>42%</td>
<td>27%</td>
<td>15%</td>
<td>8%</td>
<td>3%</td>
<td>3%</td>
<td>4.22</td>
</tr>
<tr>
<td>Instructor's interest:</td>
<td>53%</td>
<td>26%</td>
<td>7%</td>
<td>5%</td>
<td>2%</td>
<td>7%</td>
<td>4.56</td>
</tr>
<tr>
<td>Amount learned:</td>
<td>39%</td>
<td>27%</td>
<td>20%</td>
<td>8%</td>
<td>3%</td>
<td>2%</td>
<td>4.09</td>
</tr>
<tr>
<td>Relevance and usefulness of homework:</td>
<td>37%</td>
<td>17%</td>
<td>27%</td>
<td>12%</td>
<td>3%</td>
<td>3%</td>
<td>3.75</td>
</tr>
</tbody>
</table>

For median calculation: 5 = Excellent 4 = Very Good 3 = Good 2 = Fair 1 = Poor 0 = Very Poor

The numbers on each row say something about rating, i.e., how the students rated something. So, the random variable is rating.

- **Quantitative (discrete)**
- **Qualitative (categorical)**

Interpret: center ~ 3 or 4, spread ~ 1, 1.5, shape ~ skewed (to ...)

Average = \( \frac{27 + 33}{2} = 30 \% \) of students say Excellent or Very Good, etc.
These are some other shapes that you may come across.

**Bell-shaped**

**Skewed**

**Bimodal**

A histogram is a plot of frequencies (or relative frequencies, ...) of different values of one variable.

*Not* some variable as a function of time!

*Not* some variable (e.g., demand) as a function of some other variable (e.g., supply).

For example:
- $x =$ magnitude of earthquakes
- $x =$ population of cities on the planet
- $x =$ length of words in a book
- $x =$ casualties of wars, for different wars

E.g., $\log \text{freq} = m \log (x) + b$

$\log (\text{freq}) = \log (b x^m)$

Leave out all constants.
Suppose you see this histogram in an article. But the labels on the y-axis are missing. What proportion of times is x strictly positive?

(a) 100%  
(b) less than 100%  
(c) Cannot Tell because it looks exponential  
(d) Cannot Tell because it depends on bin size.

It was pointed out in class that x=0 is a bit ambiguous. So, if this were a test question, I would accept both (a) and (b).

But I'm going to discard this question because it's the first one, and there were technical difficulties.

When poll is active, respond at PollEv.com/marzban  
Text MARZBAN to 22333 once to join
For each of the following shapes, come up with at least 1 example of a quantity \( x \) (a random variable) whose histogram you expect to be approximately:

a) Bell-shaped (symmetric)

b) Skewed (one way or the other)

c) Exponential-looking

d) Bimodal

Describe the quantity clearly, and explain in words why you expect the particular shape. If you have data to support your expectation, then go ahead and show the histogram. (For this problem, \( x \) may be continuous or discrete.)

**hw. lect3-2:** In the above lecture note, there exists at least one random variable that when considered as quantitative, has an exponential-looking histogram. Identify one of them, and plot its rel. freq. hist. (By hand).

**hw. lect3-3** Suppose the histogram of \( x \) is "exponential-looking", and when we look at \( \log(\text{freq}) \), it looks linear: like this. What is the relationship between \( \text{freq.} \) and \( x \)?