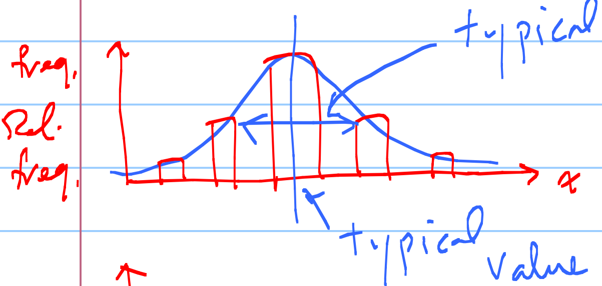
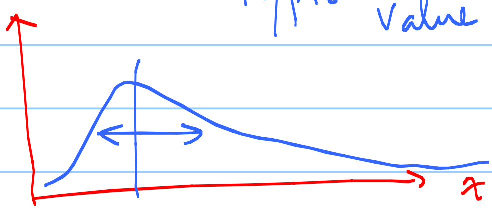
Lecture 3 (Ch. 1)

We have been talking about histograms.
These are some examples that you may come across



bell-shaped

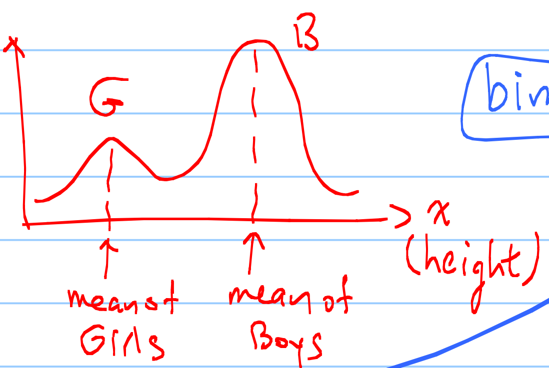


skewed to the right.

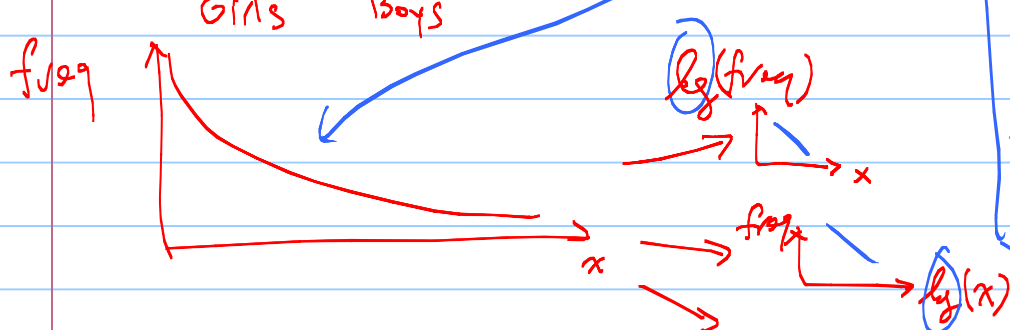
NOT some variable as a function of time!

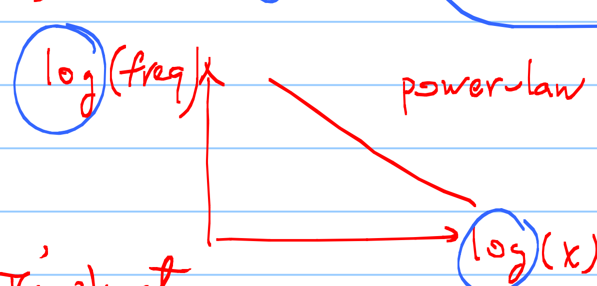
NOT some variable (e.g. demand) as a function of some other variable (e.g. supply).

A hist is a plot of freq. (or rel. freq, ...) of different values of ONE variable.



bimodal





power-law

Eg.

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

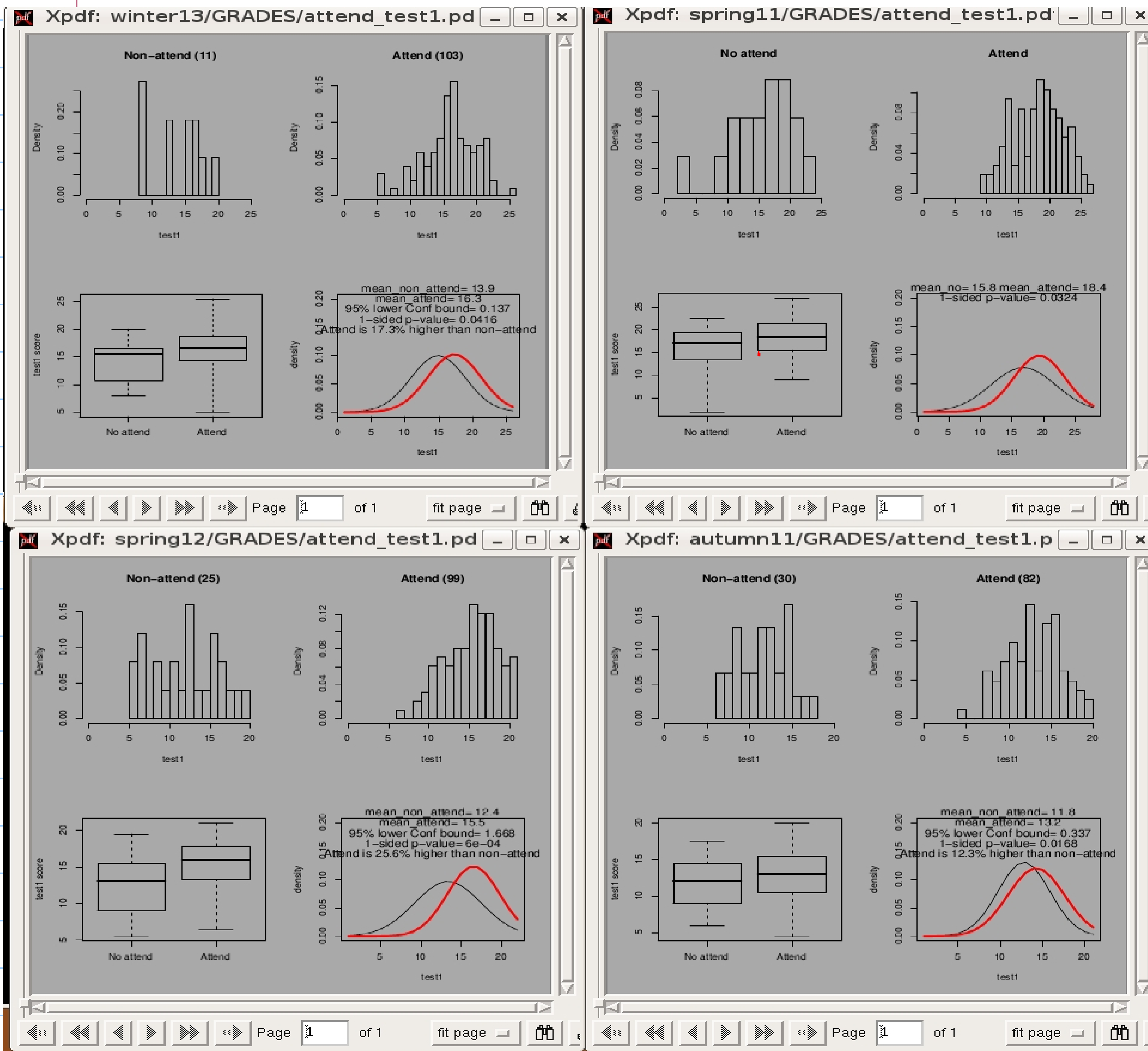
$\log(\text{freq}) = m \log(x) + b$

$= \log x^m + \log b' = \log(b' x^m)$

$\therefore \text{freq} \sim x^m$

FYI only.

Here is one more use of a histogram that should interest all of you. Just concentrate on the hists; you will learn about the rest, later.



All of this suggests that attending 390 lectures is associated with higher test grades. This is from only 4 quarters, but the same pattern exists for every quarter!

Now, a HUGE, but TRICKY, and IMPORTANT Concept

Distributions.

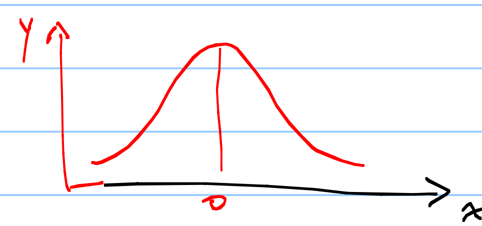
A histogram pertains to data.

But there is something else (called distribution) that looks like a histogram, but is not. In fact, distributions have nothing to do with data. So, for now, forget about data.

purely mathematical

In statistics, distributions are used to represent the population, while histograms are used to describe the sample (data). Later, we are going to learn how to tell something about the pop. (ie. distr.) from a sample (ie. histogram). But, a priori, dists and hists are completely unrelated.

Example: $y \sim f(x) \sim e^{-\frac{1}{2}x^2}$



For continuous x, $f(x)$ = density function
For discrete x $p(x)$ = mass function

generally called distribution

To be more precise,

Defn: A density function $f(x)$, or a mass function $p(x)$, must satisfy

1) $f(x) \geq 0$ || $p(x) \geq 0$

2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$ || $\sum_{\text{all } x} p(x) = 1$ (ie. $\sum_{i=1}^{n} p_i = 1$)

e.g. $p(\text{apple}) + p(\text{orange}) + \cdots = 1$

x = fruit type.

Example: $f(x) \sim e^{-\frac{1}{2}x^2}$, $-\infty < x < \infty$, is <u>not</u> a dist, because

$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\, dx =$ remind yourself how to do such integrals $= \sqrt{2\pi} \neq 1$

So, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ <u>is</u> a dist. (also $f(x) \geq 0$ ✓)

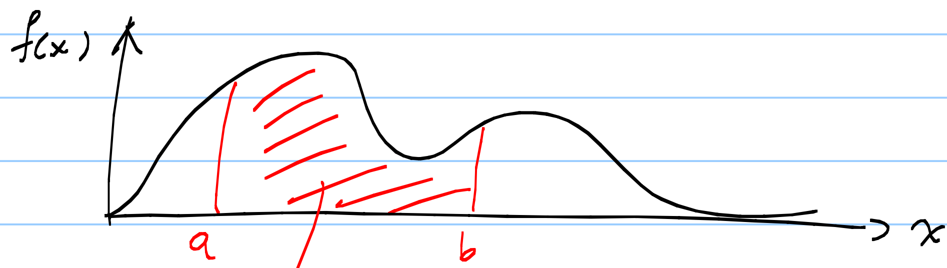
Example: $f(x) = kx^8(1-x)$, $0 < x < 1$ dist?

$\int_{-\infty}^{\infty} f(x)\, dx = \int_0^1 kx^8(1-x)\, dx = k \cdot \frac{1}{90} \neq 1$ unless $\underline{k = 90}$.

So, $f(x) = 90x^8(1-x)$ is a distr. (also $f(x) \geq 0$ ✓)

Like histograms, distributions can be used to compute the (theoretical, or mathematical, or expected) proportion of times that x falls between any 2 numbers. not from data.

E.g.



Q The proportion of x values between a, b $(a < b)$?

A $\int_a^b f(x)\, dx$ || $\sum_{x=a}^{b} p(x) = p(\text{apple}) + p(\text{kiwi}) + \dots$

$=$ area $\qquad$ $=$ area

Eg. for $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ $\qquad$ $\text{prop}(a < x < b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$

for $f(x) = 90x^8(1-x)$ $\qquad$ " $= \int_a^b 90x^8(1-x)\, dx$

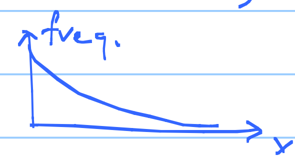
hw - lect 3

For each of The following shapes, come-up with atleast 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", ie.



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 categorical.)