Probability vs. Statistics

Statistics is "a branch of mathematics which uses formulas to collect, describe, analyze, and interpret quantitative data."

At some point, you will all have to do some data analysis, which is statistics and thus not really within the scope of this course. But if you are like most people, you will do it badly.

One book I urge you all to read is: E. Tufte's "The Visual Display of Quantitative Information" which through examples clearly presents the difference between good + bad mathematical graphics.

Let's look at some examples of each...

Chances Are...

The theory of probability lets us calculate the likelyhood of complex events.

Of all areas of mathematics I am familiar with, it is most important to have careful definitions and work things out according to them is probability, since you can easily get the wrong answer.

A Probability Space is a set Ω of all things which can happen along with a probability $Pr(\omega)$ to each elementary event $\omega \in \Omega$.

A Probability $Pr(\omega)$ is a non-negative real number such that $\sum_{\omega \in \Omega} Pr(\omega) = 1$

A Probability Distribution is a function which distributes the total probability of 1 among all the events

Ex: "A family has two kids. When you knock at the door, a boy answers. What is the probability the other child is also a boy?

The probability space is:

$$\{b, b\}, \{b, G\}, \{G, b\}$$
$$\tfrac{1}{3} \qquad \tfrac{1}{3} \qquad \tfrac{1}{3}$$

Thus the probability is $1/3$!

An event is a subset of the probability space Ω.

The probability of an event A:

$$P_r(w \in A) = \sum_{w \in A} P_r(w)$$

A Random variable is a function defined on the elementary events of a probability space.

Ex: $S(w)$ be the number of spots on a pair of dice.

$P(w)$ be the product of spots on a pair of dice.

	1	2	3	4	5	6
1	2,1	3,2	4,3	5,4	6,5	7,6
2	3,2	4,4	5,6	6,8	7,10	8,12
3	4,3	5,6	6,9	7,12	8,15	9,18
4	5,4	6,8	7,12	8,16	9,20	10,24
5	6,5	7,10	8,15	9,20	10,25	11,30
6	7,6	8,12	9,18	10,24	11,30	12,36

S, P

S	2	3	4	5	6	7	8	9	10	11	12
$P_r(S=s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

P	1	2	3	4	5	6	8	9	10	12	15	16	18	20	21	24	25	30	36
$P_r(P=p)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$		$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

X + Y are independent random variables over the same space if $P_r(X=x + Y=y) = P_r(X=x) \cdot P_r(Y=y)$ for all events x, y.

S + P are not independent: pick any x, y:

$$x = 5,\ y = 15 \rightarrow P_r(S=5 + P=15) = 0 \neq \frac{4}{36} \cdot \frac{2}{36}.$$

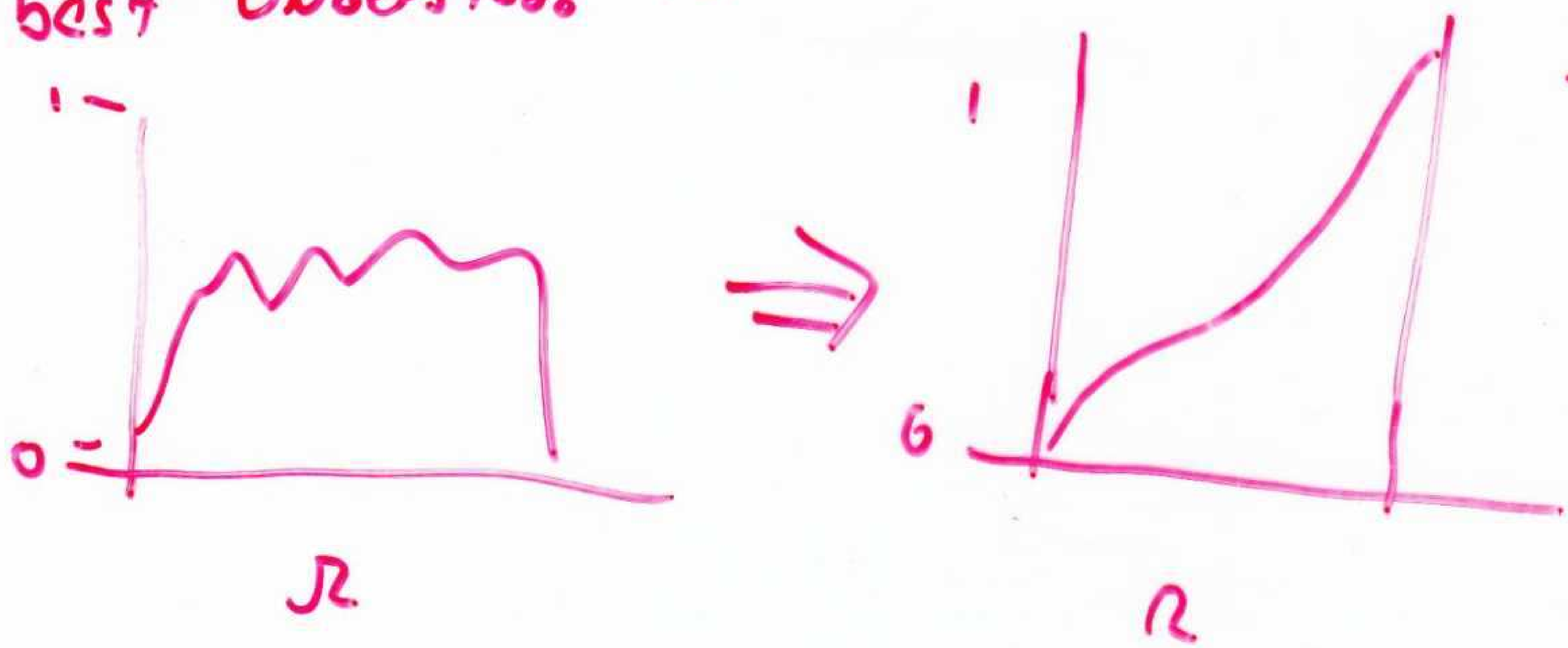
"Averages of Random Variables"

The mean of a random real-valued variable, or the expected value $E_X = \sum_{\omega \in \Omega} X(\omega) \cdot P_r(\omega)$

The median of X is the set of all x such that

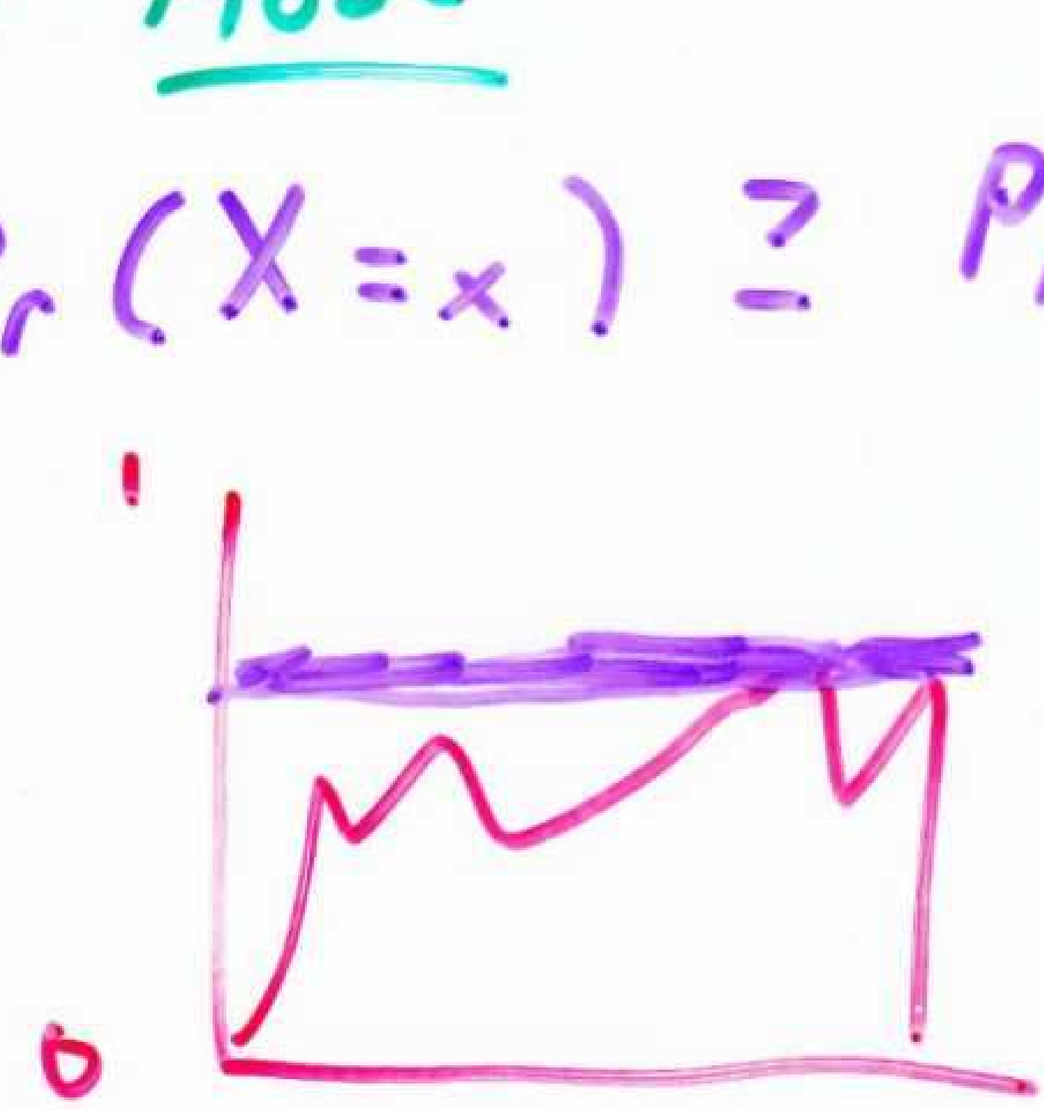
$$P_r(X \leq x) \geq 1/2 \quad \text{and} \quad P_r(X \geq x) \geq 1/2$$

best understood in terms of the cumulative distribution function



The mode of X is the set of all x such that

$$P_r(X = x) \geq P_r(X = x') \text{ for all } x' \in X(\Omega)$$



Properties of Expected Values

Since $EX = \sum_{\omega \in \Omega} X(\omega) P_r(\omega)$,

$$E(X+Y) = \sum_{\omega \in \Omega} \big(X(\omega) + Y(\omega) \big) P_r(\omega) = EX + EY$$

$$E(\alpha X) = \sum_{\omega \in \Omega} \alpha X(\omega) P_r(\omega) = \alpha EX$$

Products have a nice form $\underline{\text{IF}}$ X + Y are independent:

$$E(XY) = \sum_{\omega \in \Omega} X(\omega) Y(\omega) \cdot P_r(\omega)$$

$$= \sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega)}} xy \cdot P_r(X = x \text{ and } Y = y)$$

$$= \sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega)}} xy \; P_r(X = x) P_r(Y = y)$$

Here is where we make use of independence!

$$= \Big(\sum_{x \in X(\Omega)} P_r(x = X) \, x \Big) \Big(\sum_{y \in Y(\Omega)} P_r(y = Y) \, y \Big)$$

$$= (EX)(EY)$$

$\blacksquare$