Slide

- $>$ means ‘is greater than
- $\geq$ means ‘is greater than or equal to
- $<$ means ‘is less than
- $\leq$ means ‘is less than or equal to
- $\neq$ means ‘is not equal to
- $\approx$ or $\simeq$ means ‘is approximately equal to

A simple data set

Suppose that we have a data set with n observations. For each observation, a measure is recorded. Conventionally the measures are denoted x unless a more suitable notation is available. A subscript can be used to indicate which observation the measure is for. Hence we would write a data set as follows; $(x_1, x_2, x_3, x_1 \ldots x_n)$ (i.e. the first, second, third ... nth observation).

Summation

The summation sign $\sum$ is commonly used in most areas of statistics. Given $x_1 = 3, x_2 = 1, x_3 = 4, x_4 = 6, x_5 = 8$ find:

$$(i) \sum_{i=1}^{i=n} x_i \qquad (ii) \sum_{i=3}^{i=4} x_i^2$$

$$(i) \sum_{i=1}^{i=n} x_i = x_1 + x_2 + x_3 + x_4 + x_5$$

$$= 3 + 1 + 4 + 6 + 8 = \mathbf{22}$$

$$(ii) \sum_{i=1}^{i=n} x_i^2 = x_3^2 + x_4^2 = 9 + 16 = \mathbf{25}$$

When all elements of a data set are used, a simple version of the summation notation can be used. $\sum_{i=1}^{i=n} x_i$ can simply be written as $\sum x$

Example

Given that $p_1 = 1/4, p_2 = 1/8, p_3 = 1/8, p_4 = 1/3, p_5 = 1/6$ find:

- $\sum_{i=1}^{i=n} p_i \times x_i$
- $\sum_{i=1}^{i=n} p_1 \times x_i^2$

Joint probability tables

- A joint probability table is a table in which all possible events (or outcomes) for one variable are listed as row headings, all possible events for a second variable are listed as column headings, and the value entered in each cell of the table is the probability of each joint occurrence.
- Often, the probabilities in such a table are based on observed frequencies of occurrence for the various joint events.
- The table of joint-occurrence frequencies which can serve as the basis for constructing a joint probability table is called a contingency table.

Permutations

Example 2

youtube.com/StatsLabDublin

Twitter: @statslabdublin

Permutations

Suppose a four letter code is made from the letters $\{$**a,b,c,d,e**$\}$, where repetitions are allowed and the order of the letters in the code is significant

For example **a,a,e,c** is a different code to **a,c,e,a**.

Permutations

- Let $\mathcal{U}$ be the set of all such codes.
- Let $\mathcal{V}$ be the set of all such codes beginning with a vowel.
- Let $\mathcal{P}$ be the set of all such codes which are palindromic.

(A palindromic code is a string of letters which read the same backwards as forwards, for example **a,e,c,e,a** is a 5 letter palindromic code.)

Permutations

How many elements are there in the set $\mathcal{U}$?

(i)	(ii)	(iii)	(iv)

Permutations

How many elements are there in the set $\mathcal{V}$?

(i)	(ii)	(iii)	(iv)

Permutations

How many elements are there in the set $\mathcal{P}$?

(i)	(ii)	(iii)	(iv)

Permutations

How many elements are there in the sets $\mathcal{V}$ and $\mathcal{P}$?

(i)	(ii)	(iii)	(iv)

Empty

Probability Trees

- Two gamblers, A and B, are playing each other in a tournament to win a jackpot worth $\$6,000$.
- The first gambler to win 5 rounds, wins the tournament, and the jackpot outright.
- Each player has an equal chance of winning each round. Also, a tie is not possible.
- The tournament is suspended after the seventh round. At this point A has won 3 rounds, while B has won 4.
- They agree to finish then and divide up the jackpot, according to how likely an outright victory would have been for both.

How much money did A end up with?

Probability Trees

- Consider that A needed to win two more rounds, while B only need to win one more.
- One could suppose that B was twice as likely as A to win the jackpot.
- That would mean that the shares of the jackpot would be $\$2,000$ for A and $\$4,000$ for B.

Probability Trees

- Consider that A needed to win two more rounds, while B only need to win one more.
- One could suppose that B was twice as likely as A to win the jackpot.
- That would mean that the shares of the jackpot would be $\$2,000$ for A and $\$4,000$ for B.
- WRONG!

Probability Trees

- At the end of the seventh round, A had a 25% chance of winning the jackpot.
- A's share of the jackpot is the $\$1,500$.
- B had a 75% chance of winning, so gets $\$4,500$.

```
Cx= 75 ; Cy= 40
Ax= 25 ; Ay= 100
Bx= 75 ; By= 160
Dx= 125 ; Dy= 210
Bx= 75 ; By= 160
Ex= 125 ; Ey= 110

Xs=c(Cx,Ax,Bx,Dx,Bx,Ex)
Ys=c(Cy,Ay,By,Dy,By,Ey)
plot(Xs,Ys,pch=18,col="red",ylim=c(0,240),xlim=c(0,200))

lines(Xs,Ys,col="red",lwd=2)

text(38,70,"B wins",col="purple",font=2)
text(38,134,"A wins",col="purple",font=2)

text(88,190,"A wins",col="purple",font=2)
text(88,134,"B wins",col="purple",font=2)
```

```
#############################################

text(14,108,"A at 3",col="tomato",font=2)
text(14,92,"B at 4",col="tomato",font=2)


text(Cx+17,Cy,"P = 0.50",col="violetred2",font=2)
text(Dx+17,Dy,"P = 0.25",col="violetred2",font=2)
text(Ex+17,Ey,"P = 0.25",col="violetred2",font=2)
```

Contingency tables are used to examine the relationship between subjects' scores on two qualitative or categorical variables.

Example

- Consider the hypothetical experiment on the effectiveness of early childhood intervention programs described in another section.
- In the experimental group, 73 of 85 students graduated from high school. In the control group, only 43 of 82 students graduated. These data are depicted in the contingency table shown below.