Stat 202 - 2015 ~~8~~ XD W ~~4~~ 2 - ~~Tuesday~~ Mon (pg ~~1~~ 2)

New Normal Quantile Plot

QQ-plot in statcrunch)

Question What transformation makes the data N(0,1) distributed? In otherwords what transformation of the data makes the new variable N(0,1)?

The transformation that makes the data N(0,1) distributed will be linear if and only if the original data is Normally distributed.

So here's a way to tell if data are normally distributed.

Compute and Plot the transformation that makes it N(0,1). If the plot is a line the original data is Normal. If not it isn't. More sensitive to the eye than juding if density curve is bell shaped.

The plot is the Normal-Quantile Plot or QQ plot.

Stat 202-2015 ~~S1~~ W2 - ~~Tuesday~~ ~~Wed~~ Mon Pg 2

To get an intution about what a QQ plot looks like with Normal data we want to get some Normal data and plot it.

Hint: It won't be a perfect line even if data are Normal because the data are random.

We need to see plots to interpret plots.

Where are we going to get the data?

There are computer programs that output numbers that are Normally distributed.

In all other situations you are never sure its normal.

So we are going to use StatCrunch → data → simulate to make normal data.

(More will be said about using computers to generate random numbers in a few minutes).

Stat 202-2015 ~~SD~~ W~~2~~ - ~~Tuesday~~ ~~Wed~~ Mon 153

StatCrunch

Simulate 200 N(0,1) dynamic seed

Plot QQ plot

Do it again

Plot QQ Plot

mean

Simulate N(15, 1)

Plot QQ plot

Note: Old variable on vertical axis.

~~Uniform~~ - mention because on homework

Try with t-distribution if time permits and judging attention of class.

$\boxed{\text{Show link from website}}$

Stat 202 2015 ~~8~~ W~~8~~ - ~~Tues~~ Mon [illegible] pg ~~5~~ 4

Lets talk about using ~~the~~ a computer to generate random numbers

Computers are not random. Every time you run a program with the same input you get the same results

The solution is to program a a pseudorandom number generator

All random numbers in a computer are actually pseudorandom numbers

A Pseudorandom number generator will give you a stream of numbers that are indistinguisable from a random sequence (with the specified distribution) at least in theory.

There are different algorithms for doing this, some are better than others, none are perfect. Actually they are really really good. The only people who worry about their imperfections are mathematicans who study the algorithms

I said computers normally get the same result every time they run a program with the same input.

That's true of random number generators as well; you get the same stream of numbers everytime you run the program

But you have the option of selecting a seed. Every seed gives a different stream of numbers. And there a billions of possible seeds (any integer that can be represented by the computer will do).

In stat crunch if you pick a fixed seed you will get the same stream of numbers

Try it 3, 4 The nice thing about this is that with a fixed seed you can always go back and get the same seed.

Helpful for reproducing errors — debugging eg if you were programming a game.
If I tell you what seed to use, I'll know what answer you should get
(good for exams)

If you choose a dynamic seed StatCrunch will pick the seed for you based on the system clock so that every time you run it you get a different seed.

So we have talked about pseudorandomness. It is time to talk about randomness

§4.1

quoted from book:

We call a phenomenon random if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions

Coin Toss: 1/2 heads 1/2 tails
Die toss: 1/6 for each face
SAT scores of ~~Heights~~ of students: Normal (bell curve) ~~female or male ← might not be~~

The probability of any outcome of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions

1/2, for Heads
1/6 for 1 on dice / harder for SAT scores
we'll get to it.

Stat 202 - 2015 ~~SD~~ W~~4~~ - ~~Wed~~ Mon (Pg 7)

Review

Normal Quantile Plot (QQ Plot)

Plot the transformation that makes your data (old variable) and ~~normal~~ standard normal (new variable)

If plot is a line your original data was normal. Otherwise it wasn't

Won't be a perfect line because the data are random.

Homework gives you the idea and intuition

Generating pseudorandom numbers with compute

Want perfectly normal data, to test QQ plot so we use a computer program.

Random numbers are unpredictable

Pseudo random numbers ~~are~~ look for all essentially ~~[illegible]~~ purposes like random numbers but they are predictable if you know algorithm and seed

Each seed gives a different stream of random numbers

Any integer that the computer can represent can be used as a different seed

A dynamic seed in StatCrunch sets the seed according to the system clock and picks a different seed each time.

Chapter 4. We call a phenomenon random if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions

The probability of any outcome of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions

Coin toss P(Heads) = .5
P(tails) = .5

New

§4.2 Probability Models

A <u>probability model</u> is a description of a random phenomenon in the language of mathematics.

A probability model has two parts

book uses <u>list</u>

* A "list" of possible outcomes of the phenomenon
* The probability of each outcome

A "list" is not quite the right word here. The right word is <u>set</u>.

What is a set? How does it differ from a list?

A <u>set</u> is a <u>collection of objects</u>
A list is a less precise concept but list often means an enumeration (item 1) (item 2) etc has an order, set doesn't.

Examples of Sets:

$\boxed{\text{people in this room}}$

$\boxed{\text{numbers between 1 and 50}}$

Sets are written as

{people in this room}

A set comes with the concept

"is an element of" written $\in$

True or False?

Sean Carver $\in$ {people in this room}

Sean Carver is an element of the set of people in this room

True or False

False → Barak Obama $\in$ {people in this room}

B. O. is an element of the set of people...

BO $\notin$ { } true.

A set comes with the concept

"is a subset of" $\subset$

$A \subset B$ means every element of A is also an element of B

T or F {Women in this room} $\subset$ {people in the room}

{Women at AU} $\subset$ {people in this room}

False $\not\subset$

The concept of set comes with the concept of empty set $\emptyset$

$\{$giraffes in this room$\} = \emptyset$

(there are no giraffes in this room!

$\emptyset \subset A$ for Every set A

The empty set is a subset of every set (every element of $\emptyset$ is also an element of A)

$A \subset A$ for every set A

(every element of A is also an element of A).

Back to probability

The <u>sample space</u> S of a random phenomenon is the set of all possible outcomes of the phenomenon

Example $\{H, T\}$ for a coin toss

Warning: There is a certain amount of arbitraryness in assigning a sample space.

Lets say the coin lands on a table.

You could record just heads and tails in which case the sample space would be

$$\{H, T\}$$

OR you could record both heads and tails and the (x, y) coordinates of where the coin comes to rest on the table.

(H, x, y) and (T, x, y) would be different outcomes for each possible value of x, and y.

The ~~Experiment~~ Phenomenon would be the same what you record would be different

The sample space depends on what you record.

An event is a subset of the sample space.

Book's definition An event is an outcome or set of outcomes.

Outcomes of coin toss recording only heads and tails are

$$H, ~~and~~ T$$

The events are

$$\{H, T\} \; \{H\}, \{T\}, \emptyset$$

However the book doesn't distinguish between

$$H \text{ and } \{H\}$$

Makes a mathematician cringe

Both OK in terms of exam, homework

With subsets there is the notion of union and intersection

$A = \{\text{men in this room}\}$

$B = \{\text{people in this room wearing a blue shirt}\}$

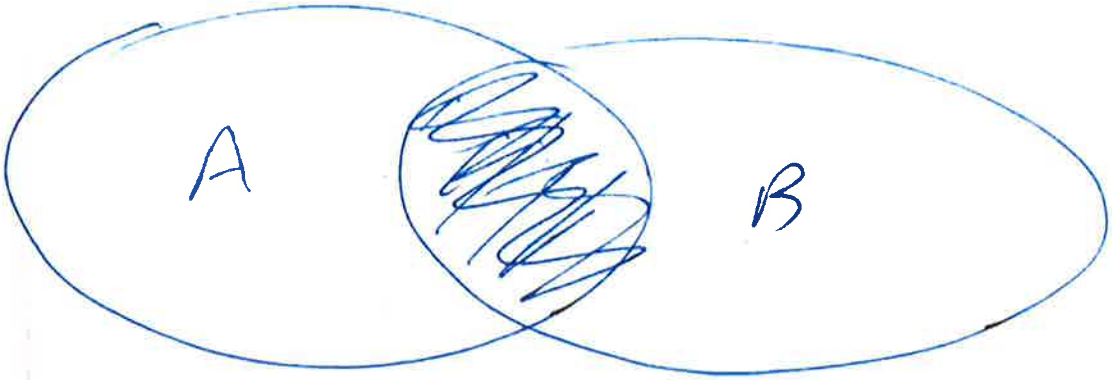
$\boxed{A \cup B}$ A Union B

would be

$A \cup B = \{$every person in this room who is *either* a man *or* wearing a blue shirt$\}$

thus the women wearing blue shirts would be elements of $A \cup B$

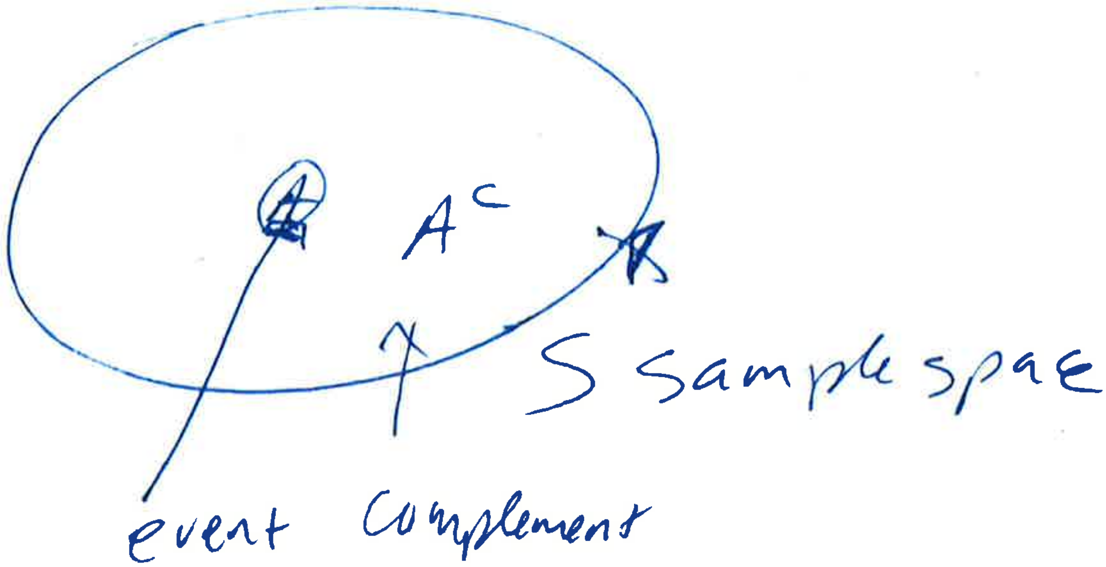
$\boxed{A \cap B}$ A intersection B



would be

$A \cap B$ {every person in this room who is **both** a man **and** wearing a blue shirt}

Finally there is the concept of complement



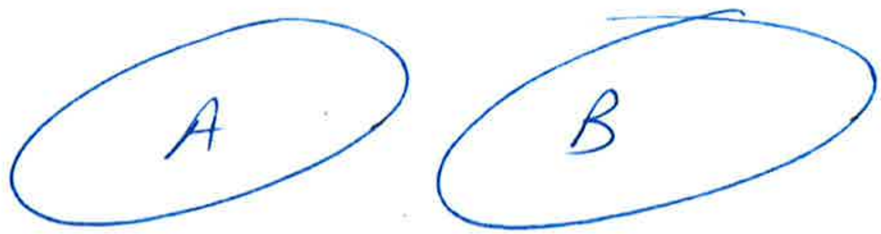
If A is an event

A^c is the set of every outcome that is not in A.

$$\{H\}^c = \{T\} \qquad \{T\}^c = \{H\}$$

$$\{H,T\}^c = \emptyset \qquad \emptyset^c = \{H,T\}$$

There is the notion of disjoint set



A and B are disjoint if they have no elements in common (intersection is empty set)

Now we are ready for probability

The probability is a number assigned to all events.

To be a legitimate probability assignment it must obey certain Rules.