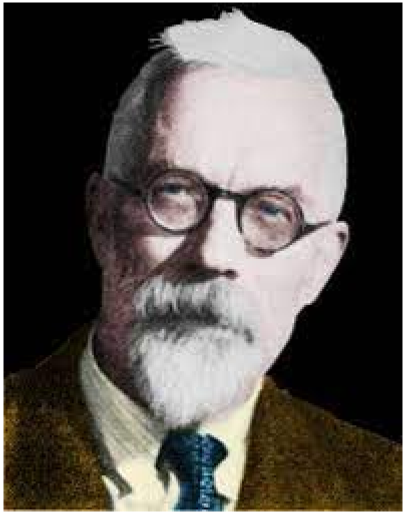
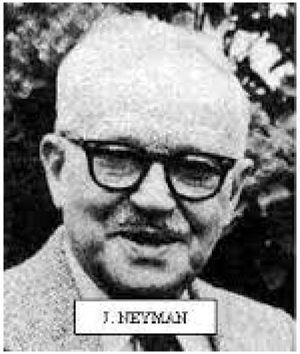
Fisher Tradition & Probability

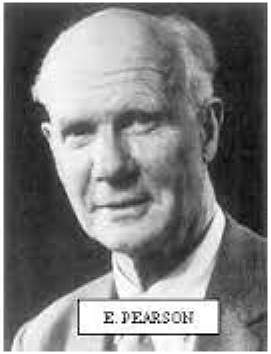
"Fisher" tradition



- Set up a statistical null hypothesis (note that null does NOT mean "nil")
- Report the exact level of significance
- Do not use a "conventional" level, do not talk about accepting/rejecting hypotheses, do not pass GO, and do not collect $200
- Use this procedure *only* if you know very little about the problem at hand

Neyman-Pearson





- Set up 2 hypotheses, and design a study based no the "rejection region" for each hypothesis
- If data is within the rejection range for H1, accept H2. Otherwise, accept H1. Note that accepting it doesn't mean you *believe* it...just that you act as though it was so
- Utility is limited to situations where there is a clear difference in hypotheses, when you can make a rational decision about when to accept vs. when to reject H1 and H2

So who won?

The two ideas melded together somehow into something that neither camp would be too excited by: 1) Set up a null hypothesis, where null almost always means "chance" 2) Make a yes-no decision about that hypothesis 3) Repeat

We want to know:

- what is the probability that we would get the values evidenced (*or those more extreme*) given our null hypothesis
- ☝️ $P(Data + |H_0)$ aka a **p-value**
- assumes, among other things, that the **null hypothesis is *exactly* true**, that you have a **random sample**, and that the scores are **independent**

Probability

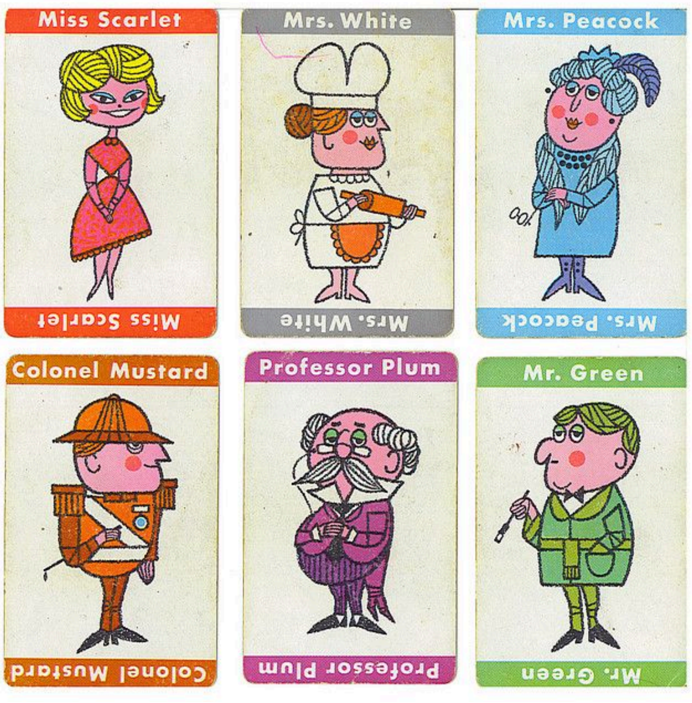
Sample Space & Assumptions

Our sample space is the range of possible values for a random variable. 6 Clue characters.

Assumption 1) Sum of all the probabilities of all outcomes needs to equal 1. $P(S) = 1$

Assumption 2) The probability of an event occurring must be between 0 and 1. $0 \leq P(event) \leq 1$

P(Miss Scarlet)



- P(Miss Scarlet) = N of events / sample size
- P(Miss Scarlet) = 1 Miss Scarlet / 6 characters
- P(Miss Scarlet) = 1/6

P(Female)

- P(Female) = N of events / sample size
- P(Female) = 3 females / 6 characters
- P(Female) = 3/6 = .5

Complement

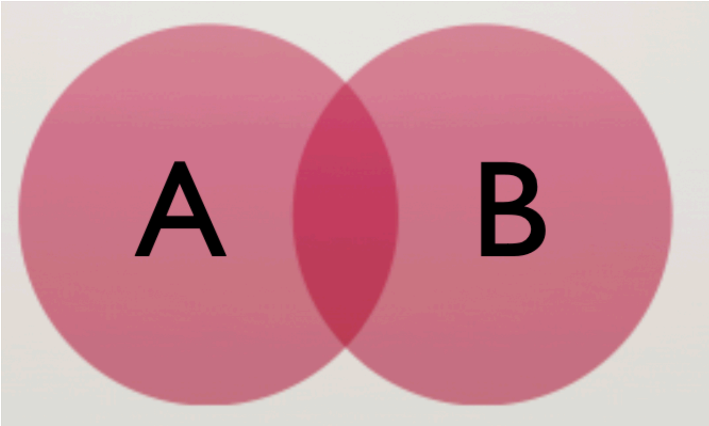
- The probability that the event does *not* occur
- $1 - P(event)$

P(NOT Female)

- P(NOT Female) = N of events / sample size
- P(NOT Female) = 3 not females / 6 characters
- P(NOT Female) = 3/6 = .5

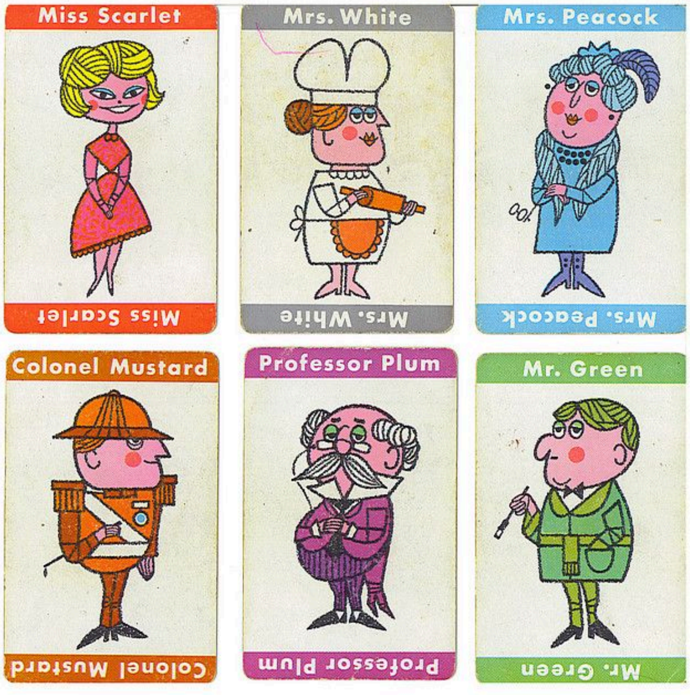
Unions

- The possbility of A ***or*** B occurring
- All elements that are in one of A or B
- $P(A \cup B)$



- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

P(Female or Holding Something)



P(Female or Holding Something)

P(Female)

- P(Female) = 3/6

P(Holding Something)

- P(Holding Something) = 4/6

P(Female & Holding Something)

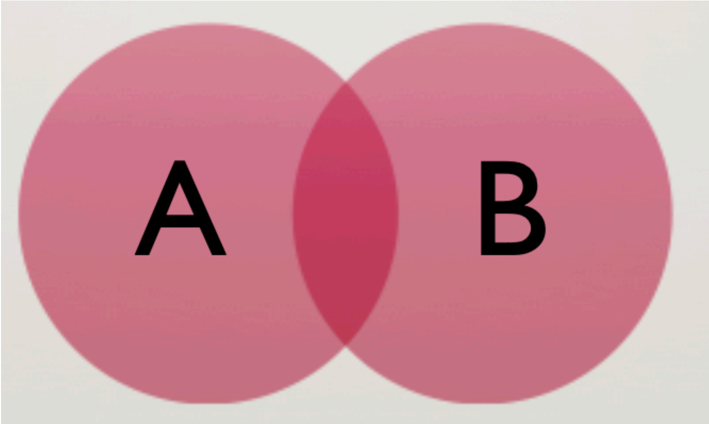
- P(Female & Holding Something) = 2 females with stuff / 6 characters
- P(Female & Holding Something) = 2/6

P(Female OR Holding Something)

- P(Female or Holding Something) = 3/6 + 4/6 - 2/6
- P(Female or Holding Something) = **5/6**

Intersection

- The probability of A *and* B occurring
- $P(A \cap B) = P(A) \times P(B|A)$



P(Baker & Female)

- P(Baker) = 1/6
- P(Female GIVEN there is a baker) = 1 baker that's female = 1
- P(Baker & Female) = 1/6 * 1 = 1/6

Intersection

- The probability of A *and* B occurring
- $P(A \cap B) = P(A) \times P(B|A)$
- P(Baker & Female) has *dependent events*; the occurrence of Event A changes the probability of Event B
- *Independent* events would be that the occurence of Event A does NOT impact the occurence of Event B
- *If independent*, $P(A \cap B) = P(A) \times P(B)$

Independence of observations is one of the criteria for having interpretable p-values!

- 2 games of Clue
- Finding the murderer for game 1 doesn't help you find the murderer for game 2
- P(Murderer in Game 1 & Murderer in Game 2) = $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

Permutations

All the ways n objects can be arranged

$N = 2$

```
##    O1 O2
## 1   b  a
## 2   a  b
```

Permutations

All the ways n objects can be arranged

$N = 3$

```
##    O1 O2 O3
## 1   c  b  a
## 2   b  c  a
## 3   c  a  b
## 4   a  c  b
## 5   b  a  c
## 6   a  b  c
```

Permutations

$N = 4$

```
##    01 02 03 04
## 1   d  c  b  a
## 2   c  d  b  a
## 3   d  b  c  a
## 4   b  d  c  a
## 5   c  b  d  a
## 6   b  c  d  a
## 7   d  c  a  b
## 8   c  d  a  b
## 9   d  a  c  b
## 10  a  d  c  b
## 11  c  a  d  b
## 12  a  c  d  b
## 13  d  b  a  c
## 14  b  d  a  c
## 15  d  a  b  c
## 16  a  d  b  c
## 17  b  a  d  c
## 18  a  b  d  c
## 19  c  b  a  d
## 20  b  c  a  d
## 21  c  a  b  d
## 22  a  c  b  d
## 23  b  a  c  d
```

Permutations

The number of permutations for n objects is:

$$n! = n(n-1)(n-2)\ldots$$

BUT, if you're looking for the number of r choices from n:

$$\frac{n!}{(n-r)!}$$

There are 6 Clue characters, but we are now 100% sure that 2 of them are the culprits. How many possibilities are there?

- $n = 6$, $r = 2$
- $\frac{6!}{(6-2)!} = \frac{6!}{4!}$
- $\frac{720}{24} = 30$

Combinations

1 Permutation = Miss Scarlet & Professor Plum. Another permutation is Professor Plum and Miss Scarlet. What if we don't care who comes first and who comes second? The paring of Plum/Scarlet should be good enough!

When order doesn't matter, we count the number of **combinations**

$$\frac{n!}{(n-r)!r!}$$

- n = 6, $r = 2$
- $\frac{6!}{(6-2)!2!} = \frac{6!}{4!2!}$
- $\frac{720}{24*2} = \frac{720}{48} = 15$
- We have a 1/15 chance of getting the correct 2 culprits out of 6 characters

Combinations on Combinations

We split up the class into 2 different games of Clue. In the first game, we assume 2 culprits. In the second game, we assume 3 culprits. The intersection of these combinations is the probability that we get both the 2 culprits correct in game 1 and the 3 culprits correct in game 2.

Game 1

- n = 6, $r = 2$
- $\frac{6!}{(6-2)!2!} = \frac{6!}{4!2!}$
- $\frac{720}{24*2} = \frac{720}{48} = 15$
- 1/15

Game 2

- n = 6, $r = 3$
- $\frac{6!}{(6-3)!3!} = \frac{6!}{3!3!}$
- $\frac{720}{6*6} = \frac{720}{36} = 20$
- 1/20

$$\frac{1}{15} \times \frac{1}{20} = \frac{1}{300}$$

Bernoulli Trials

- Exactly 2 possibilities, **success** or **failure**
- Flipping a coin. Heads = Success, Tails = Failure
- Let's say we flip a coin twice. Our possibilities are: TT, TH, HT, HH...

Bernoulli Trials

How many ways are there to get 0 Heads out of 2 coin flips?

- $\frac{2!}{(2-0)!0!} = 1$

How many ways are there to get 1 Heads out of 2 coin flips?

- $\frac{2!}{(2-1)!1!} = 2$

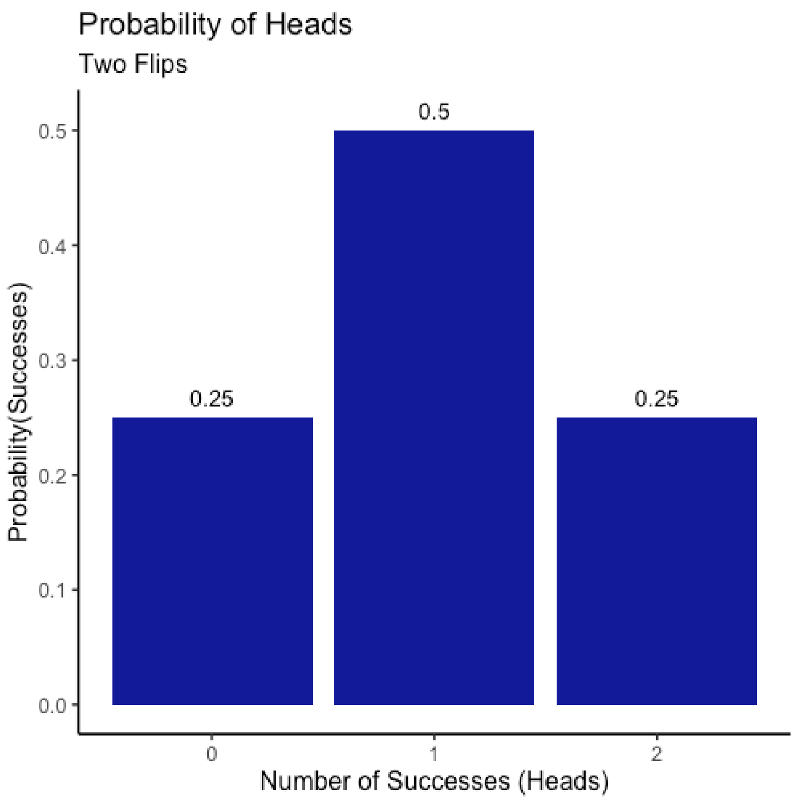
How many ways are there to get 2 Heads out of 2 coin flips?

- $\frac{2!}{(2-2)!2!} = 1$

Bernoulli Trials

How many outcomes are there total?

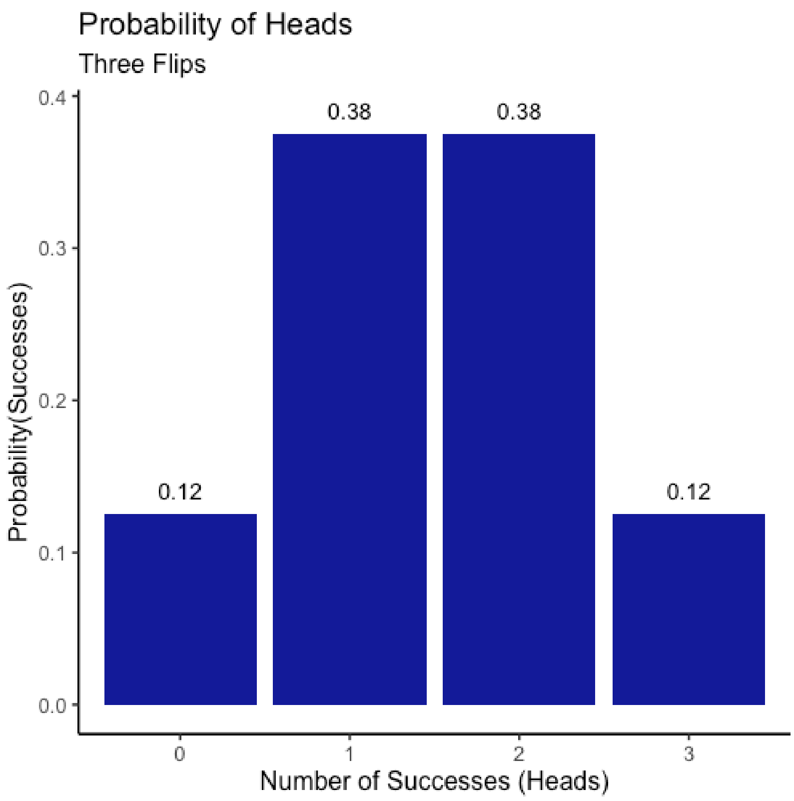
- 1 + 2 + 1 = 4
- Probabilities = 1/4, 2/4, 1/4



Bernoulli Trials

Three coin flips?

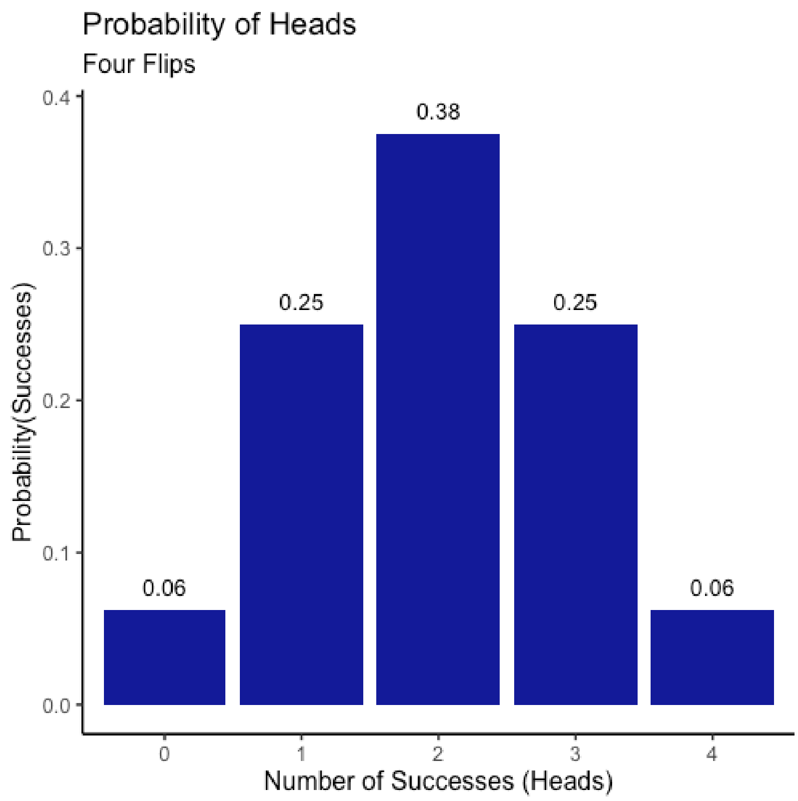
- TTT, TTH, THT, THH, HTT, HTH, HHT, HHH



Bernoulli Trials

Four coin flips?

Probability Mass Function



Counting Rule: Combinations

- Previous we used r for choices. Let's be more specific. Let k be the number of *successes*
- Total number of ways of getting k successes out of n trials, irrespective of order:

$$\frac{n!}{k!(n-k)!} = \binom{n}{k}$$

- Called **binomial coefficient**
- Go back a few slides...we just did this...a lot

What is the probability of getting 3 heads out of 5 coin flips?

```
Tails Heads Heads Tails Heads
```

This can be represented in binomial form. First, we need to choose which value represents a "success". Here, we'll use **Heads**.

```
NotHeads Heads Heads NotHeads Heads
```

The probability of that particular sequence is:

$$P(NotHeads)P(Heads)P(Heads)P(NotHeads)P(Heads)$$

$$P(Heads)^3P(NotHeads)^2 = (\frac{1}{2})^3(\frac{1}{2})^2 = 0.03125$$

What is the probability of getting 3 heads out of 5 coin flips?

But a specific sequence of independent outcomes is just one way we could have X successful trials out of N

- We need to know **how many possible ways** we could get X successes in N trials
  - HHHTT, HTHTH, TTHHH etc...

Remaining part of the equation is the combination rule for probability theory, $\left(\frac{N}{k}\right)$, it tells us how many different ways this can happen

```
H H H T T
H T H H T
H T H T H
H H T H T
H H T T H
```

```
T T H H H
T H T H H
T H H T H
T H H H T
H T T H H
```

$$\frac{n!}{k!(n-k)!} = \frac{5!}{3!(5-3)!}$$

$$\frac{5!}{3!2!} = \frac{120}{12} = 10$$

What is the probability of getting 3 heads out of 5 coin flips?

$$
\begin{aligned}
P(X = \text{a } head, \text{three times} | p_{.5}, n = 5) \\
&= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \frac{5!}{3!(5-3)!} (\frac{1}{2})^3 (\frac{1}{2})^2 \\
&= (10)(.03125) \\
&= .3125
\end{aligned}
$$

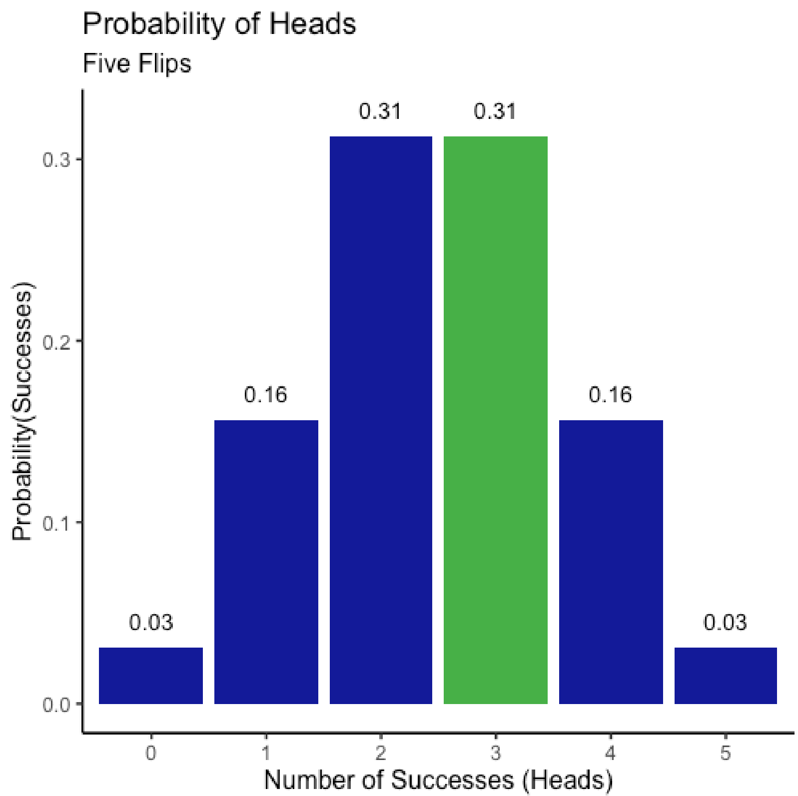
Or in R

```r
dbinom(x = 3, size = 5, prob = .5)
```

```
## [1] 0.3125
```

```r
data.frame(heads = 0:5, pr
```

```
##   heads    prob
## 1     0 0.03125
## 2     1 0.15625
## 3     2 0.31250
## 4     3 0.31250
## 5     4 0.15625
## 6     5 0.03125
```



Next time...

Continuing with the Binomial Distribution