Introductory Statistics
5: Fisher's Exact Test

Richard Veale

Graduate School of Medicine
Kyoto University

https://youtu.be/arEqCiqC-eI

Lecture Video at above link

Summary

1) What is Statistical Hypothesis Testing?

2) How to test 2 variables for statistical independence.

- Fisher’s exact test
- Chi-squared test (next week)

Expected vs Observed

Expected

Are you Man?

Like Ramen?

	Yes	No	TOTAL
Yes	4.8	1.2	6
No	3.2	0.8	4
TOTAL	8	2	10

Observed

Are you Man?

Like Ramen?

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

Those are pretty close…

..But are they close enough to say that we observed statistically independent results?

Expected vs Observed

Expected

Are you Man? (columns) / Like Ramen? (rows)

	Yes	No	TOTAL
Yes	4.8	1.2	6
No	3.2	0.8	4
TOTAL	8	2	10

Observed

Are you Man? (columns) / Like Ramen? (rows)

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

Those are pretty close…

Only off by 0.2 in all the squares…

Review:
Definition of Statistical Independence

How to compute the expected values, assuming statistical independence:

For two statistically independent events A and B:

P(A?B) = P(A) ;no influence of B on A
P(B?A) = P(B) ;no influence of A on B

Thus with P(A?B) = P(A&B) / P(B)
follows P(A) = P(A&B) / P(B)

solved for $P(A \cap B)$: P(A&B) = P(A) · P(B)
and: N(A&B) = N(A)· N(B) / N ; counts

Definition of Statistical Independence (in case you want traditional notation...)

How to compute the expected values, assuming statistical independence:

For two statistically independent events A and B:

$P(A|B) = P(A)$;no influence of B on A
$P(B|A) = P(B)$;no influence of A on B

Thus with $P(A|B) = P(A \cap B) / P(B)$
follows $P(A) = P(A \cap B) / P(B)$

solved for $P(A \cap B)$: $P(A \cap B) = P(A) \cdot P(B)$
and: $N(A \cap B) = N(A) \cdot N(B) / N$; counts

Review: Expected vs Observed

Expected

Are you Man?

Like Ramen?

	Yes	No	TOTAL
Yes	4.8	1.2	6
No	3.2	0.8	4
TOTAL	8	2	10

$N(A \cap B) = N(A) \cdot N(B) / N$

$4.8 = 6 \cdot 8 / 10$

$4.8 = 48 / 10$

Observed

Are you Man?

Like Ramen?

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

Hypothesis Testing

Expected

Are you Man?

Like Ramen?

	Yes	No	TOTAL
Yes	4.8	1.2	6
No	3.2	0.8	4
TOTAL	8	2	10

Observed

Are you Man?

Like Ramen?

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

5 – 4.8 = 0.2

The values are off by 0.2.

What would convince you that this is statistically significant?

Hypothesis Testing

Expected

Like Ramen? / Are you Man?

	Yes	No	TOTAL
Yes	4.8	1.2	6
No	3.2	0.8	4
TOTAL	8	2	10

Observed

Like Ramen? / Are you Man?

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

The values are off by 0.2.

What would convince you that this is statistically significant?

Maybe 0.2 is a “normal” amount to be different.

→ What if I tell you: we did the experiment 100 times and 80 times, it is off by 0.2.

Hypothesis Testing

Expected

Like Ramen? / Are you Man?

	Yes	No	TOTAL
Yes	4.8	1.2	6
No	3.2	0.8	4
TOTAL	8	2	10

Observed

Like Ramen? / Are you Man?

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

The values are off by 0.2.

What would convince you that this is statistically significant?

Maybe 0.2 is a “normal” amount to be different.

→ What if I tell you: we did the experiment 100 times and 80 times, it is off by 0.2.

...80% of the time! That seems very common, right?

In that case, 0.2 is “normal”, i.e. not statistically significant.

Hypothesis Testing

-Sometimes it will be off by 0.1.

-Sometimes it will be off by 0.3.

-Sometimes it will be off by 0.7.

Hypothesis Testing

-Sometimes it will be off by 0.1.

-Sometimes it will be off by 0.3.

-Sometimes it will be off by 0.7.

Usually, we don't care how often it is off by *exactly* 0.2

→ We care how often it is off by *0.2 or more*

Hypothesis Testing

Let's say you were to meet your friend John at the bus stop at 20pm.

It's getting late and later.
10 minutes late…
20 minutes late…

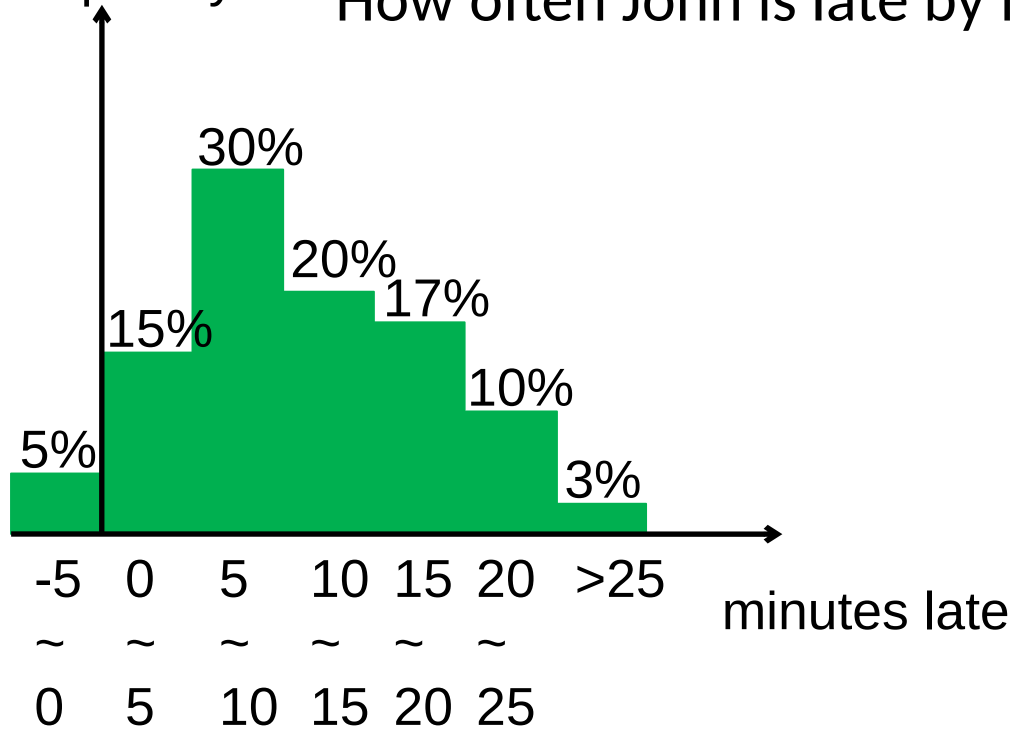
Should you be worried?

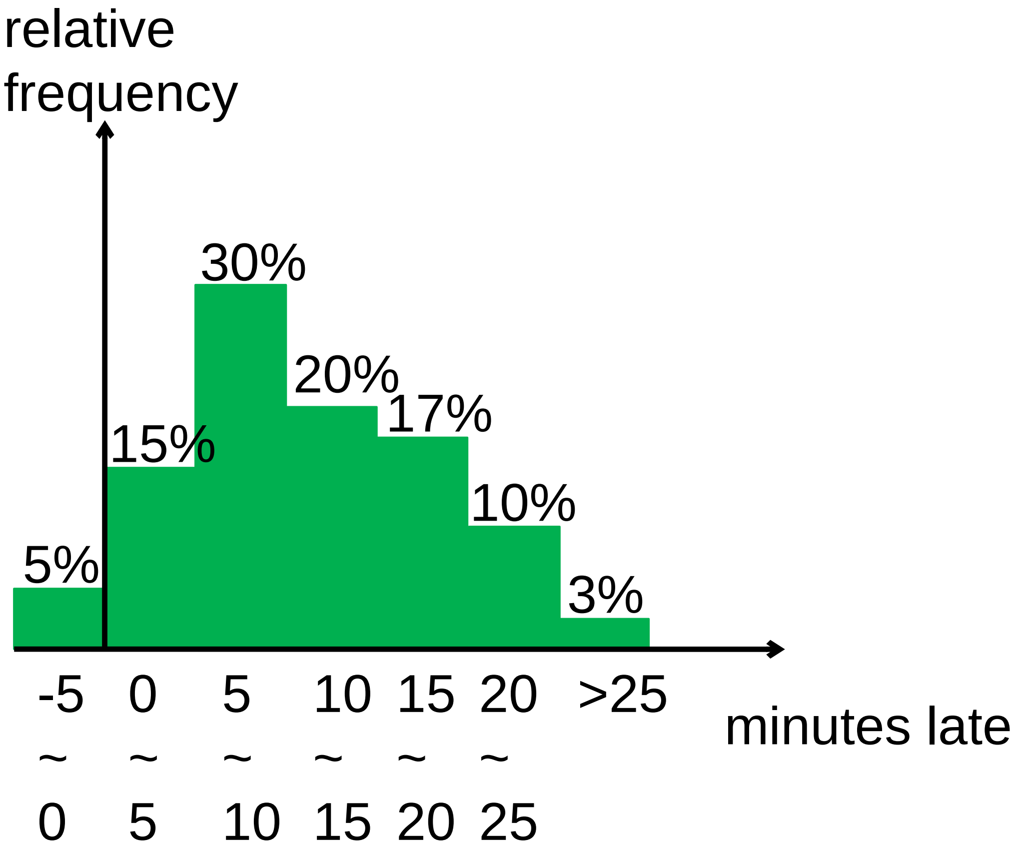
Should you go to his home and see
if he's okay? But then you might miss him when he comes to the bus stop.

Should you wait a little more? But what if he had an accident at home?

You rely on your experience:

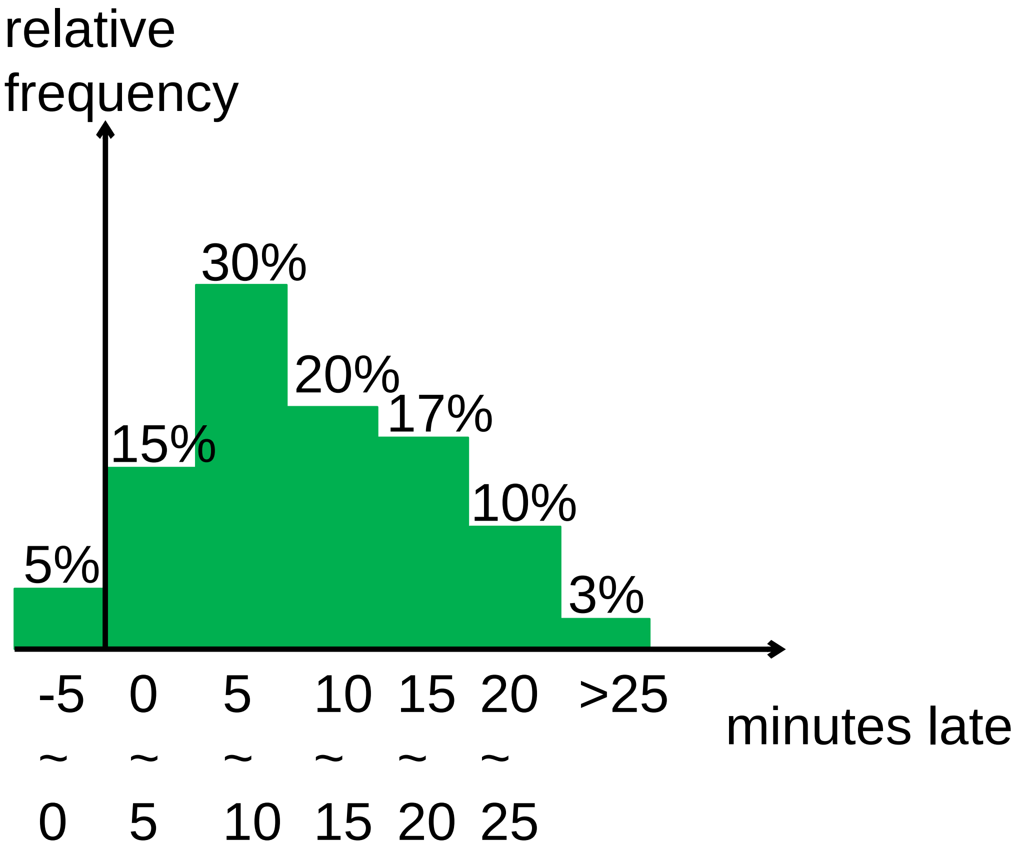
How often John is late by how much





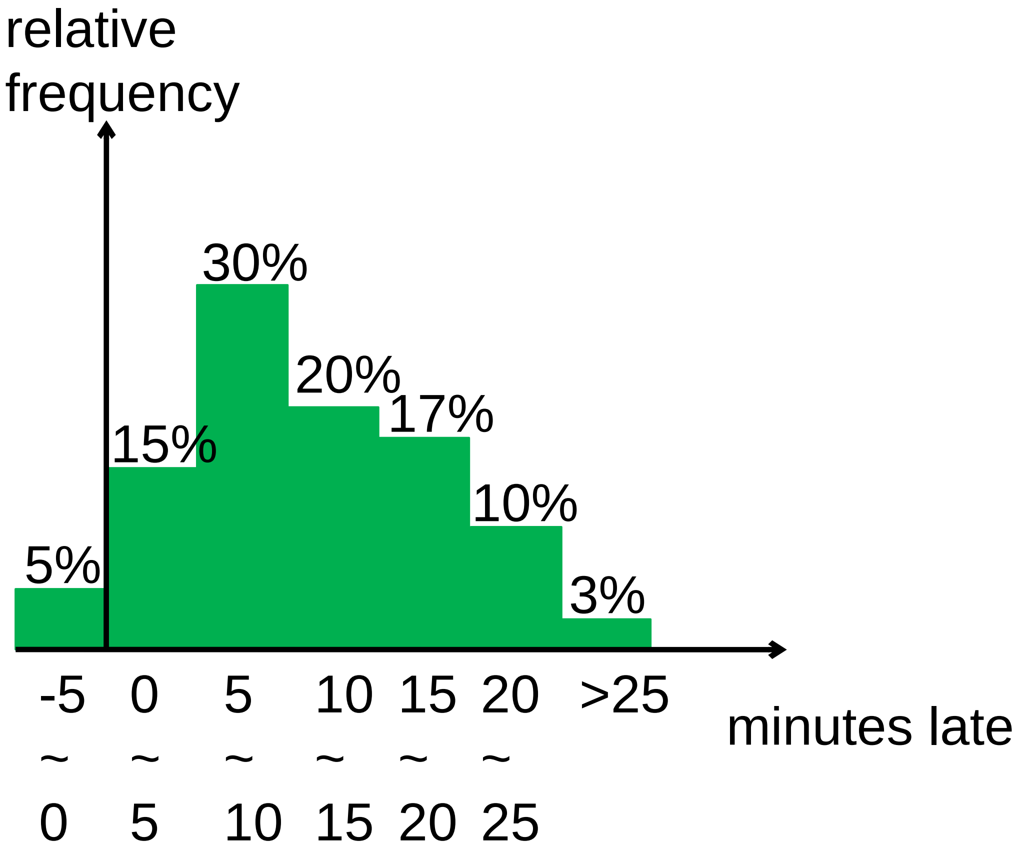
Null Hypothesis (H_0): John is just late as usual.

His behavior today follows the usual distribution.



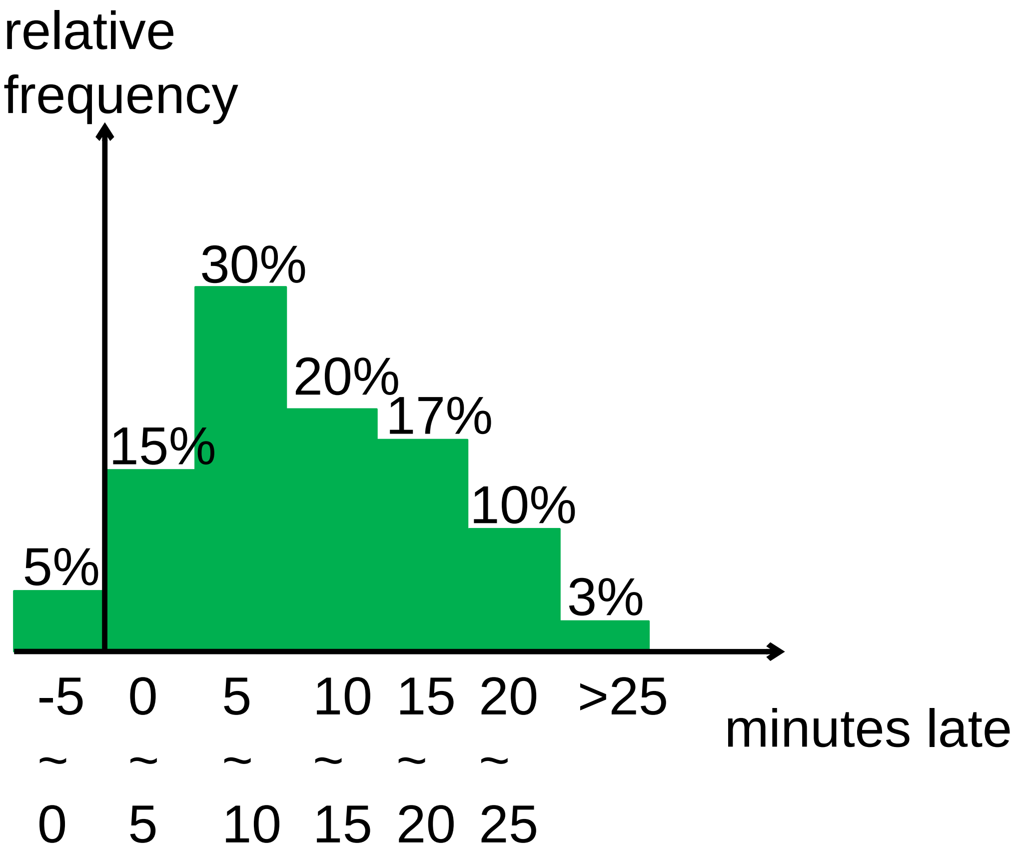
Alternative Hypothesis (H_a): something is odd.

His behavior today does not follow the usual distribution.

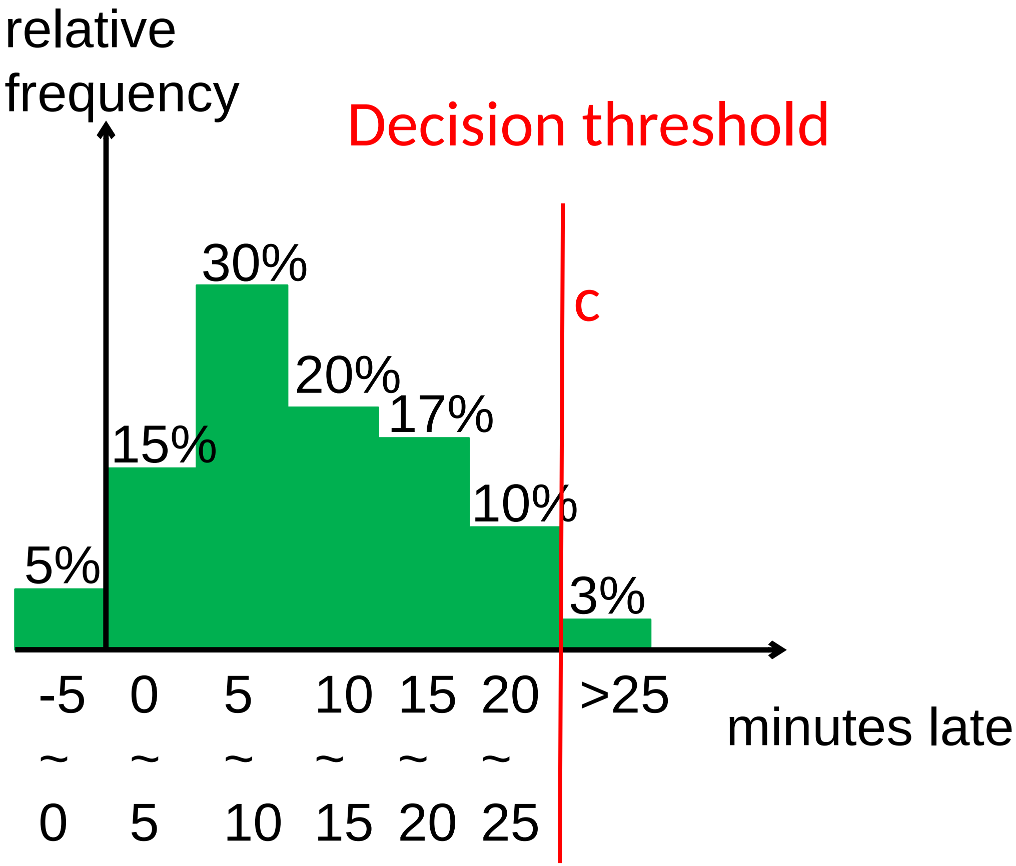


Type I error: Rejecting the Null Hypothesis (H_0) even though it is true.

Consequence: we leave our meeting place too early.

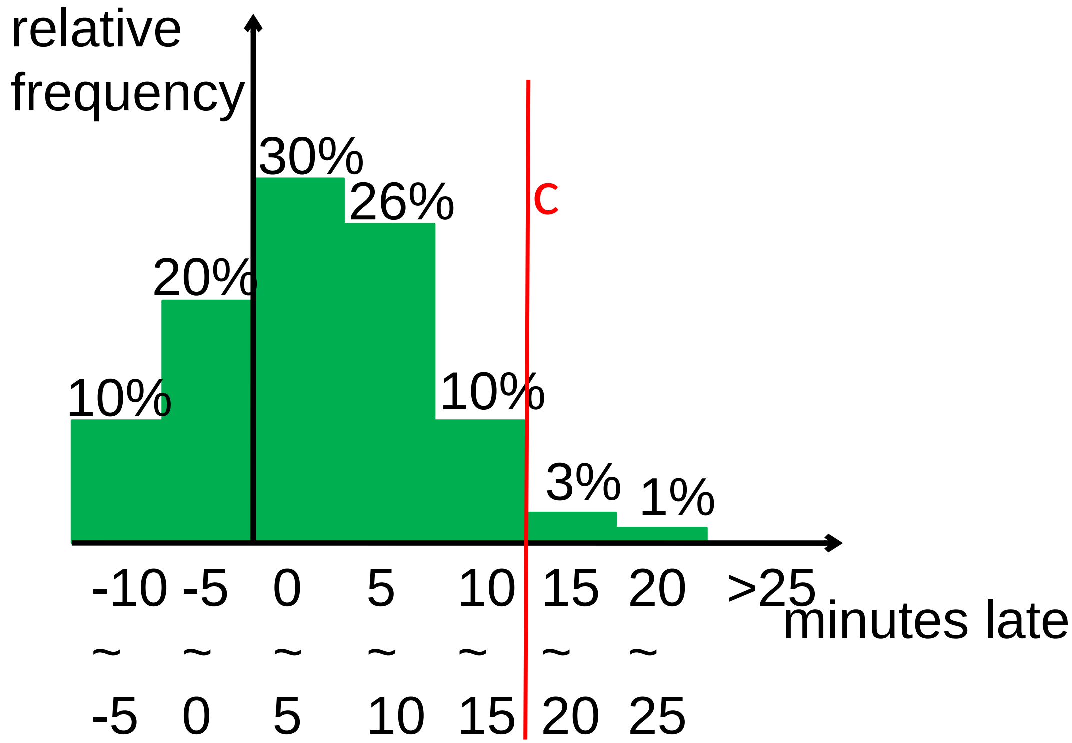


Type II error: Not rejecting the Null Hypothesis (H_0) even though it is false. Consequence: we are waiting for too long.



We want to keep the probability for a type I error low, say equal or below 5%, i.e., α=5%, so we wait for at least 25 minutes (critical delay c = 25 min).

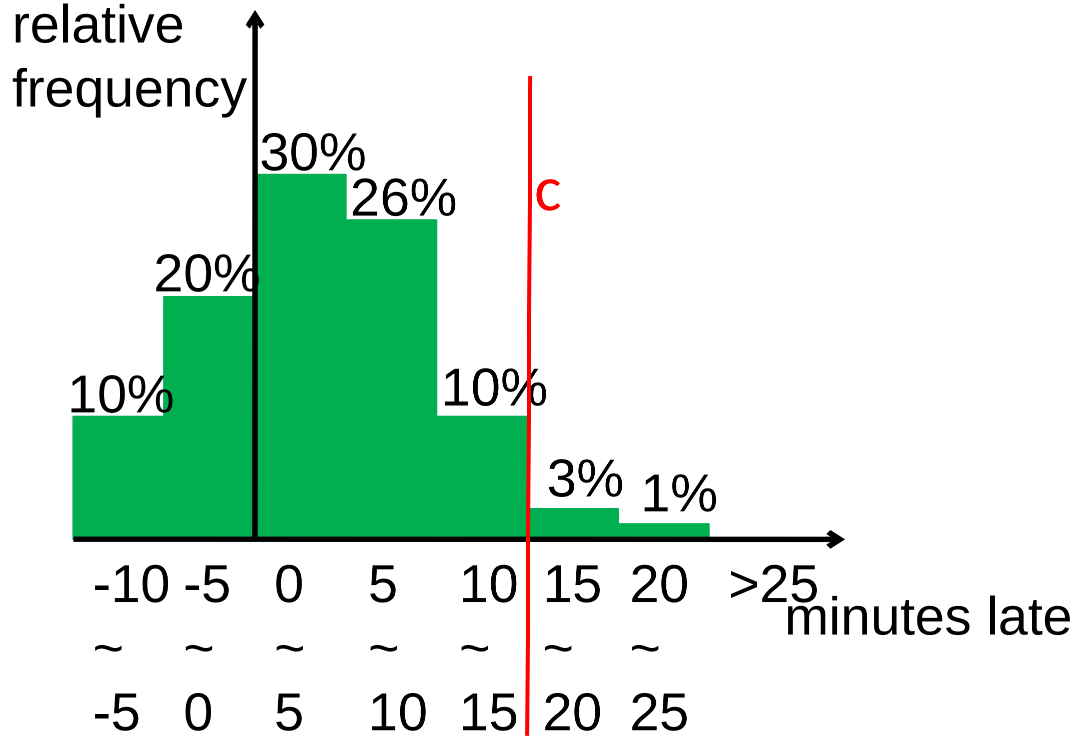
In case you are waiting for Jane…

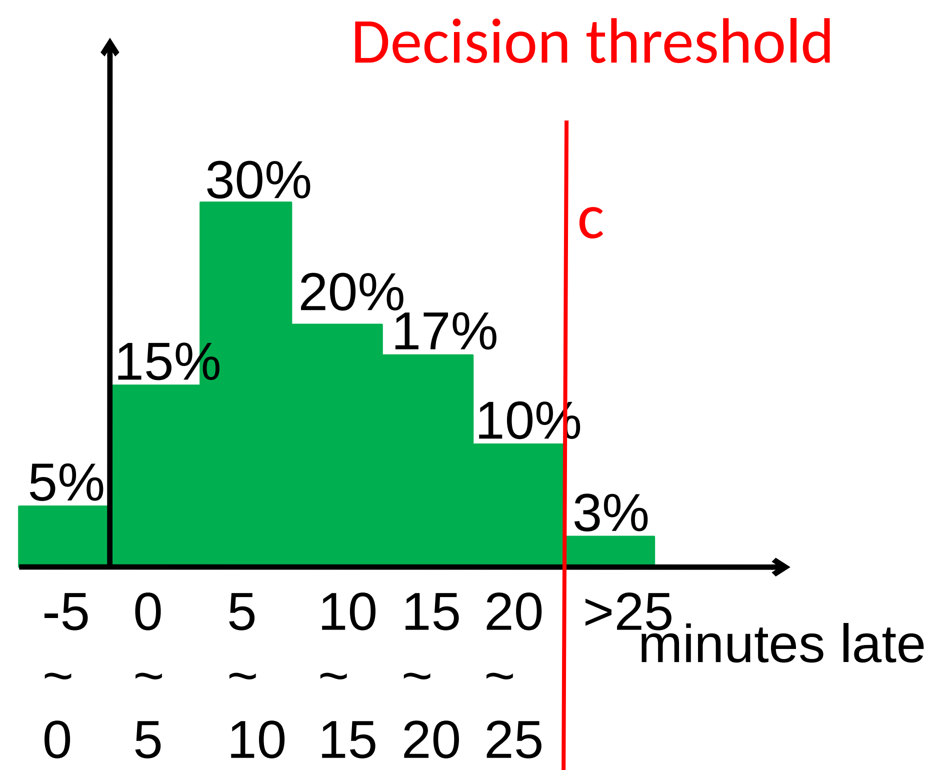


She has a different distribution of arrival times (shorter delays), so we adjust our decision threshold and wait only 15 minutes (c = 15 min).

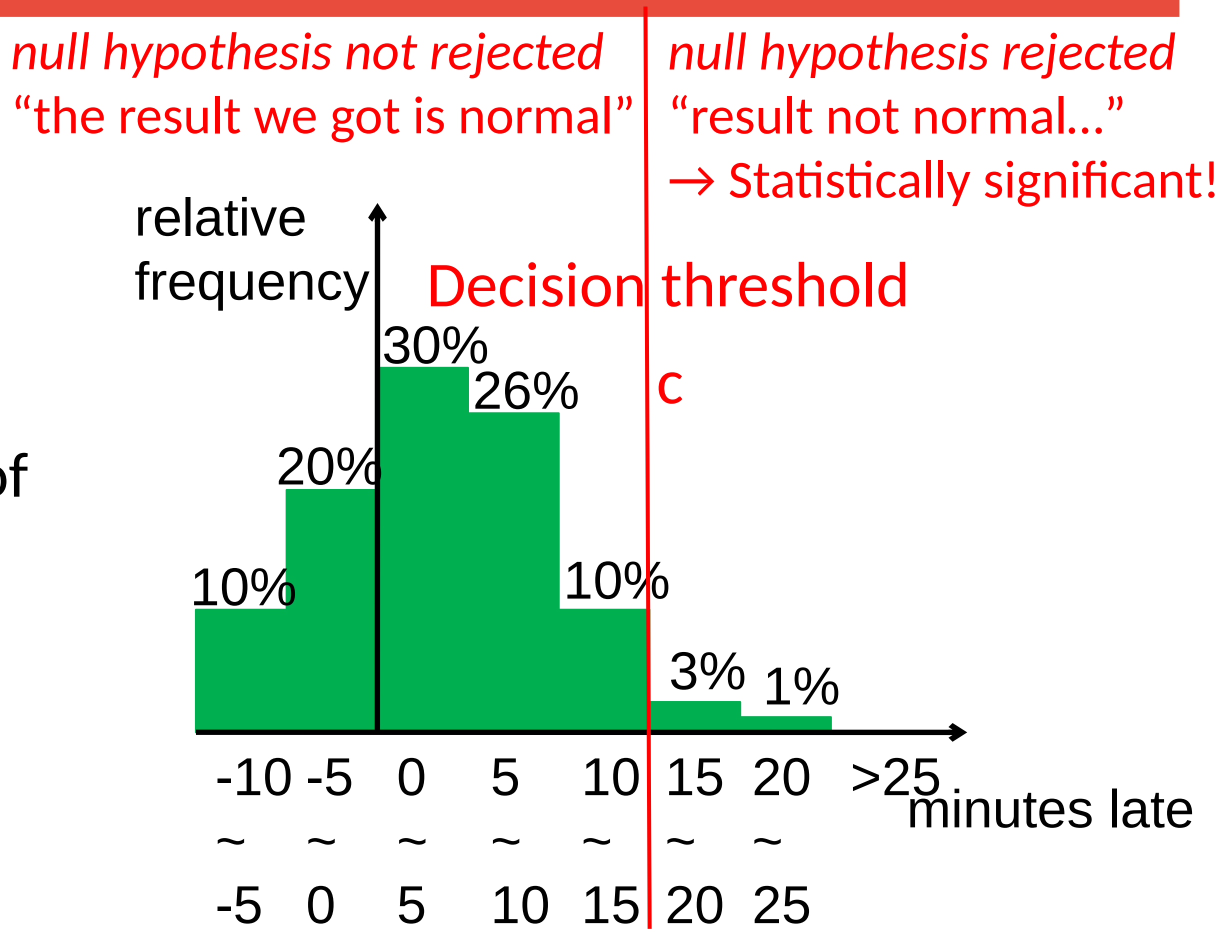
Two different distributions...

Different cutoff for
5% >= cutoff





So, we compute a single test statistic (here arrival time) and compare it to the distribution of the test statistic under the null hypothesis.



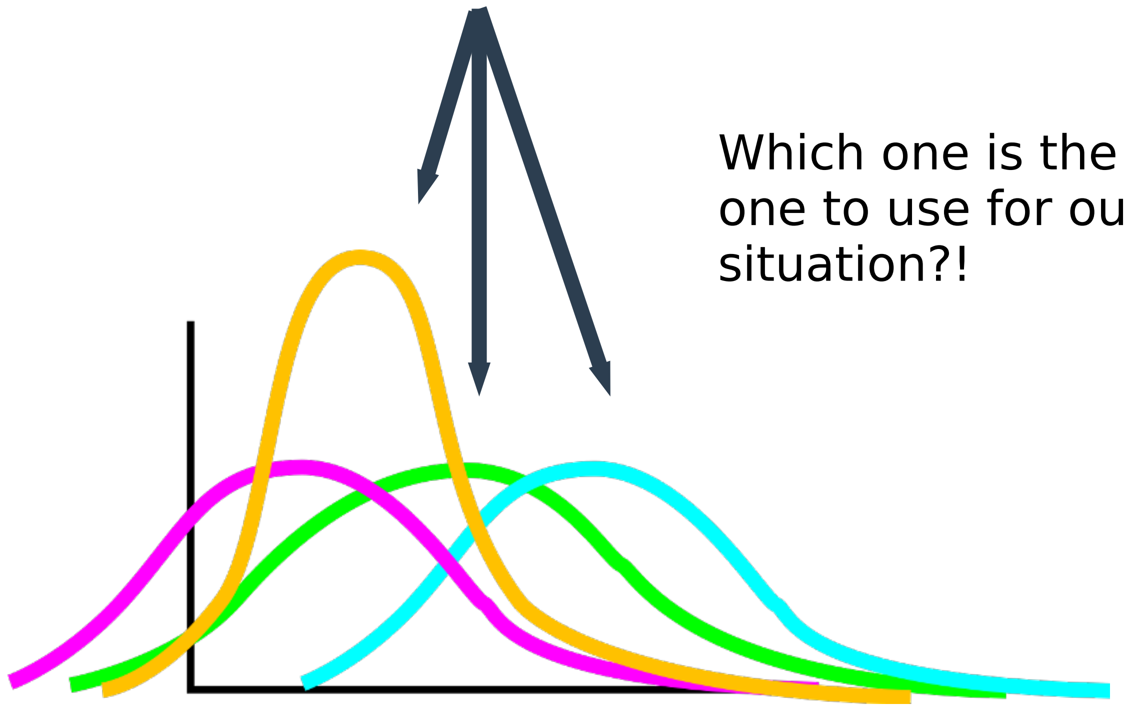
P(type I error) = P(rejecting H_0 when it is true) = α = P(delay ≥ c | H_0)

What distribution to use…?

Are you Man?

Like Ramen?

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10



Which one is the correct one to use for our situation?!

We can list *every possibility*...

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

→ If we know one table entry, we can compute the rest from the marginals

This is called the:
hypergeometric distribution

We can list *every possibility*...

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

→ If we know one table entry, we can compute the rest from the marginals

	Yes	No	TOTAL
Yes	4	1	6
No	3	1	4
TOTAL	8	2	10

What if we found only 4 men liked ramen?

We can list *every possibility...*

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

→ If we know one table entry, we can compute the rest from the marginals

	Yes	No	TOTAL
Yes	4	1	6
No	3	1	4
TOTAL	8	2	10

Rows and columns don't add up any more!

We can list *every possibility*...

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

→ If we know one table entry, we can compute the rest from the marginals

	Yes	No	TOTAL
Yes	4	2	6
No	3	1	4
TOTAL	8	2	10

Rows and columns don’t add up any more!

We can list *every possibility*...

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

→ If we know one table entry, we can compute the rest from the marginals

	Yes	No	TOTAL
Yes	4	2	6
No	3	1	4
TOTAL	8	2	10

Rows and columns don’t add up any more!

We can list *every possibility...*

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

→ If we know one table entry, we can compute the rest from the marginals

	Yes	No	TOTAL
Yes	4	2	6
No	3	0	4
TOTAL	8	2	10

Rows and columns don't add up any more!

We can list *every possibility*...

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

→ If we know one table entry, we can compute the rest from the marginals

	Yes	No	TOTAL
Yes	4	2	6
No	3	0	4
TOTAL	8	2	10

Rows and columns don't add up any more!

We can list *every possibility...*

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

→ If we know one table entry, we can compute the rest from the marginals

	Yes	No	TOTAL
Yes	4	2	6
No	4	0	4
TOTAL	8	2	10

OK, now it works..

We can list *every possibility...*

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

→ If we know one table entry, we can compute the rest from the marginals

	Yes	No	TOTAL
Yes	4	2	6
No	4	0	4
TOTAL	8	2	10

What if we found that 6 men liked ramen?

Again, rows and columns must add up to marginals...

	Yes	No	TOTAL
Yes	6	1	6
No	1	1	4
TOTAL	8	2	10

We can list *every possibility*...

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

→ If we know one table entry, we can compute the rest from the marginals

	Yes	No	TOTAL
Yes	4	2	6
No	4	0	4
TOTAL	8	2	10

	Yes	No	TOTAL
Yes	6	0	6
No	2	2	4
TOTAL	8	2	10

We can list *every possibility...*

	Yes	No	TOTAL
Yes	3	3	6
No	5	-1	4
TOTAL	8	2	10

	Yes	No	TOTAL
Yes	8	-2	6
No	0	4	4
TOTAL	8	2	10

	Yes	No	TOTAL
Yes	1	5	6
No	7	-3	4
TOTAL	8	2	10

Actually, for our data, there is only these 3 possibilities!

If we set it to 3, we must have 3 non-men who like ramen to add up to 6...

But we only have 2 non-men! So we'd have a negative person in one of the groups...

If we set it to 8, then we would have to have 4 non-men who *don't* like ramen.

But, we only have 2...

We can't have negative people

We can list *every possibility...*

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

	Yes	No	TOTAL
Yes	4	2	6
No	4	0	4
TOTAL	8	2	10

	Yes	No	TOTAL
Yes	6	0	6
No	2	2	4
TOTAL	8	2	10

Actually, it's easier to think about if we look at the non-men...

How many ways can you separate 2 people into 2 groups A and B?

There are only 3 ways.

A – B
----
0 – 2
1 – 1
2 – 0

We can list *every possibility*...

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

	Yes	No	TOTAL
Yes	4	2	6
No	4	0	4
TOTAL	8	2	10

	Yes	No	TOTAL
Yes	6	0	6
No	2	2	4
TOTAL	8	2	10

How many ways can you separate 2 people into 2 groups A and B?

There are only 3 ways.

A – B
----
0 – 2
1 – 1
2 – 0

But this is just *possibilities*.

This is not *probability.*

We can list *every possibility*...

Here are our two non-men

..let’s see how we can distribute them.

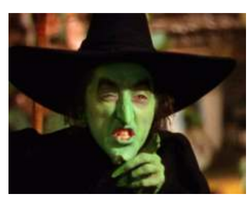
We can list *every permutation*...

2 - 0

Likes	Doesn't like
 	

1 - 1

Likes	Doesn't like
	

Likes	Doesn't like
	

0 - 2

Likes	Doesn't like
	 

We can list *every permutation*...

2 - 0

1 - 1

0 - 2

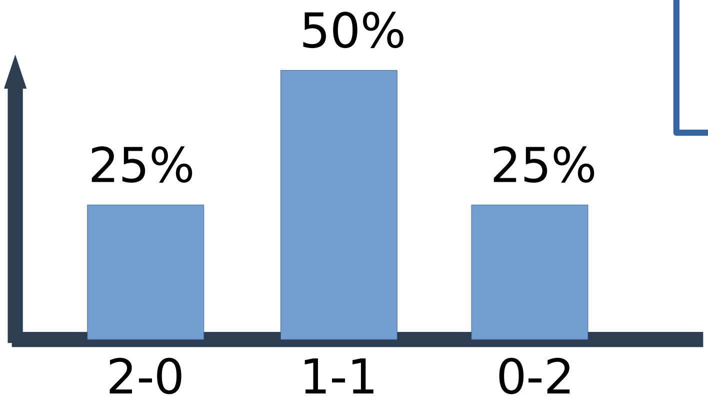
Order doesn’t matter within the groups...

We can list *every permutation*...

2 - 0

1 - 1

0 - 2



Out of 4 possible permutations…

But this is just the non-males!

We actually have 10 people...

Are you Man?

Like Ramen?

	Yes	No	TOTAL
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10










We can list *every permutation*...

6 - 4

Likes






Doesn't like




We can list *every permutation*...

6 - 4

Likes






Doesn't like




We can list *every permutation*...

6 - 4

Likes	Doesn’t like
     	   

We can list *every permutation*...

6 - 4

Likes	Doesn't like
     	   

We can list *every permutation*...

6 - 4

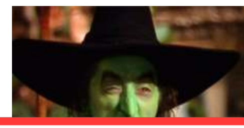
Likes	Doesn't like
     	   

We can list *every permutation*...

6 - 4

Likes	Doesn’t like



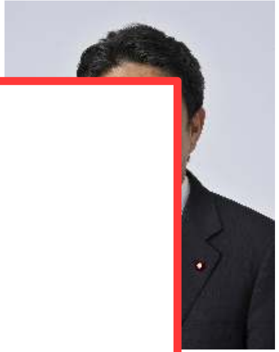



This is going to take a while...




We can list *every permutation*...

6 - 4

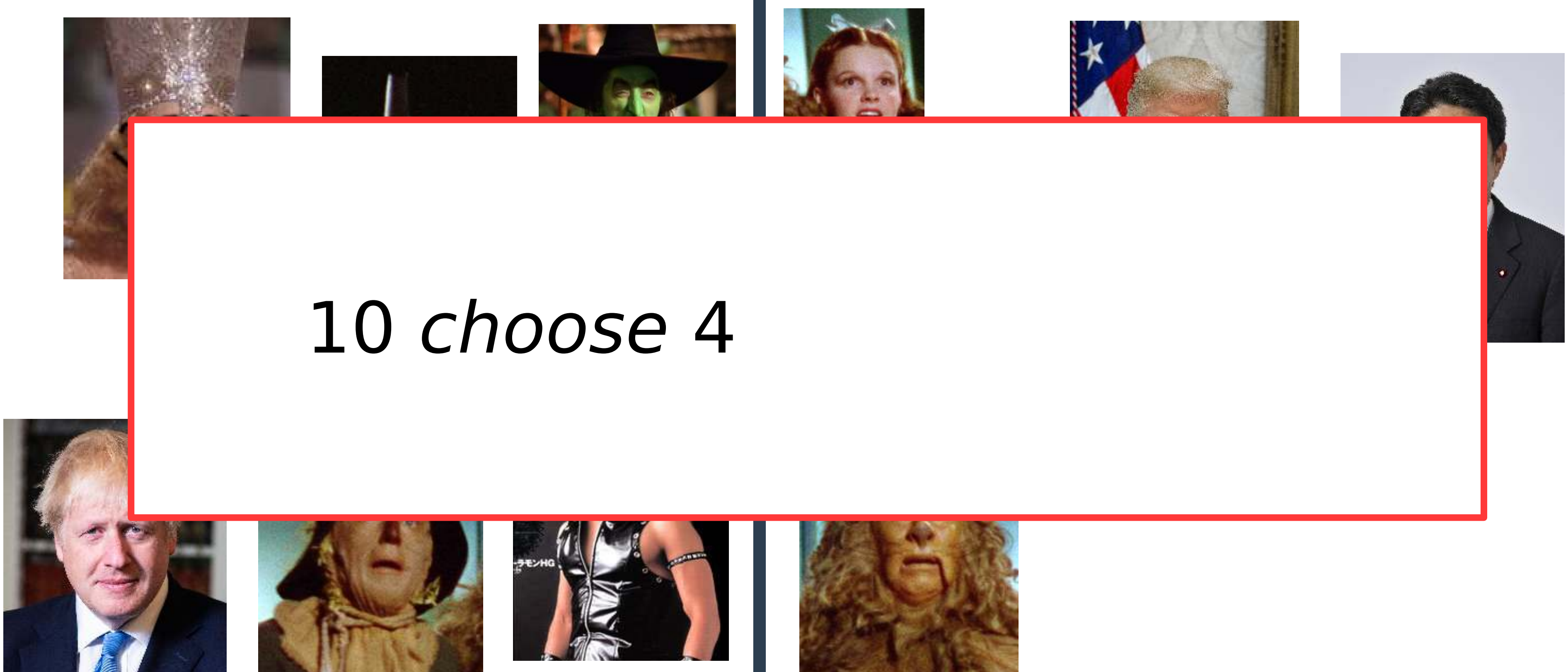
Likes	Doesn't like
     	   

N *choose* r

We can list *every permutation*...

6 - 4

Likes	Doesn't like



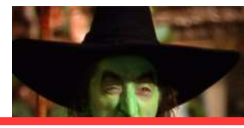
10 *choose* 4

We can list *every permutation*...

6 - 4

Likes	Doesn't like






Same as:
10 *choose* 6

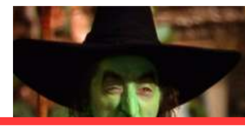



We can list *every permutation*...

6 - 4

Likes	Doesn't like

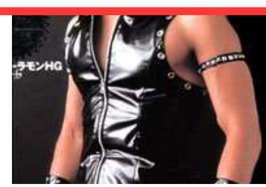





$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

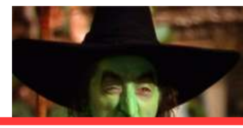
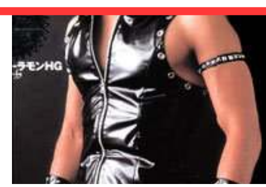
$$a! = \prod_{x=0}^{a-1} a - x$$






We can list *every permutation*...

6 - 4

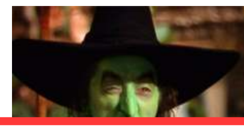
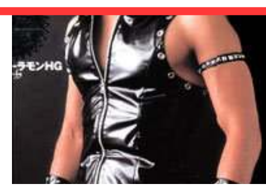
Likes	Doesn't like
     	   

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

$$a! = a \times (a-1) \times (a-2) \times (a-3) \ldots (a-(a-1))$$

We can list *every permutation*...

6 - 4

Likes	Doesn't like
  	  
  	

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

Generally, a really big number...

“Acceptable” possibilities (Match our data)

6 - 4

Unacceptable! We need 1 non-male on each side!

“Acceptable” possibilities (Match our data)

6 - 4

We need to count just the *acceptable* ones…

“Unacceptable” outcomes

Unacceptable outcomes in our situation is when there is 2 non-males in either category.
→ we need 1-1

Number of ways to choose 2 males (from our 8)

Number of ways to choose 2 non-males (from our 2)

$$\frac{\binom{8}{2} \times \binom{2}{2}}{\binom{10}{4}} = \frac{28 \times 1}{210}$$

Number of ways to choose 4 people from our 10

"Unacceptable" outcomes

Unacceptable outcomes in our situation is when there is 2 non-males in *either* category.
→ we need 1-1

Unacceptable #1

Number of ways to choose 2 males (from our 8)

Number of ways to choose 2 non-males (from our 2)

$$\frac{\binom{8}{2} \times \binom{2}{2}}{\binom{10}{4}} = \frac{28 \times 1}{210} = 0.133333$$

Number of ways to choose 4 people from our 10

“Unacceptable” outcomes

Unacceptable outcomes in our situation is when there is 2 non-males in *either* category.
→ we need 1-1

Unacceptable #2

Number of ways to choose 4 males (from our 8)

Number of ways to choose 0 non-males (from our 2)

$$\frac{\binom{8}{4}\times\binom{2}{0}}{\binom{10}{4}} = \frac{70\times 1}{210} = 0.333333$$

Number of ways to choose 4 people from our 10

“Unacceptable” outcomes

So, unacceptable outcome probability is:

0.333333 + 0.133333 = 0.4666666

Probability of “successful” outcome is:

1.0 – 0.4666666 = 0.5333333

Check it…

Acceptable number

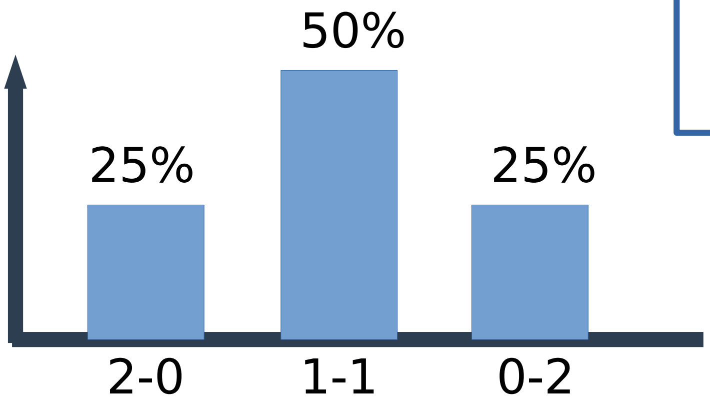
Number of ways to choose 1 non-males (from our 2)

Number of ways to choose 3 males (from our 8)

$$\frac{\binom{8}{3}\times\binom{2}{1}}{\binom{10}{4}}=\frac{56\times 2}{210}=0.533333$$

Number of ways to choose 4 people from our 10

Why…? Shouldn't it be 50/50?



Why…? Shouldn't it be 50/50?

2 - 0

Likes	Doesn't like
 	

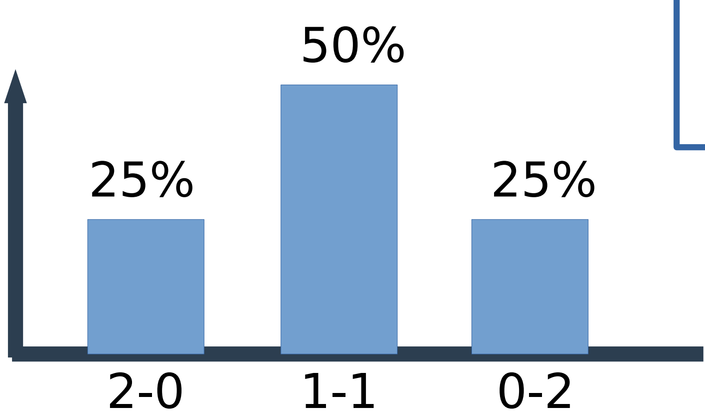
1 - 1

Likes	Doesn't like
	

Likes	Doesn't like
	

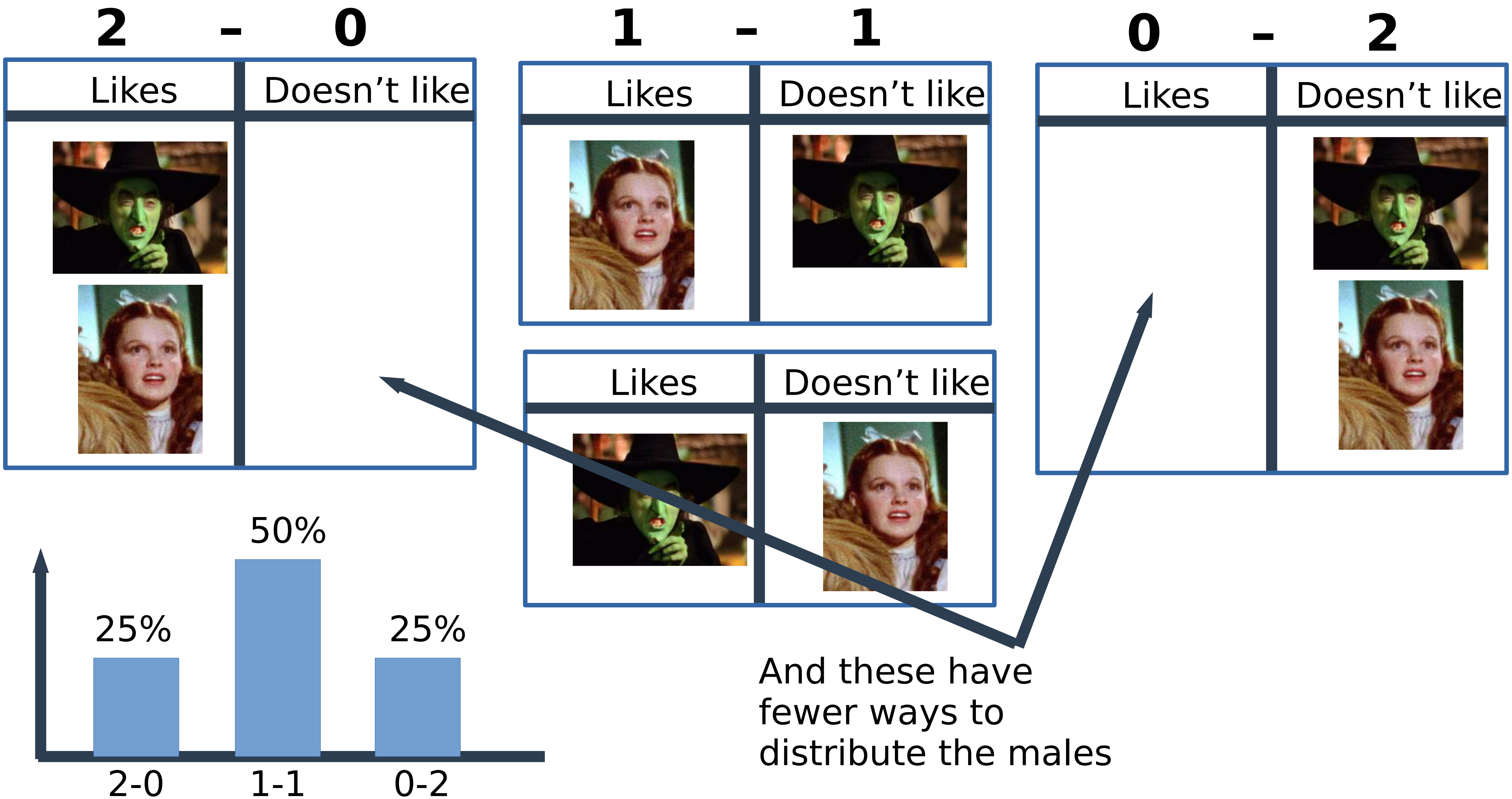
0 - 2

Likes	Doesn't like
	 



Hidden in these is more possible ways to distribute the males

Why…? Shouldn't it be 50/50?

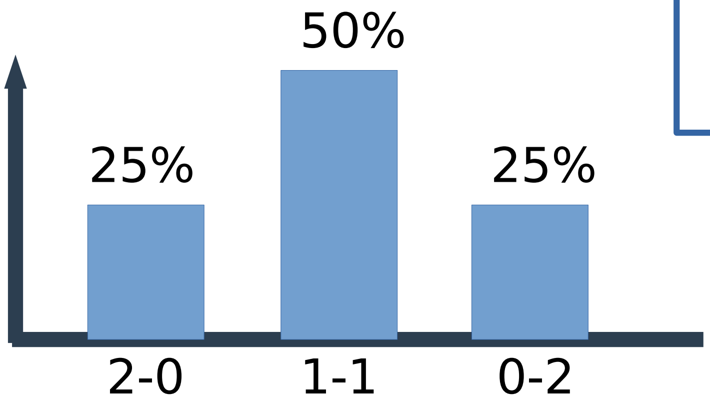


Why…? Shouldn’t it be 50/50?

2 - 0

1 - 1

0 - 2



Because:

$$\binom{8}{3}\times\binom{2}{1} > \binom{8}{4}\times\binom{2}{2} + \binom{8}{2}\times\binom{2}{0}$$

Why…? Shouldn't it be 50/50?

2 - 0

Likes	Doesn't like
 	

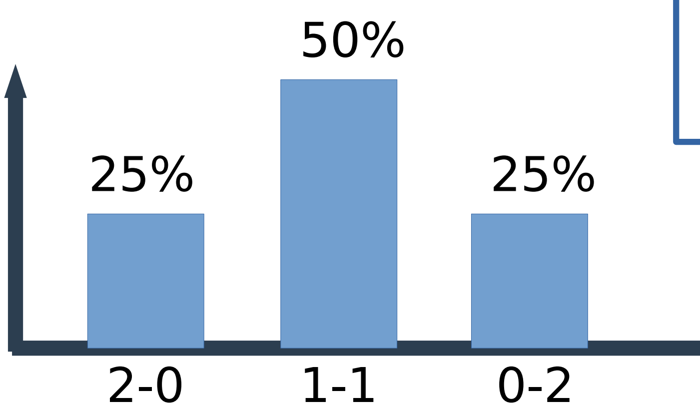
1 - 1

Likes	Doesn't like
	

Likes	Doesn't like
	

0 - 2

Likes	Doesn't like
	 



Because:

$$56 \times 2 > 70 \times 1 + 28 \times 1$$

Probability of a given outcome x:

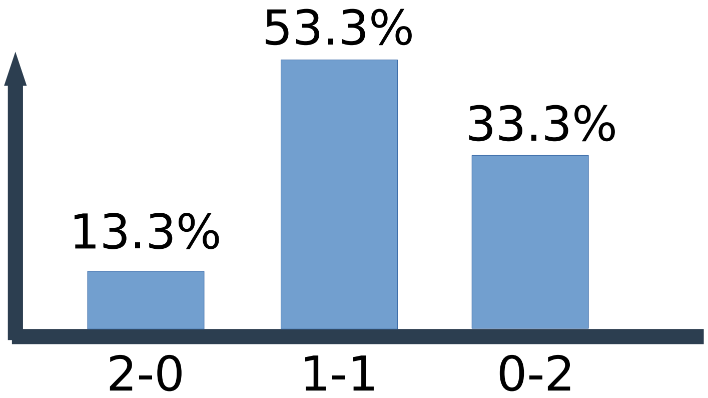
	Group Y	Group !Y	Total
Group A	x	M-x	M
Group !A	K-x	N-M-K+x	N-M
Total	K	N-K	N

$$P(X=x)=\frac{\binom{M}{x}\times\binom{N-M}{K-x}}{\binom{N}{K}}$$

This defines the *hypergeometric distribution (*for 2x2 tables)

Probability of drawing (w/o replacement) “x successes” in K draws where you have M objects of that feature and total population N.

So our actual distribution...



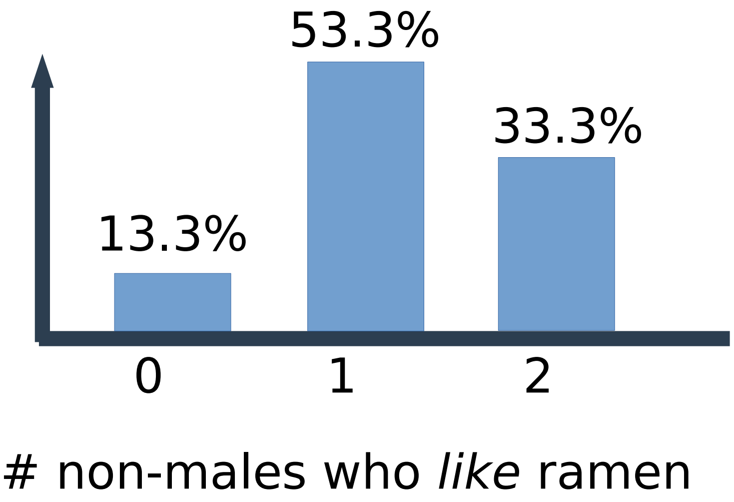
Non-male dislikes ramen ↔ Non-male likes ramen

Now we're in business.

We can use this to do statistics!

Before we did any of this, we should have chosen a hypothesis, and a cutoff…

"Non-men like ramen less than men"
Type 1 Error rate: 15% (usually you should use at MOST 5%)

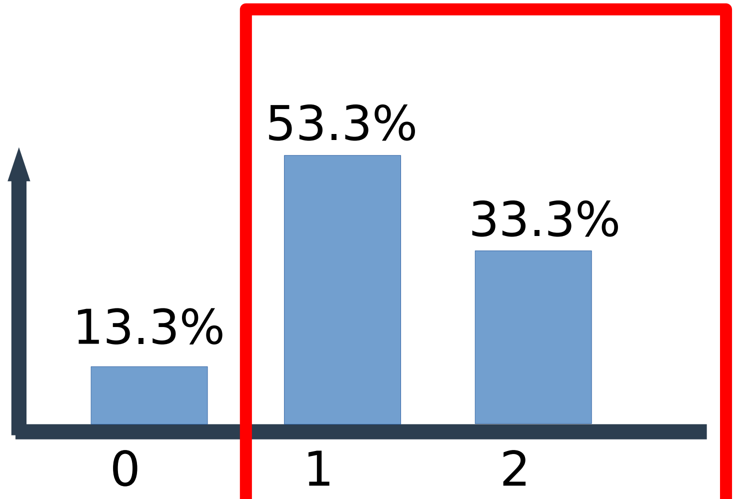


Now we're in business.

We can use this to do statistics!

Before we did any of this, we should have chosen a hypothesis, and a cutoff…

"Non-men like ramen less than men"
Type 1 Error rate: 15% (usually you should use at MOST 5%)



Probability of getting result *equal to* or *greater than* the one observed = 86.66%

non-males who *like* ramen

Now we’re in business.

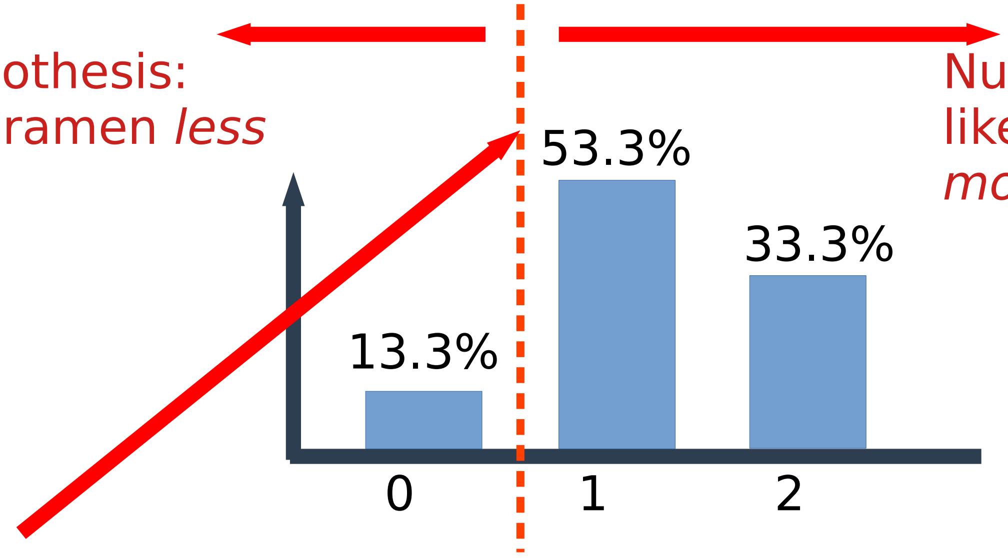
We can use this to do statistics!

Before we did any of this, we should have chosen a hypothesis, and a cutoff…

Hypothesis: “Non-men like ramen less than men”
Type 1 Error rate: 15% (usually you should use at MOST 5%)

Alternate hypothesis: non-men like ramen *less* than men

Null hypothesis: non-men like ramen *equal to* or *more* than men



non-males who *like* ramen

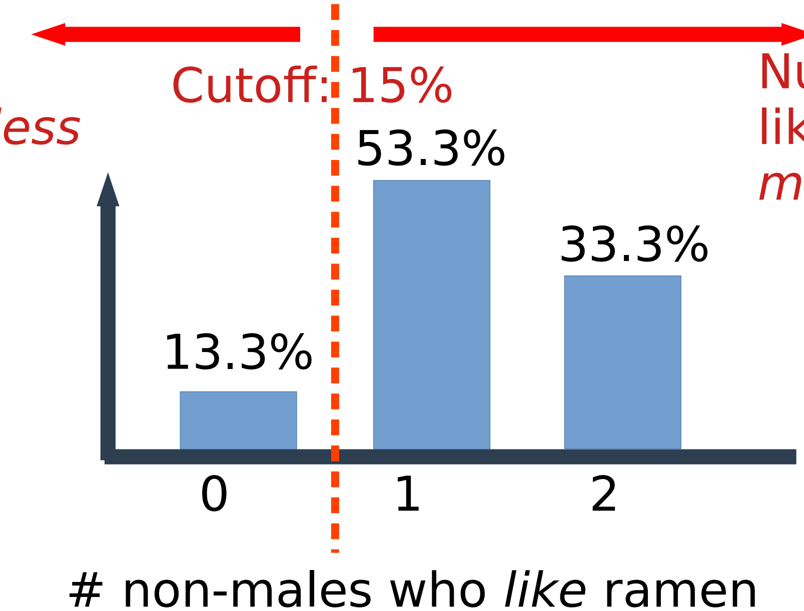
Cutoff so that 15% or less is below

Now we're in business.

Where does our statistic fall?

Hypothesis: Non-men like ramen less than men
Type 1 Error rate: 15% (usually you should use at MOST 5%)

Alternate hypothesis: non-men like ramen *less* than men



Null hypothesis: non-men like ramen *equal to* or *more* than men

Now we’re in busines

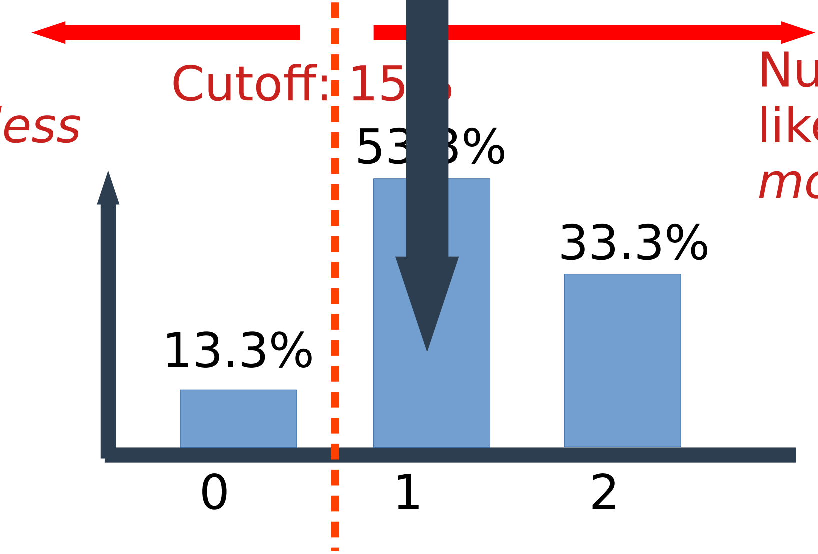
Like ramen?	Yes	No	TOTAL
	Is Male?		
Yes	5	1	6
No	3	1	4
TOTAL	8	2	10

Where does our statistic fall?

Hypothesis: Non-men like ram n less than men
Type 1 Error rate: 15% (usually ou should use at MOST 5%)

Alternate hypothesis: non-men like ramen *less* than men

Null hypothesis: non-men like ramen *equal to* or *more* than men



\# non-males who *like* ramen

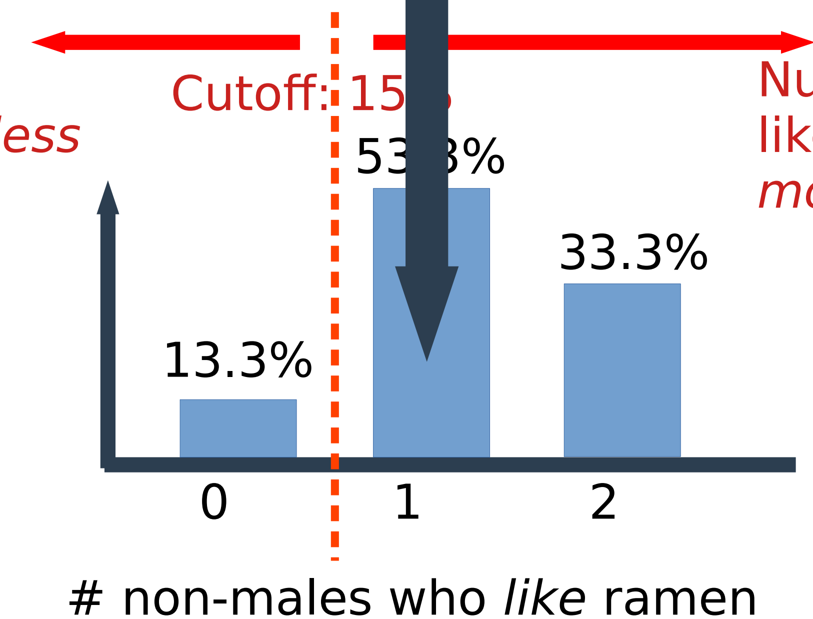
Now we’re in business.

So, we accept null...

Hypothesis: Non-men like ram n less than men
Type 1 Error rate: 15% (usually ou should use at MOST 5%)

Alternate hypothesis: non-men like ramen *less* than men

Null hypothesis: non-men like ramen *equal to* or *more* than men



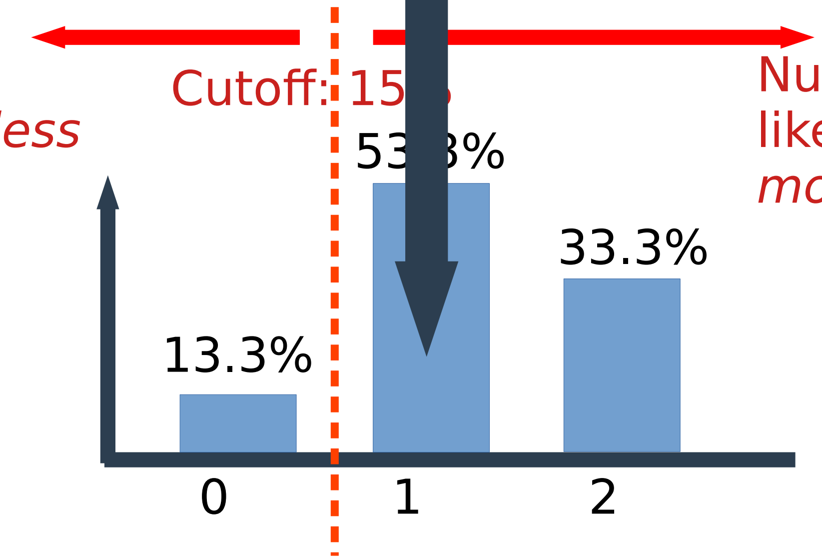
Now we’re in business.

This is called One-tailed test

Hypothesis: Non-men like ramen less than men
Type 1 Error rate: 15% (usually you should use at MOST 5%)

Alternate hypothesis: non-men like ramen *less* than men

Null hypothesis: non-men like ramen *equal to* or *more* than men



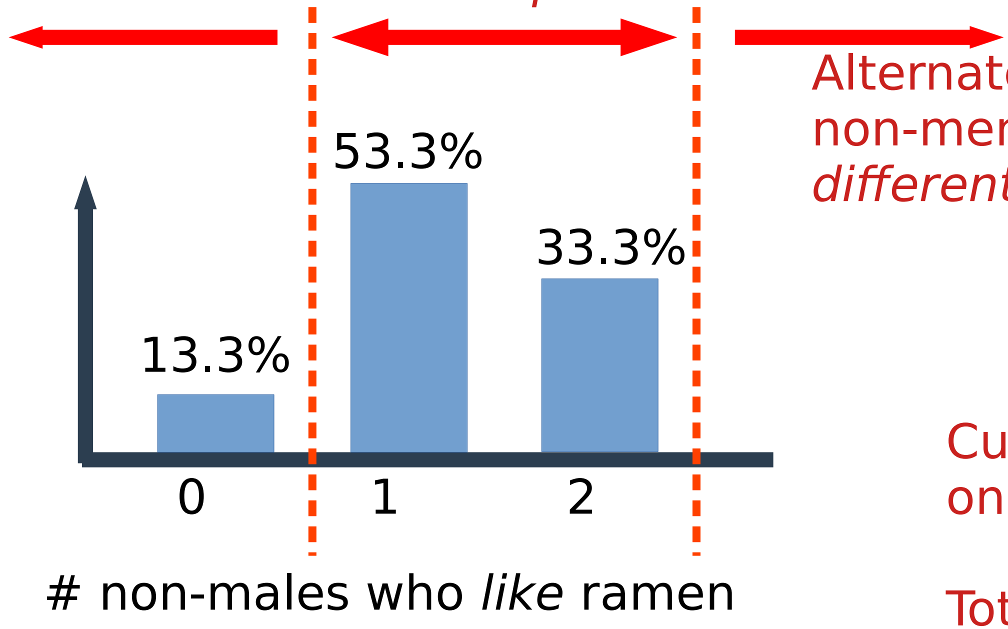
non-males who *like* ramen

Two-tailed test

What if we only care that they are different? → Two-tailed test

Null hypothesis: non-men like ramen *equal to men.*

Alternate hypothesis: non-men like ramen *different* than men



Alternate hypothesis: non-men like ramen *different* than men

Cutoff: 15% on each side

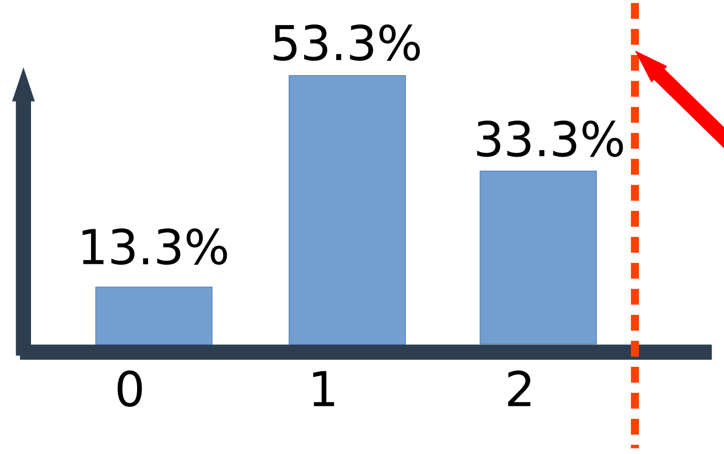
Total: 30%

Men like ramen more?

Men like ramen more? One-tailed in other direction…

Null hypothesis: non-men like ramen less then or equal to men

Alternate hypothesis: non-men like ramen *more than* men



non-males who *like* ramen

Cutoff so that 15% or less is above

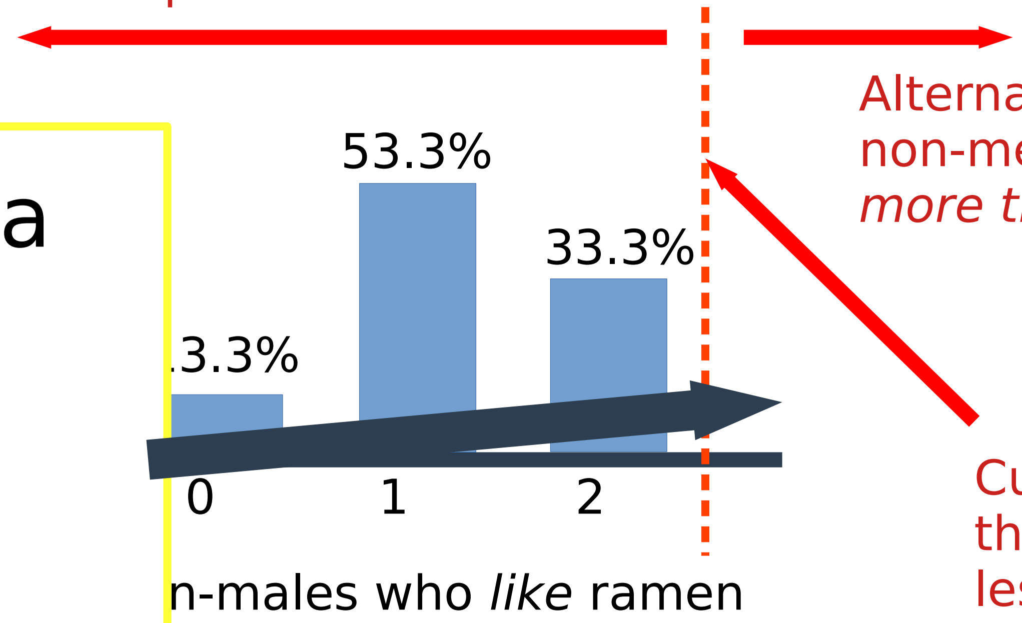
Men like ramen more?

Men like ramen more? One-tailed in other direction…

Null hypothesis: non-men like ramen less then or equal to men

Alternate hypothesis: non-men like ramen *more than* men

With our data size, we will never reject the null hypothesis…



Cutoff so that 15% or less is above

Summary: Fisher's Exact Test

1) Define your null hypotheses, decide whether you use 1-tailed versus 2-tailed test, and your alpha (cutoff) level (e.g., $\alpha=0.05$).

2) Collect data and create a 2x2 contingency table.

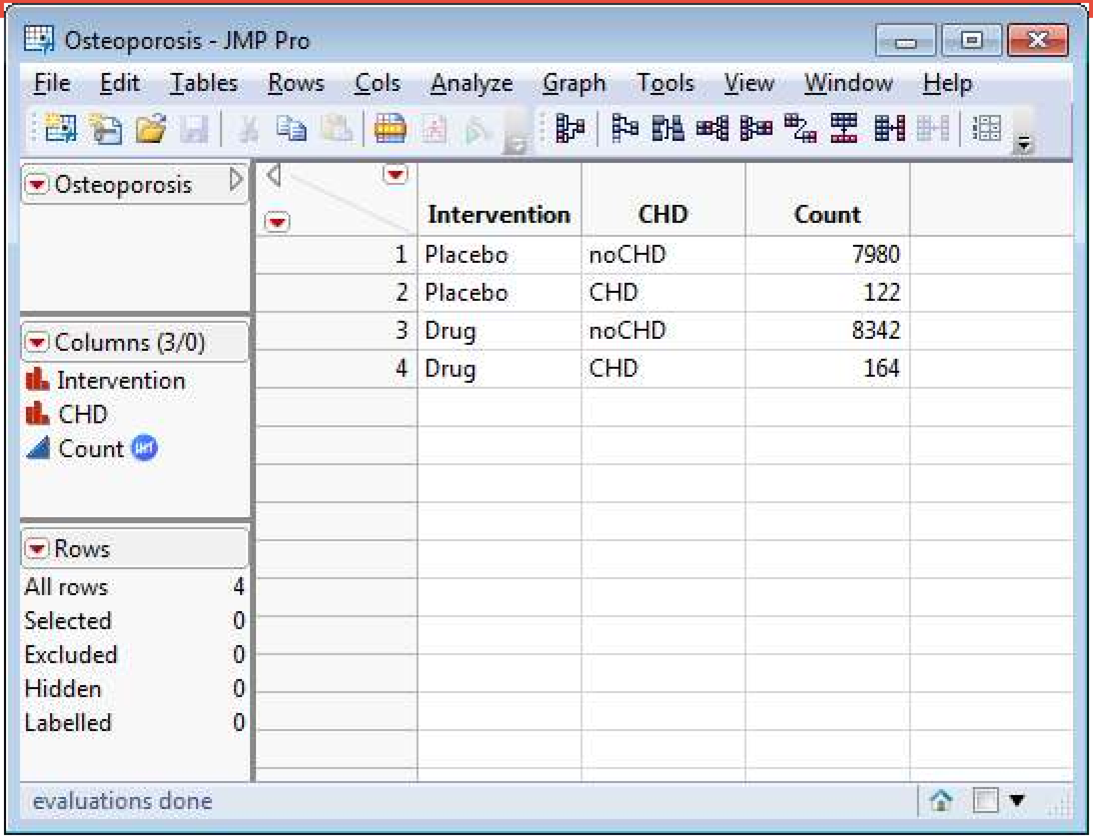
3) Using the hypergeometric distribution calculate the exact probabilities of $P(X=x)$ under the null hypothesis.

4) If under the null hypothesis the probability to observe a result as extreme or more extreme as the one observed is below your alpha level, reject the null hypothesis. Otherwise, we cannot reject the null hypothesis.

Fischer’s Exact Test in JMP

Fischer’s exact test and Chi-squared test are computed together…

Make a contingency table in JMP…

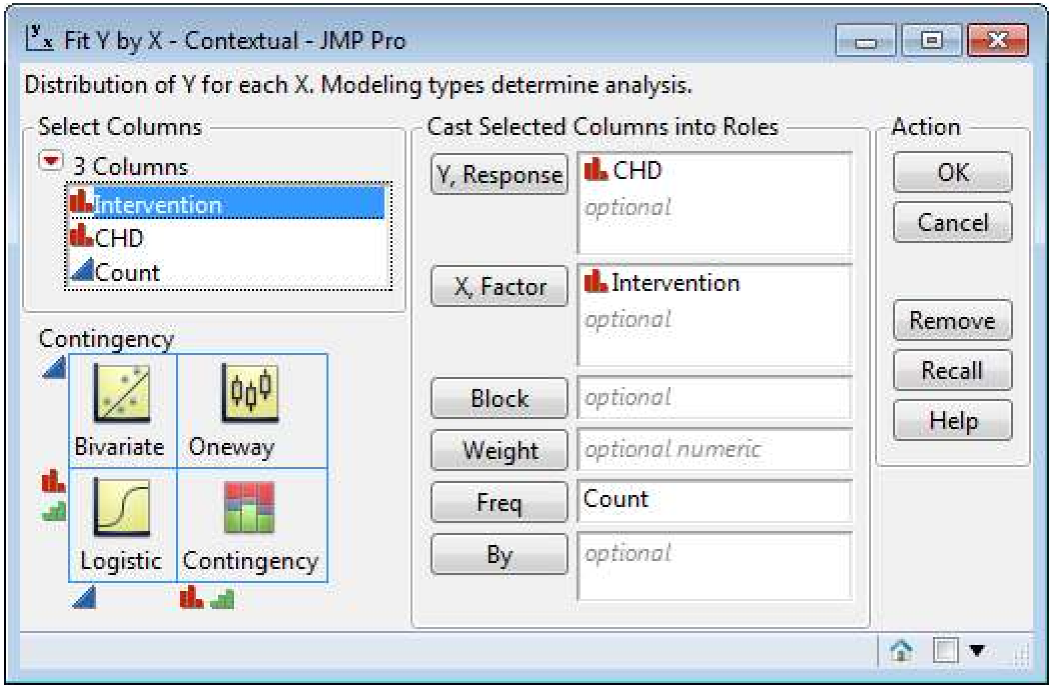


You can directly create a contingency table in JMP

→ Usually rows stand for individual cases/patients/participants.

For a 2x2 contingency we need a third column and do “Preselect Role”-> “Freq”.

Fisher’s exact test in JMP

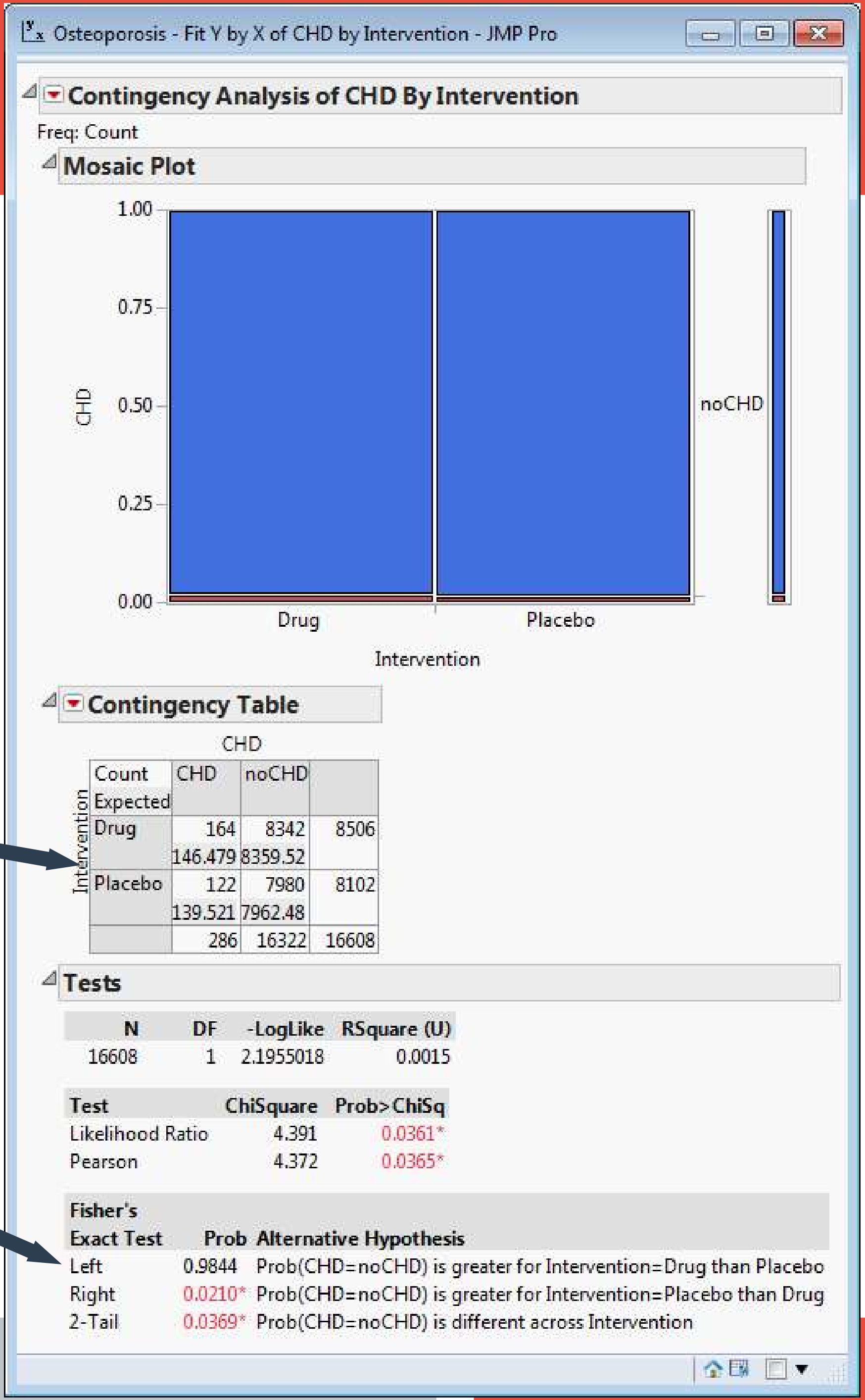


Under “Analyze”, we choose “Fit Y by X” and define the roles.

In JMP

We see our contingency table, we can display the expected values.

Below we see the result of the Fisher's exact test.



Count / Expected	CHD	noCHD	
Drug	164 146.479	8342 8359.52	8506
Placebo	122 139.521	7980 7962.48	8102
	286	16322	16608

N	DF	-LogLike	RSquare (U)
16608	1	2.1955018	0.0015

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	4.391	0.0361*
Pearson	4.372	0.0365*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9844	Prob(CHD=noCHD) is greater for Intervention=Drug than Placebo
Right	0.0210*	Prob(CHD=noCHD) is greater for Intervention=Placebo than Drug
2-Tail	0.0369*	Prob(CHD=noCHD) is different across Intervention

Fischer’s vs Chi-squared

Fischer’s exact test can always be used when you would use Chi-squared test (and is “more accurate”)

When you have many possibilities (or more than 2x2 design) it becomes too complex to compute Fischer’s exact…

So we “approximate” the exact distribution with a continuous one (the chi-squared distribution).