Topics covered in these slides

- Using real world datasets
- A simple way to make an Ogive, Ogives are good at displaying percentiles visually
- Explaining p-values by getting many samples and making an Ogive of p-values

Section 1
 Finding relationships in datasets with more than one column (more than one variable)

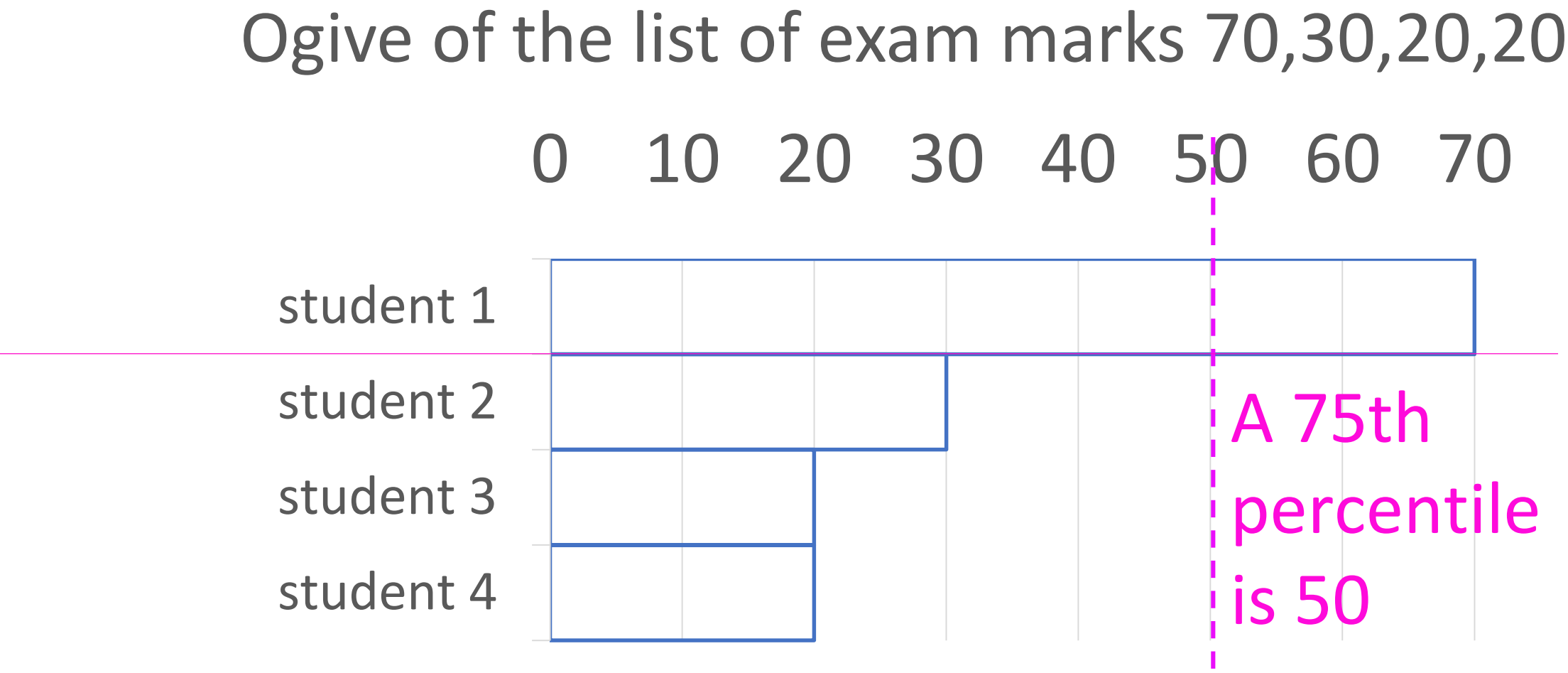
You can download this dataset from a government website, ALL properties first rented in NSW in 2018

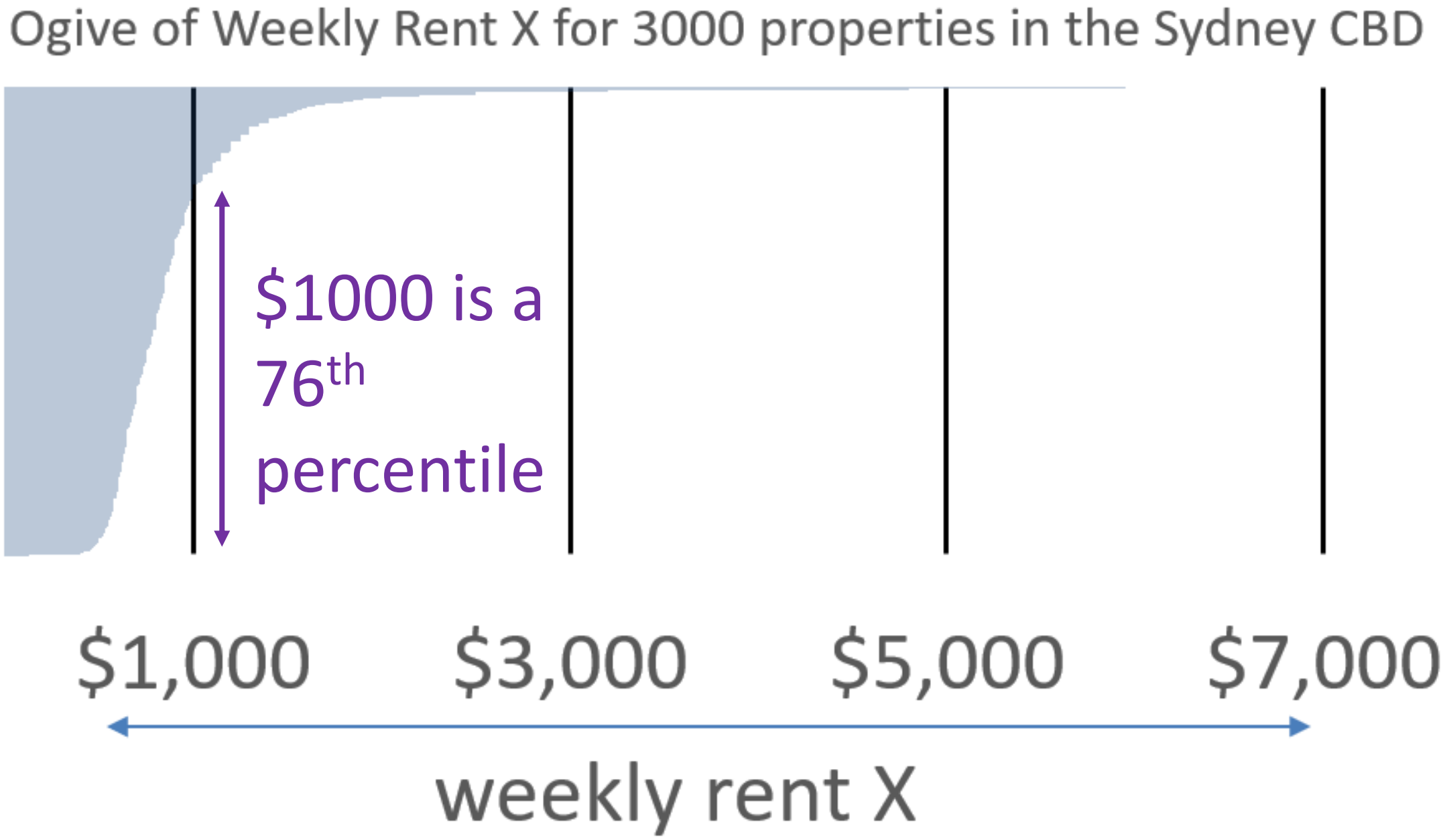
	A	B	D	E
1	Post code	Dwelling Type	Weekly Rent	In Sydney?
2	2007	F	$850	Sydney
3	2021	F	$760	Sydney
4	2106	F	$595	Not Sydney
5				dney
6	2145	F	$510	Sydney
262220	2850	H	$395	Not Sydney

OMG 260,000 rows

summary of all dwellings	F	H	total
Not Sydney			
Average of Weekly Rent	$345.73	$419.90	$394.86
Count of Dwelling Type	30,039	58,944	88,983
Count of Dwelling Type%	33.76%	66.24%	100.00%
Sydney			
Average of Weekly Rent	$578.69	$646.62	$600.03
Count of Dwelling Type	118,812	54,424	173,236
Count of Dwelling Type%	68.58%	31.42%	100.00%
Total Average of Weekly Rent	**$531.68**	**$528.74**	**$530.41**
Total Count of Dwelling Type	**148,851**	**113,368**	**262,219**
Total Count of Dwelling Type%	**56.77%**	**43.23%**	**100.00%**

You can easily make ogives in excel, just highlight a list of sorted numbers and insert a bar chart




Ogive of Weekly Rent X for 3000 properties in the Sydney CBD
$1000 is a 76th percentile
$1,000
$3,000
$5,000
$7,000
weekly rent X

By definition the ***p*-value** or **probability value** is the probability of obtaining test results at least as extreme as the results actually observed during the test, assuming that H_0 is correct

A good way of checking you understand this is proving that if H_0 is true and you get many samples and find the p-value for each sample then close to

25% of p-values will be less than 0.25

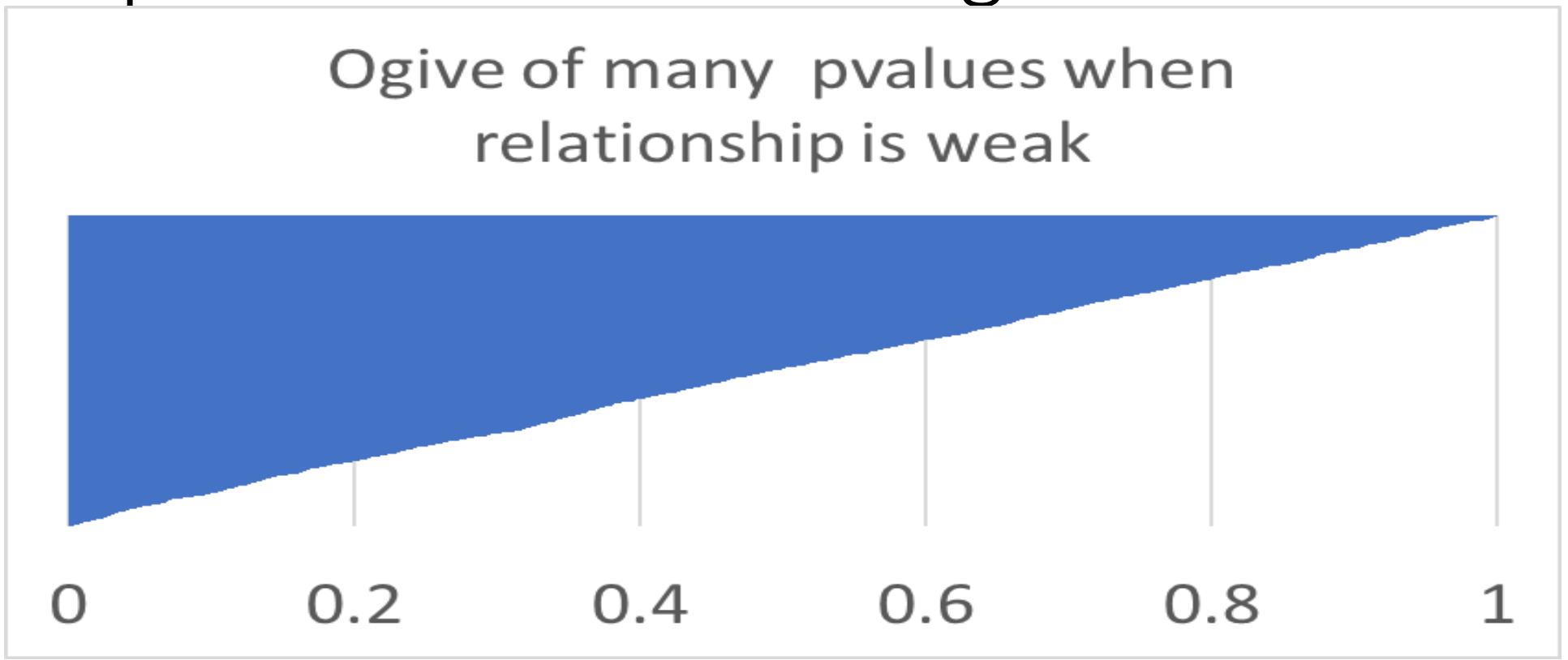
5% of p-values will be less than 0.05

K% of p-values will be than K/100

etc

How many lecturers understand this ??

If there is only a weak relationship the ogive of p-values will be a straight line.



To get the Ogive you need to find many samples and find the p-value for each sample

- If you have a sample you can measure evidence for claims about the population using a p-value
- Students do not need to learn the formulas they can use online calculators they just have to enter the sample statistics or the lecturer can make excel spreadsheets that automatically calculate them
- It is easy to use excel to get many samples and find the p-value for each sample

It is easy to split a population into many samples using excel, look at the following example

	A	B	C	D	E	F
1	Dwelling Type	Sample		Dwelling Type	Sample	rand()
2	Flat	sample 1		Flat	sample 1	0.15685
3	Flat	sample 1		Flat	sample 2	0.26605
4	Flat	sample 2		Flat	sample 2	0.39724
5	Flat	sample 2		Flat	sample 4	0.40355
6	Flat	sample 3		Flat	sample 3	0.83585
7	Flat	sample 3		Flat	sample 3	0.91367
8	House	sample 4		House	sample 4	0.9664
9	House	sample 4		House	sample 1	0.98886

If you understand this dataset you only expect the columns A and B to be related

	A	B	C	D
1	Dwelling type?	In Sydney?	which half?	which sample ?
2	F	Sydney	1st	sample 1193
3	F	Sydney	1st	sample 1309
127626	F	Sydney	1st	sample 949
127627	F	Sydney	2nd	sample 576
127628	F	Sydney	2nd	sample 754
262220	H	Not Sydney	2nd	sample 1335

You should think column C is not related to anything

	A	B	C
1	Dwelling type?	In Sydney?	which half?
2	F	Sydney	1st
3	F	Sydney	1st
127626	F	Sydney	1st
127627	F	Sydney	2nd
127628	F	Sydney	2nd
262220	H	Not Sydney	2nd

In row 2 to row 127626 the bond was lodged in the 1st half of 2018

There is clearly a relationship between dwelling type and location .

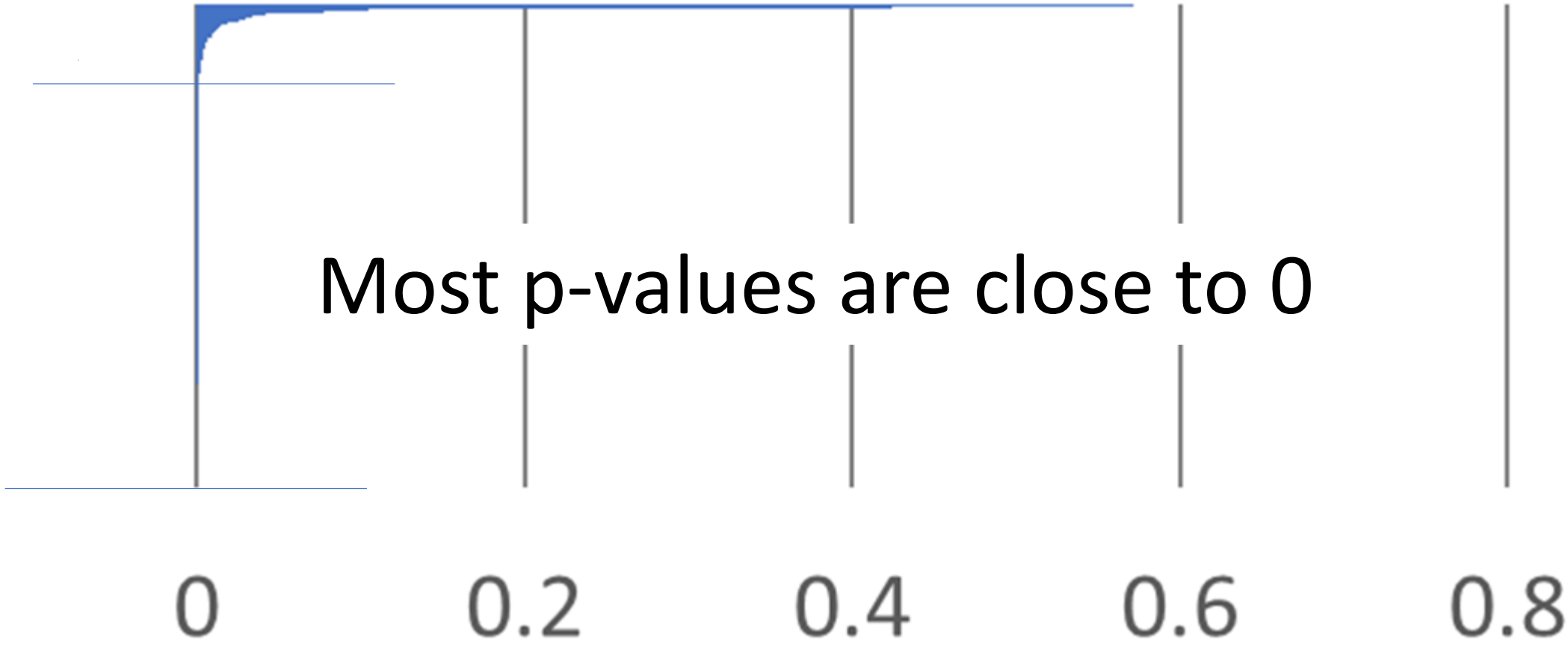
summary	F	H	Grand Total
Not Sydney			
Count	30,039	58,944	88,983
%	33.76%	66.24%	100.00%
Sydney			
Count	118,812	54,424	173,236
%	68.58%	31.42%	100.00%
Total Count	**148,851**	**113,368**	**262,219**
Total %	**56.77%**	**43.23%**	**100.00%**

You can get a single sample and find pvalue

summary	F	H	Grand Total
Not Sydney			
Count	10	21	31
%	32.26%	67.74%	100.00%
Sydney			
Count	47	22	69
%	68.12%	31.88%	100.00%
Total Count	**57**	**43**	**100**
Total %	**57.00%**	**43.00%**	**100.00%**

Entering the numbers into an online calculator you get the p-value 0.0008

An ogive of pvalue when there is a strong relationship



Note that Different samples give different answers

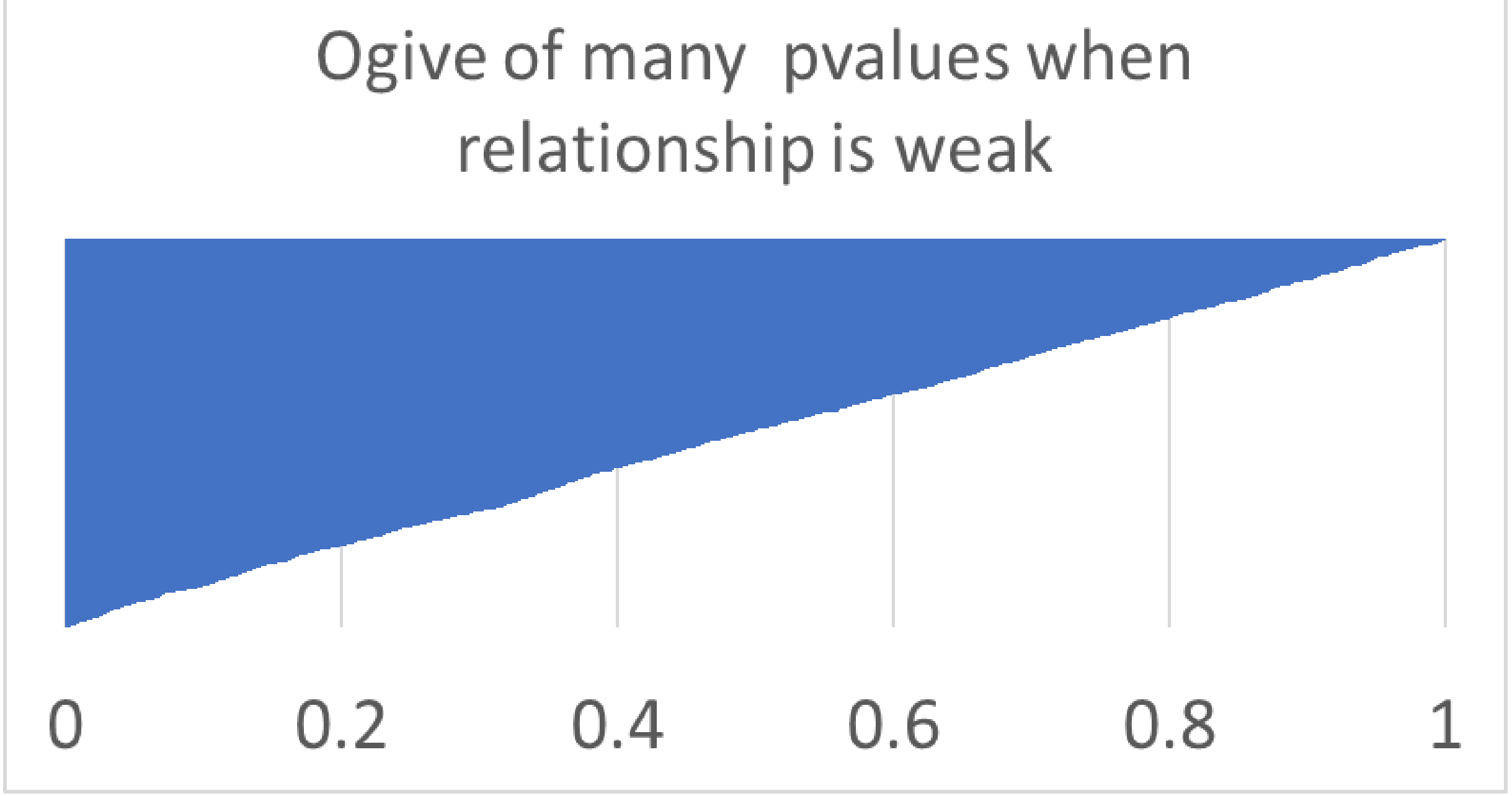
Looking at this summary there is not much of a difference between %s so there is a weak relationship

summary	1st	2nd	Total
Not Sydney			
Count	44679	44304	88983
%	50.21%	49.79%	100.00%
Sydney			
Count	82946	90290	173236
%	47.88%	52.12%	100.00%
Total Count	**127625**	**134594**	**262219**
Total %	**48.67%**	**51.33%**	**100.00%**

You can get a single sample and find the p-value

summary	Column Labels 1st	2nd	Total
Not Sydney			
Count	17	14	31
%	54.84%	45.16%	100.00%
Sydney			
Count	33	36	69
%	47.83%	52.17%	100.00%
Total Count	**50**	**50**	**100**
Total %	**50.00%**	**50.00%**	**100.00%**

Entering the numbers into an online calculator you get the p-value 0.5166


Ogive of many pvalues when relationship is weak
0
0.2
0.4
0.6
0.8
1

Example of a population with a strong relationship

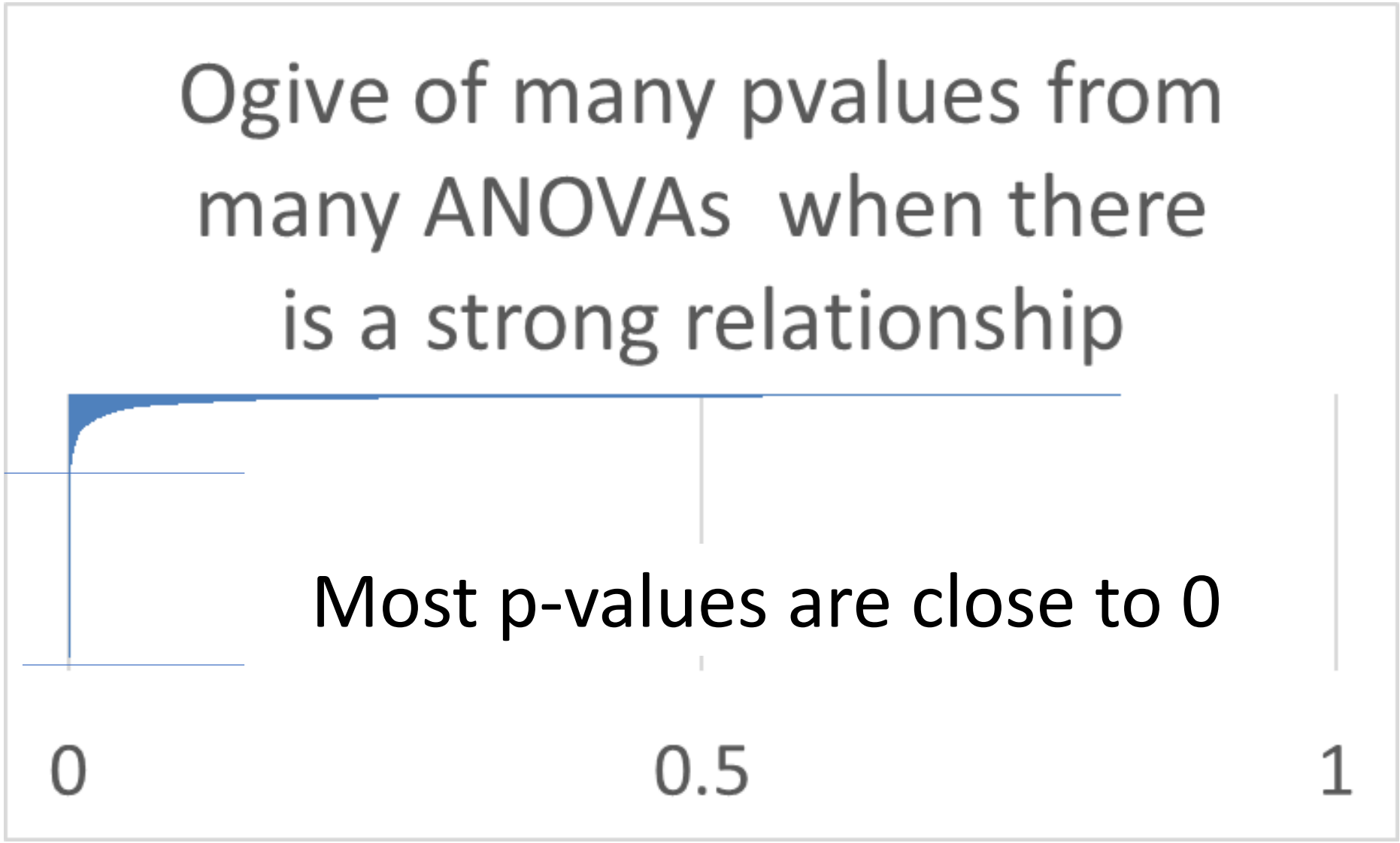
summary	Average of Rent	StdDev	count
NSF	$345.73	$164.17	30,039
NSH	$419.90	$192.76	58,944
SF	$578.69	$238.17	118,812
SH	$646.62	$358.03	54,424

Note: NSF=NonSydney Flat , SH= Sydney House

You can get a single sample and find p-value

Summary	Average Rent	Stdev	count
NSF	$366.36	$109.46	11
NSH	$470.48	$235.25	21
SF	$605.31	$316.84	49
SH	$684.74	$246.70	19
Total (ALL)	**$565.80**	**$286.29**	**100**

Entering the numbers into an online calculator you get the p-value 0.006


Ogive of many pvalues from many ANOVAs when there is a strong relationship
Most p-values are close to 0
0
0.5
1

Example of a weak relationship, note that the averages are very close, Close to $530

summary	1stF	1stH	2ndF	2ndH
Average of Weekly Rent	$534.00	$526.97	$529.52	$530.46
StdDev of Weekly Rent	$249.25	$302.40	$238.68	$309.61
Count of Weekly Rent	71,728	55,897	77,123	57,471

Note that 1st F is a Flat where the bond was lodged in the 1st half of the year

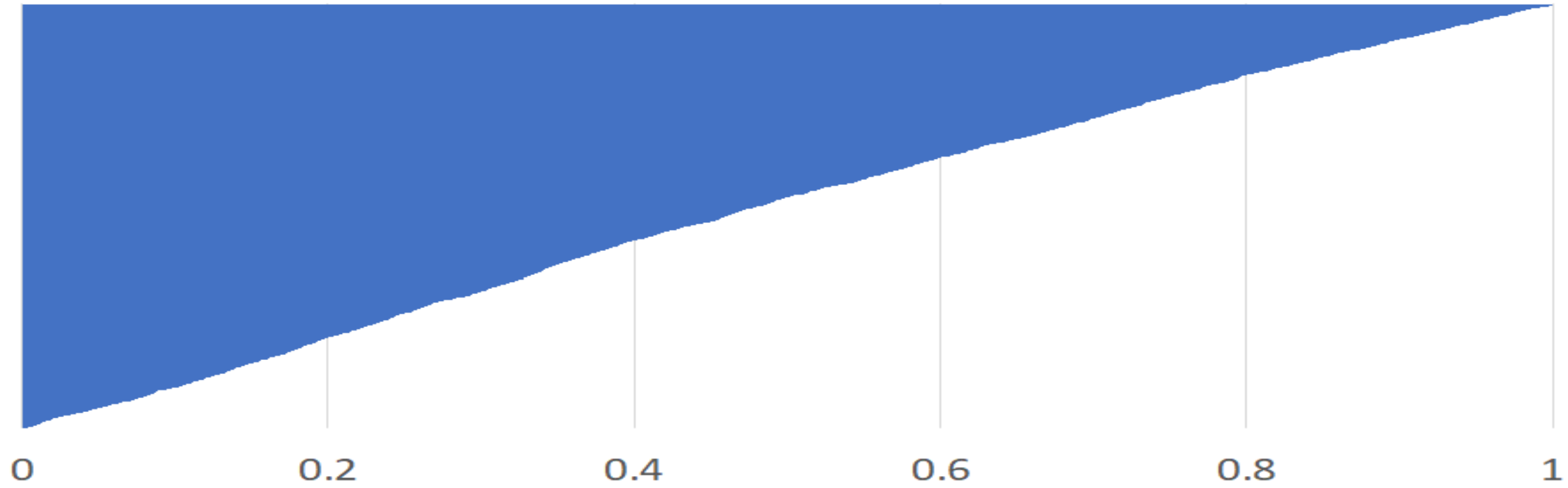
2nd H is a House and the bond was lodged in the 2nd half of the year

You can get a single sample and find a p-value

summary	Average rent	stdev	Count
1stF	$631.71	$363.19	35
1stH	$543.25	$233.32	20
2ndF	$463.20	$151.68	25
2ndH	$601.25	$289.49	20
Total(All)	$565.80	$286.29	100

Entering the numbers into an online calculator you get the p-value 0.138

The ogive of p-value when there is a weak relationship in the population, It was easy to split the population into 2600 samples and get the ANOVA p-value for each sample



Section 2

Working with a single variable (A dataset with a single column)

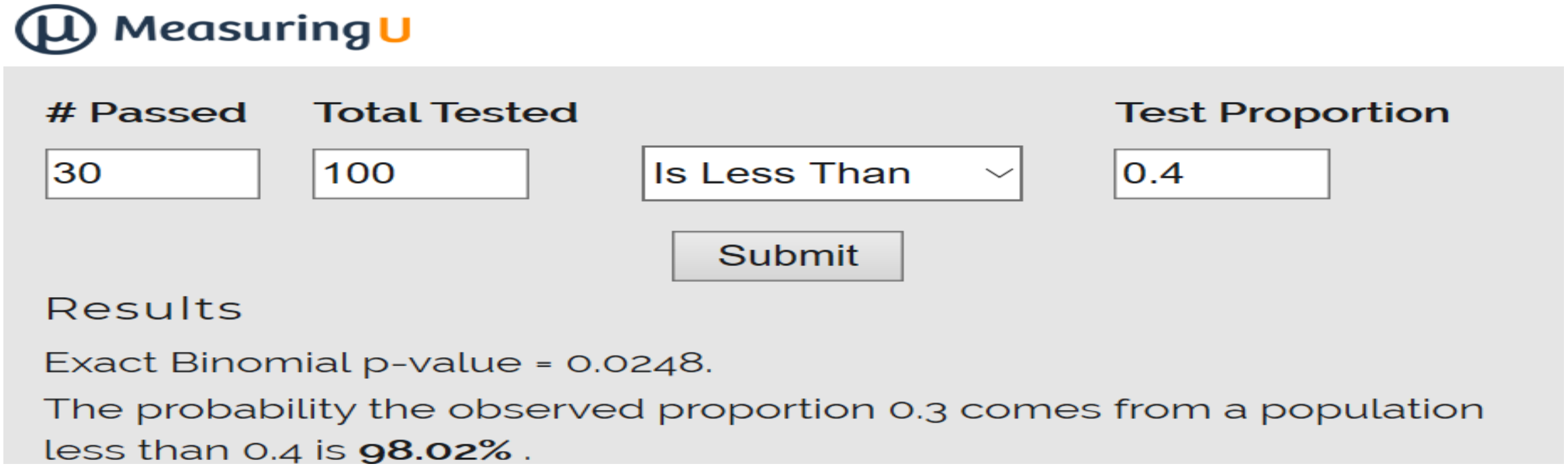
The normal way of introducing p-values using a single variable can be made even easier by

Not talking about H0 and H1

Not talking about probability, including the normal distribution

Students do not need to learn p-value formulas they can use a webpage

https://measuringu.com/onep/



Repeating slide 7: By definition the ***p*-value** or **probability value** is the probability of obtaining test results at least as extreme as the results actually observed during the test, assuming that the H_0 is correct

So if H_0 is true and you get many samples and find the p-value for each sample

25% of p-values are less than 0.25

5% of p-values are less than 0.05

K% of p-values are less than K/100

How many lecturers understand this ??

The next 5 slides will explain p-value by making an Ogive of p-values after investigating the accuracy of sampling by using a computer to find many samples.

- Suppose you are a politician and there is an election today and your “leader” wins by getting 52% of the vote.

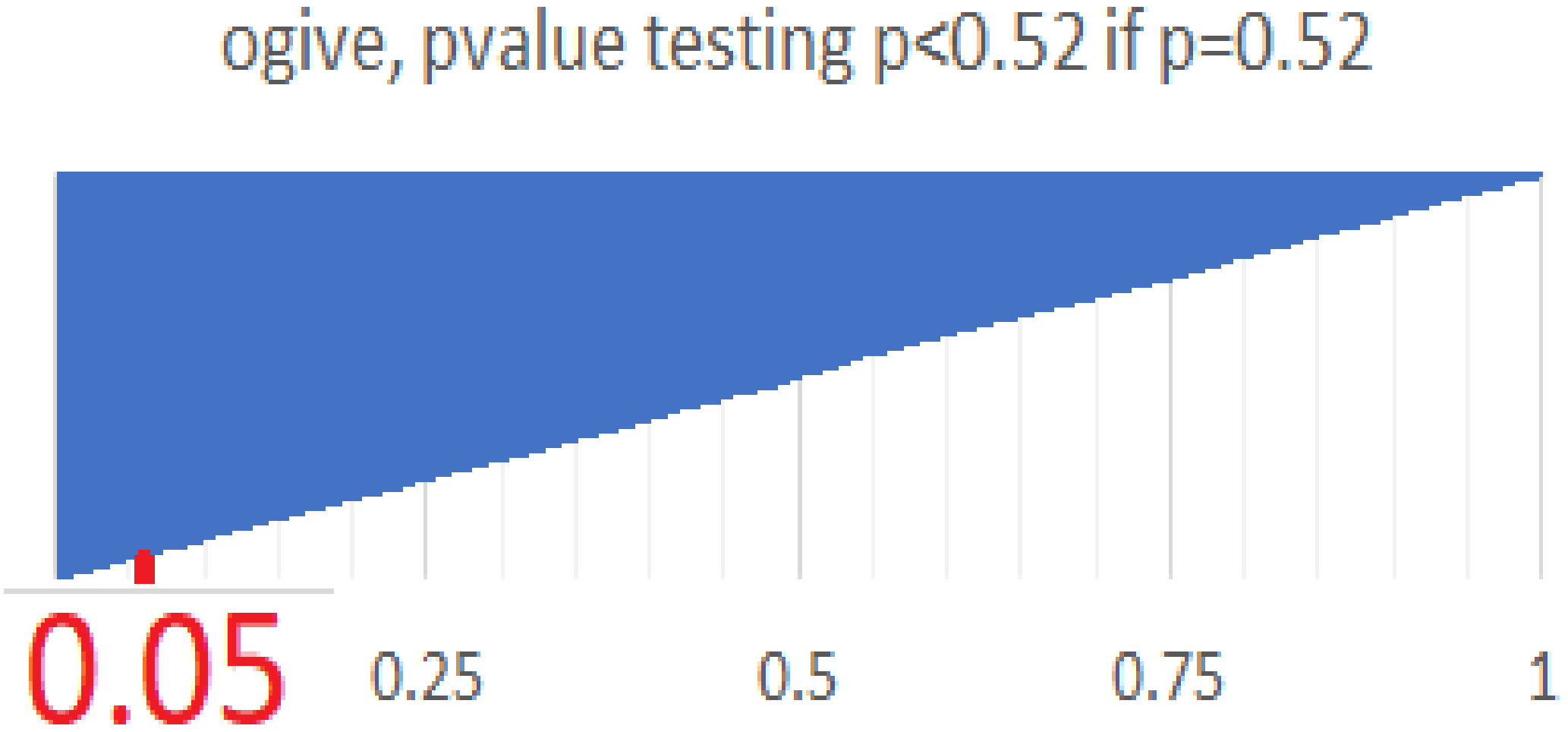
- It is very likely that you will want to abandon your leader in the future if (when) they do something silly that causes a drop in support

- You can get many samples today using a computer using population proportion is p=0.52

Which sample	no	yes	Sample size	Sample proportion	p-value testing claim "p<0.52"
Remember a computer found many samples using p=0.52					
sample 8	62	38	100	0.38	0.002537434
sample 2	55	45	100	0.45	0.080588895
sample 3	52	48	100	0.48	0.211669821
sample 5	49	51	100	0.51	0.42067765
sample 9	49	51	100	0.51	0.42067765

Which sample	no	yes	Sample size	Sample Proportion	p-value testing p<0.52
Remember a computer found many samples using p=0.52					
sample 67	53	53	106	0.5	0.340112
sample 102	55	55	110	0.5	0.337294
sample 122	56	56	112	0.5	0.335907
sample 3521	226	226	452	0.5	0.197358
sample 3543	227	227	454	0.5	0.196836

A computer was used to find 6000 samples, the p-value for each sample was calculated and summarized in the ogive below



Based on the students reactions to the fact the ogive of p-values is a straight line they can tell actually calculating p-values must be tricky because something must be making the ogive of p-value really simple, Basically a genius can imagine an infinite number of samples and work out the percentiles just using advanced maths

Usually if you have many variables you NEED to use a computer to find the p-values so the exact details are not important

There are many hypothesis tests much more complicated than testing if $p<0.52$ but you can still explain them the same way

*Use a computer to check the accuracy of taking samples by getting many samples and get the p-value for each sample, The ogive of p-values will be a straight line

*Show the students different samples , they will notice the lower the p-value the stronger the evidence

3 benefits of explaining p-values by using an Ogive of p-values

Benefit 1:

It uses the concept "an ogive lets you see percentiles" instead of the concept "probability", you need a very strong background in probability to understand what a p-value is

3 benefits of explaining p-values by using an Ogive of p-values

Benefit 2:

It uses concept of “checking the accuracy of samples by getting many samples” instead of the concept the null hypothesis H_0

3 benefits of explaining p-values by using an Ogive of p-values

Benefit 3:

It is similar to a real world examples that use things that are much simpler than p-values

In the previous slides explain the following

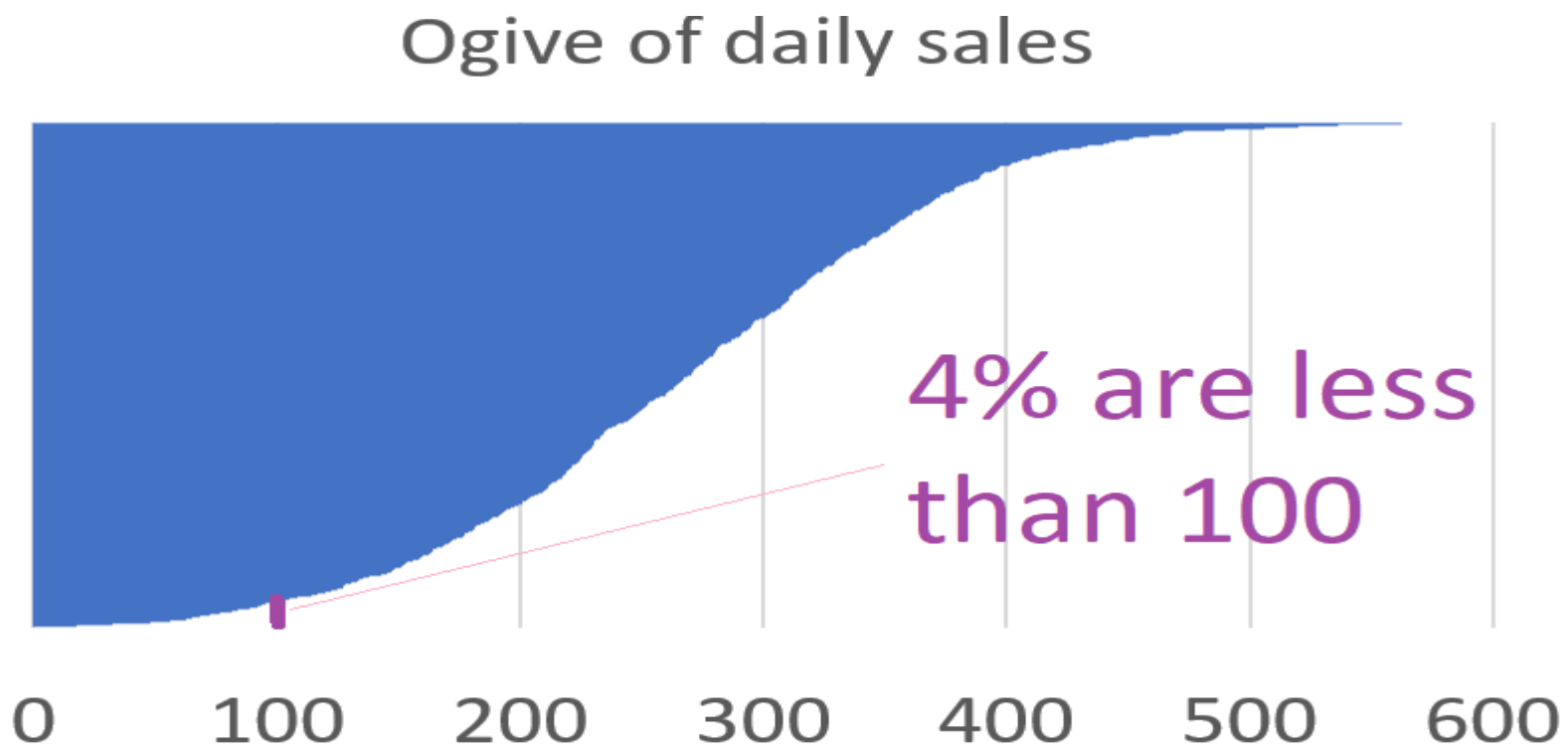
If a Political leader does something annoying you can get a single sample and use a single p-value to answer the question is “Has their popularity dropped”

The ogive of p-values based on samples before the leader did something annoying helped you interpret the p-value

The following situation will be explained by an Ogive like p-values are explained by an Ogive

"If someone in HR says something stupid on social media you have evidence sales have decreased if they unusually low the next day"

Suppose Daily sales were 100 the day after someone in HR said something stupid on social media, Based on the Ogive below you do have evidence sales has dropped.



The “there is strong evidence the sales have decreased example” given on the previous slide is not about p-values which are complex it was about daily sales

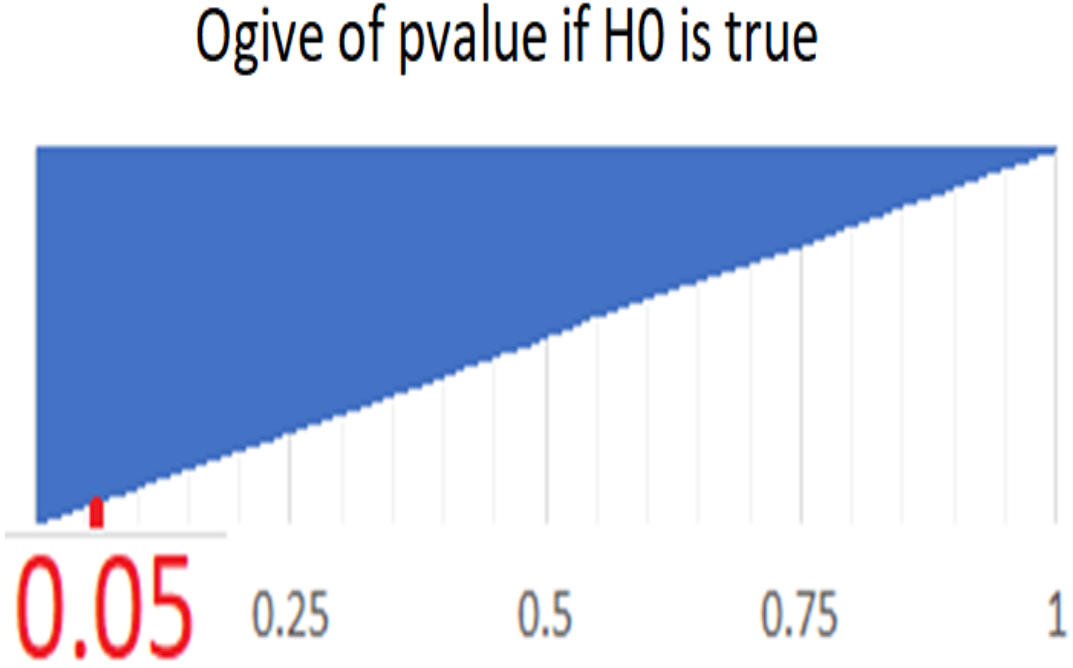
This example is useful because it shows you can use a Ogive of many values in the past to work out if a single value in the future is strong evidence that things have changed.

An ogive a past values clearly show what values were unusually low in the past

If you are using a p-value to measure evidence of a relationship an ogive of p-values when there is no relationship clearly shows what p-values are unusually low when there is no relationship

Revision of slide 7, whenever you use p-value

- If you get many samples when H_0 is true and you find the p-value for each sample, the ogive of p-value will look like



75% of p-values are less than 0.75
50% of p-values are less than 0.5
25% of p-values are less than 0.5
K% of p-values are less than K/100

Do you remember the definition of pvalue? Does the Ogive help explain it ?

Summary

A low p-value is like something unusually low that is strong evidence

Most p-values are used to measure evidence of relationships, when there is a weak* relationship in the population the Ogive of many p-values will be a straight line and most pvalues will be around 0.5 , If there is a strong relationship most pvalues will be close to 0

A relationship is weak* if the typical standard error is large enough to usually hide the relationship

Section 3 General tips on teaching inference

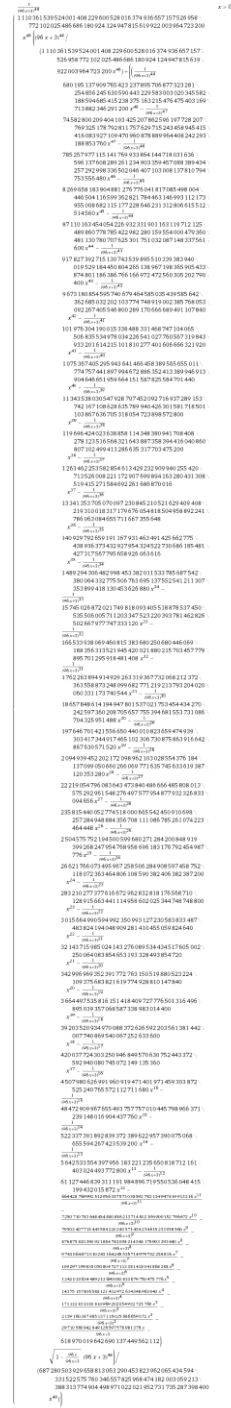
Some textbook encourage teaching inference using simulation. Personally I am sure the vast majority of students do not understand inference until they see many samples and notice how different samples give different answers. Many students can do textbook questions without understanding p-values are a way of dealing with the problems samples have standard error.

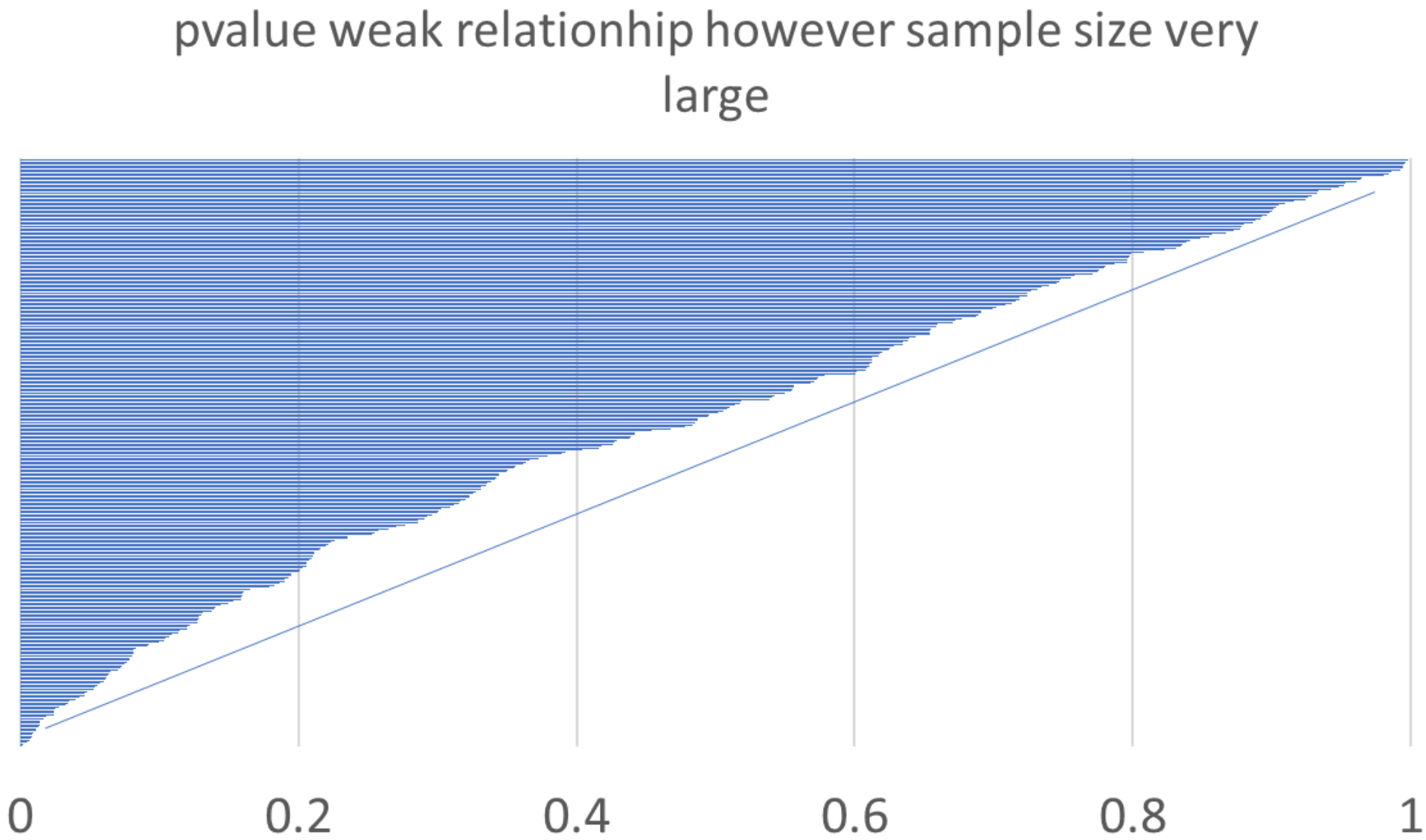
Since you need to get many samples to teach inference you also should find the p-value for each sample as well and make the ogive of p-values.

Without too much effort you can roughly explain p-value without using probability or defining H_0 and H_1.

Section 4:Additional Material

Formula for cdf of the F distribution, This was used to get the p-value




pvalue weak relationhip however sample size very large
0
0.2
0.4
0.6
0.8
1

Consider following dataset which column is unrelated to the other columns

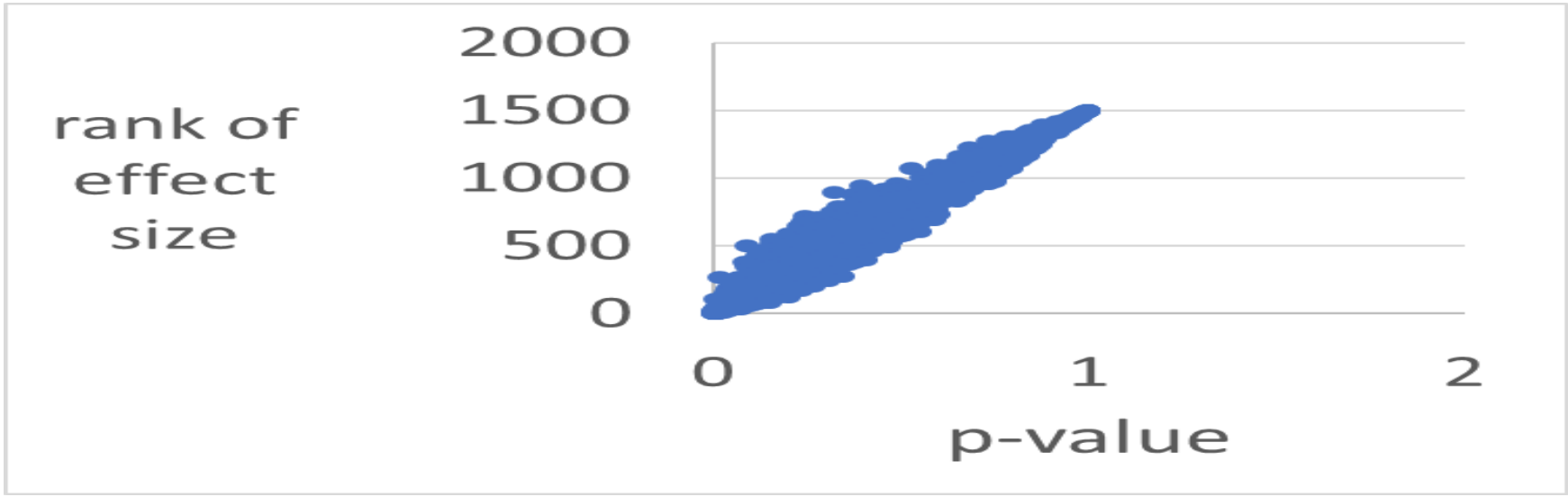
	A	B	C
1	**Dwelling Type?**	In Sydney?	coin toss?
2	F	Sydney	head
3	F	Sydney	head
4	F	Not Sydney	head
5	F	Sydney	tail
6	F	Sydney	head
262220	H	Not Sydney	tail

As you would expect there is no relationship between the variable dwelling type and coin toss

summary	F	H	Grand Total
head			
Count	74,552	56,617	131,169
%	56.84%	43.16%	100.00%
tail			
Count	74,299	56,751	131,050
%	56.70%	43.30%	100.00%
Total Count	**148,851**	**113,368**	**262,219**
Total %	**56.77%**	**43.23%**	**100.00%**

If sample size is always the same there will always be a linear relationship between rank of p-value and rank of effect size

and if H0 is true then rank of effect size will be proportional to p-value



*If you integrate a pdf you get a cdf

*A histogram is like a pdf you need the shaded area to find the percentile

*An Ogive is like a cdf , you find percentile using the height

References

Further reading

- Perezgonzalez, J. D. (2015). P-values as percentiles. Commentary on:“Null hypothesis significance tests. A mix–up of two different theories: the basis for widespread confusion and numerous misinterpretations”. *Frontiers in psychology*, *6*, 341.

A website about teaching statistics ,

https://www.causeweb.org/

You can search the site for simulation there are plenty of resources explaining the benefits of explaining inference Using simulation