

# Stats - Fundamentals of hypothesis testing Part 1

[anthony.smith@scu.edu.au](mailto:anthony.smith@scu.edu.au)

# Record Attendance

YOU SHOULD COME TO  
OUR LINGUISTICS CLUB'S  
SESQUIANNUAL MEETING.

MEMBERSHIP IS OPEN TO  
ANYONE WHO CAN FIGURE  
OUT HOW OFTEN WE MEET.



# Any Questions About Any Maths At All?





# Do Most Of Us Usually Arrive On Time?

- Last time: how we can estimate the mean arrival time of students to this tutorial?
- New question: do most of us usually arrive on time?
- How do we answer this question? What's the procedure?

# Do Most Of Us Usually Arrive On Time?

- Null hypothesis: most of us usually arrive on time
- In a form we can test: the mean arrival time is before the beginning of the tutorial
- The sample statistic corresponding to the population mean arrival time is the average arrival time to this tutorial
- Null hypothesis says mean arrival time is before start time

# Do Most Of Us Usually Arrive On Time?

- Null hypothesis says our sample mean arrival time should be at or before start time
- At 10% level the rejection region is the 10% area in the upper tail
- Does our sample mean fall into this region?
- For the sampling distribution: must estimate standard deviation from sample, sample is small ( $n < 30$ ) so use t distribution
- Using the appropriate row of the t-table, find critical t-value for limit of rejection region, compare with sample mean's t-value



# Review the Lecture

MISS LENHART COULDN'T BE HERE TODAY, SO SHE ASKED ME TO SUBSTITUTE.

MATH

MR. MUNROE

I'VE PUT OUT YOUR TESTS. PLEASE GET STARTED.

MR. MUNROE, MISS LENHART NEVER TAUGHT US THIS.

THAT'S BECAUSE MISS LENHART DOESN'T UNDERSTAND HOW IMPORTANT CERTAIN KINDS OF MATH ARE.

BUT THIS JUST LOOKS --

THIS MATERIAL IS MORE VITAL THAN ANYTHING YOU'VE EVER LEARNED

BUT --

NO BUTS.

THIS IS A MATTER OF LIFE AND DEATH.

Name: \_\_\_\_\_

1. The velociraptor spots you 40 meters away and attacks, accelerating at  $4 \text{ m/s}^2$  up to its top speed of  $25 \text{ m/s}$ . When it spots you, you begin to flee, quickly reaching your top speed of  $6 \text{ m/s}$ . How far can you get before you're caught and devoured?

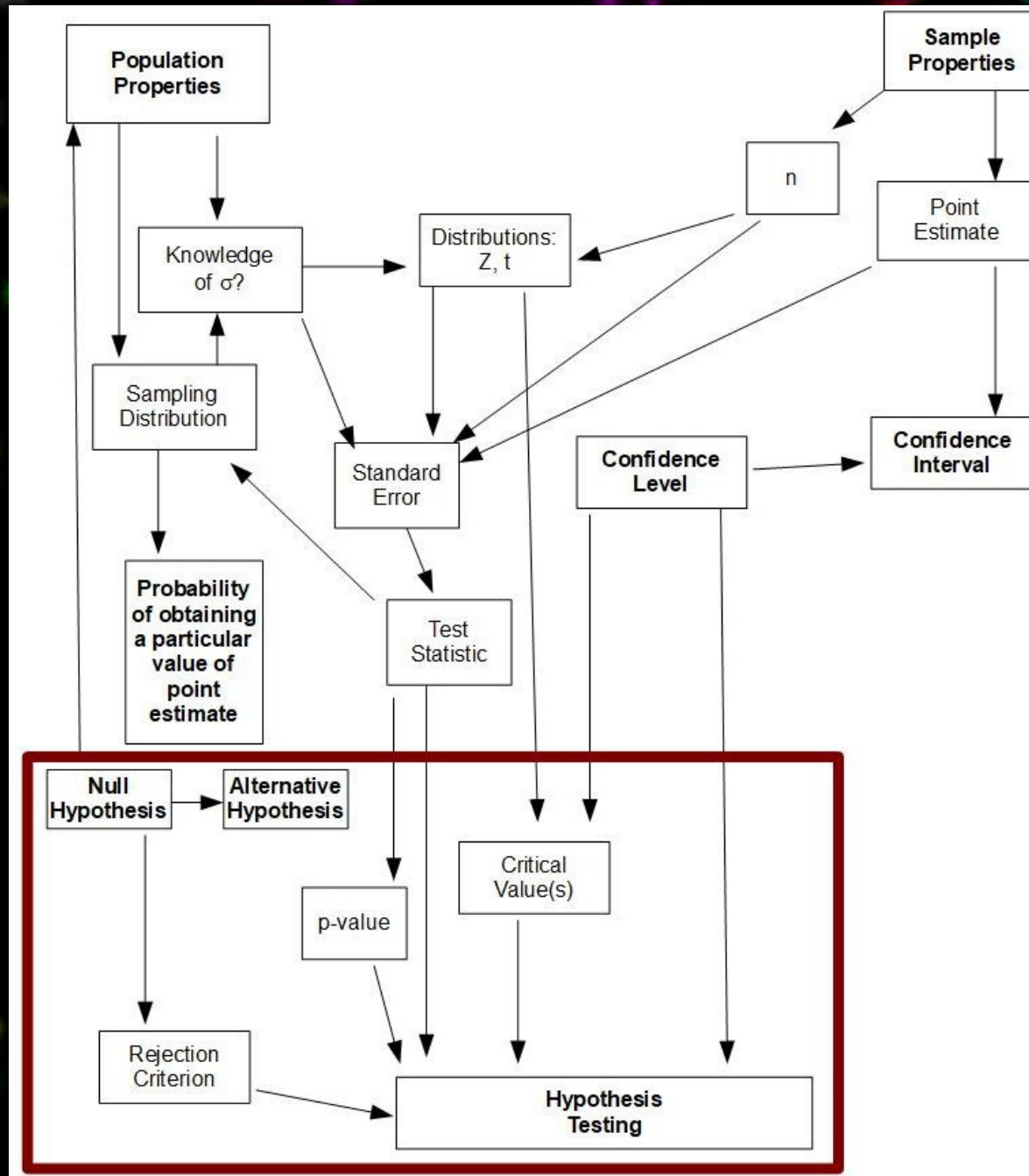


# Do Most Of Us Usually Arrive On Time?

- Null hypothesis says our sample mean arrival time should be at or before start time
- At 10% level the rejection region is the 10% area in the upper tail
- Does our sample mean fall into this region?
- For the sampling distribution: must estimate standard deviation from sample, sample is small ( $n < 30$ ) so use t distribution
- Using the appropriate row of the t-table, find critical t-value for limit of rejection region, compare with sample mean's t-value

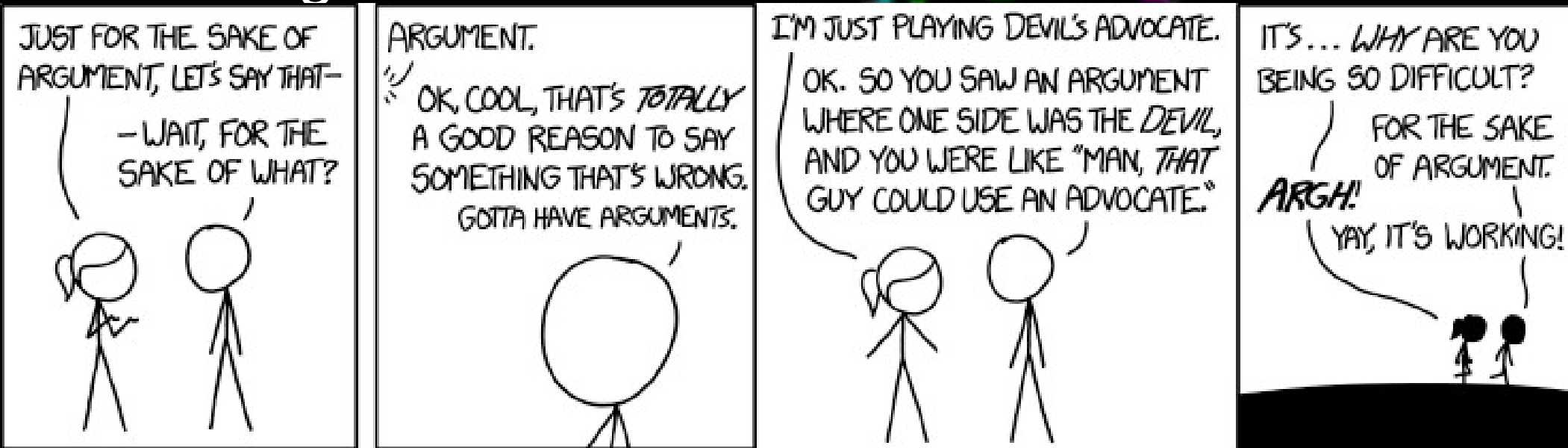


# Where Are We Now?

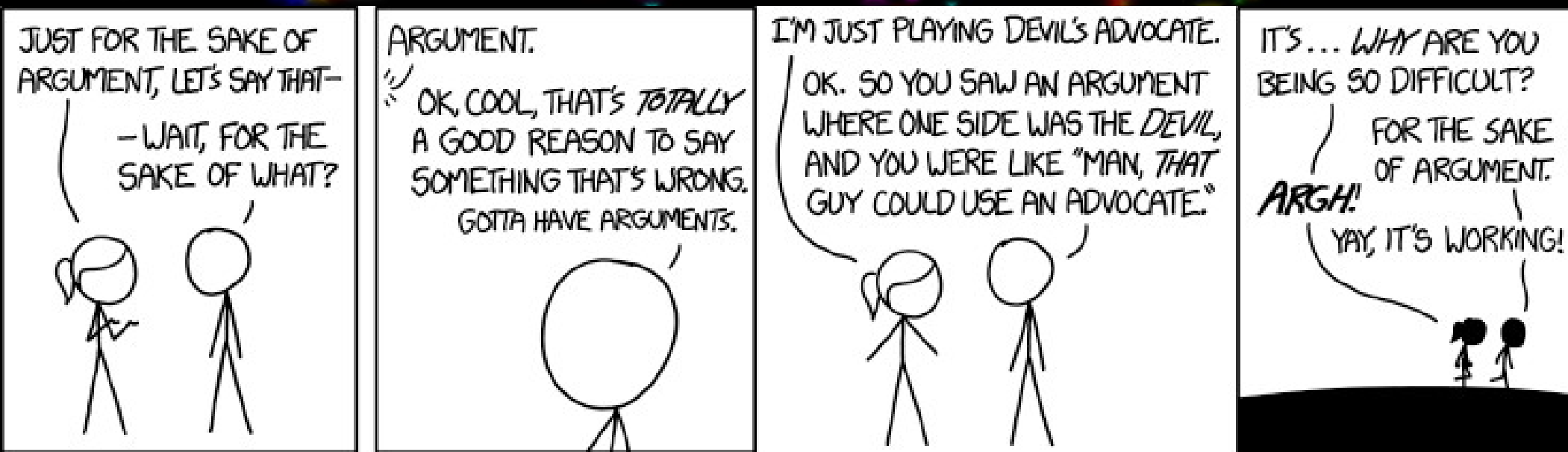


# The Sake of Argument

- What does this have to do with hypothesis testing?



# The Sake of Argument



*Title text: 'It's not actually... it's a DEVICE for EXPLORING a PLAUSIBLE REALITY that's not the one we're in, to gain a broader understanding about it.' 'oh, like a boat!' '...' 'Just for the sake of argument, we should get a boat! You can invite the Devil, too, if you want.'*



# The Meaning of Statistical Significance – '*Show me you're real*'

- Statisticians don't accept results at face value
- A result might not be “real” or “statistically-significant”
- '*Crack, baby crack – show me you're real*'
  - “Cracked Actor” (Aladdin Sane) – David Bowie
- Recall the arrival time example: why do we not necessarily reject the null hypothesis (that we mostly arrive on time) if the sample mean arrival time is after the tutorial start time?

# “Significant” – Part 1

JELLY BEANS  
CAUSE ACNE!

SCIENTISTS!  
INVESTIGATE!

BUT WE'RE  
PLAYING  
MINECRAFT!  
... FINE.



WE FOUND NO  
LINK BETWEEN  
JELLY BEANS AND  
ACNE ( $p > 0.05$ ).



THAT SETTLES THAT.

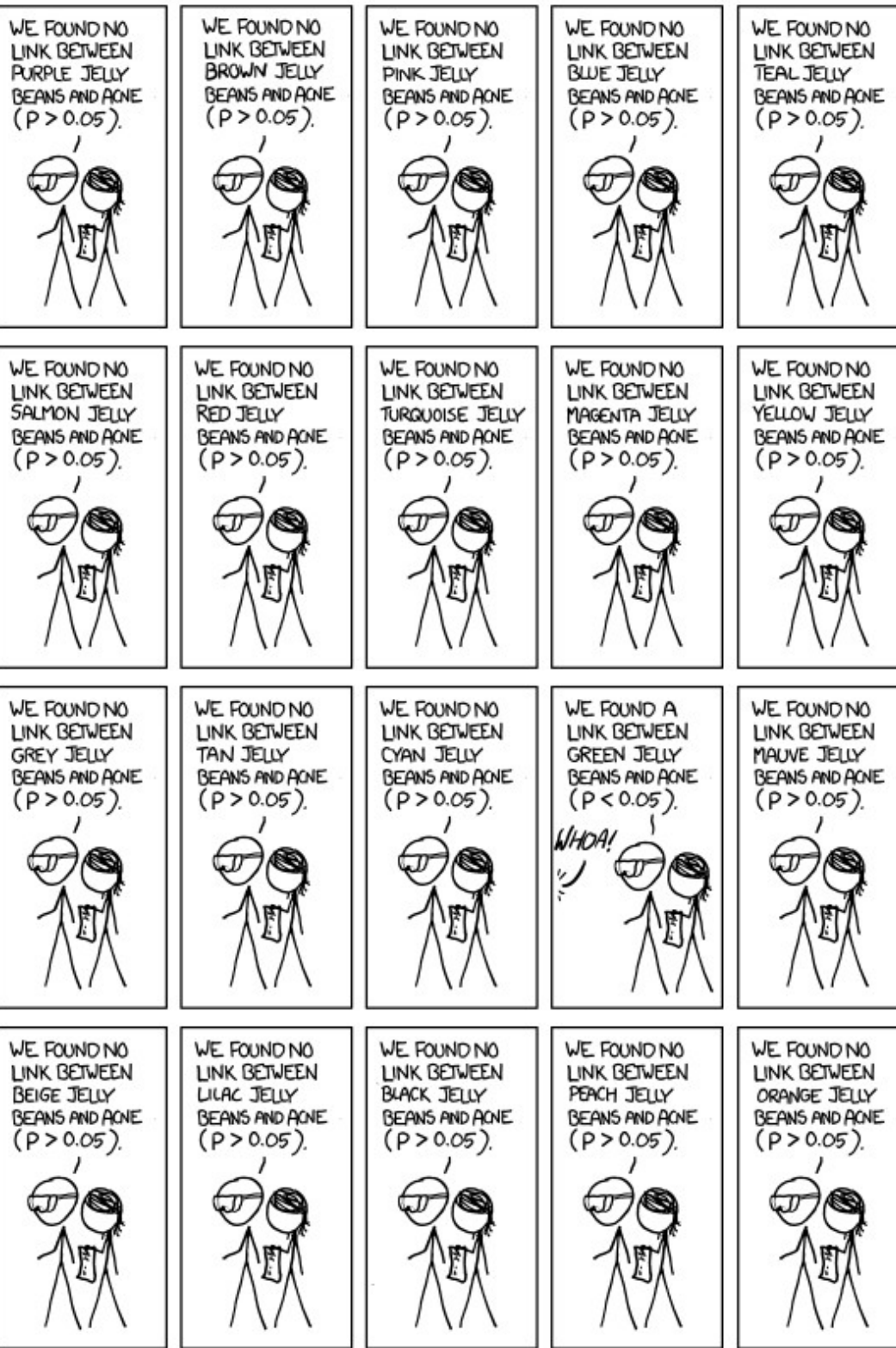
I HEAR IT'S ONLY  
A CERTAIN COLOR  
THAT CAUSES IT.

SCIENTISTS!

BUT  
MINECRAFT!



# “Significant” – Part 2





# The Meaning of Statistical Significance

- Answer: because we might have been “lucky”: the result might have been a fluke
- It might not be statistically-significant
- This is an example of how statisticians think differently to normal people
- So how do we decide whether a result is statistically-significant?

# Determining Statistical Significance

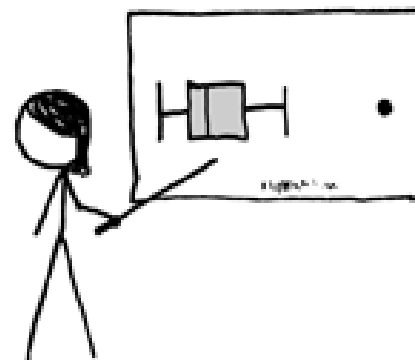
CAN MY BOYFRIEND  
COME ALONG?



I'M NOT YOUR  
BOYFRIEND!  
/ YOU TOTALLY ARE.  
I'M CASUALLY  
DATING A NUMBER  
OF PEOPLE.

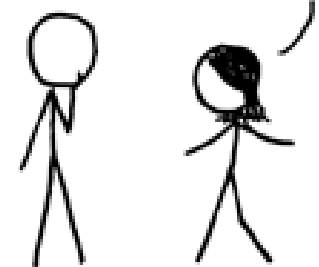


BUT YOU SPEND TWICE AS MUCH  
TIME WITH ME AS WITH ANYONE  
ELSE. I'M A CLEAR OUTLIER.



YOUR MATH IS  
IRREFUTABLE.

FACE IT—I'M  
YOUR STATISTICALLY  
SIGNIFICANT OTHER.



# The Meaning of Statistical Significance

- Answer: when the result deviates by more than some distance from what we expect
- Determining this distance is what this topic is all about!



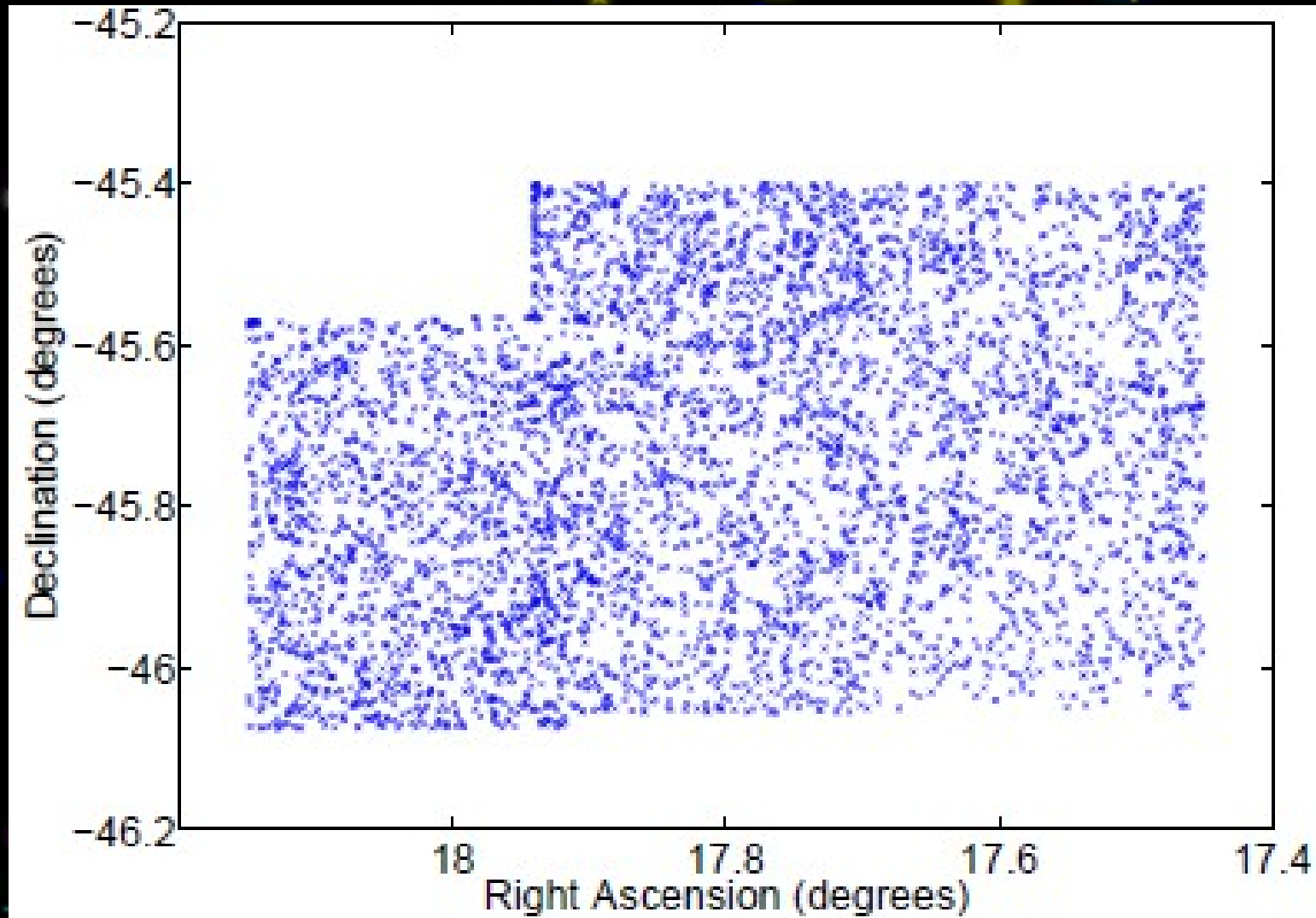
# Confidence Level vs Level of Significance

- What's the difference between these two terms?

# Confidence Level vs Level of Significance

- Answer: A confidence level is how confident we can be that the true value lies within a certain range
- But a level of significance is the significance a result must at least have to reject the null hypothesis

# Problem: Cluster-Finding



- Want to teach a computer to find the clusters automatically
- Any ideas how?

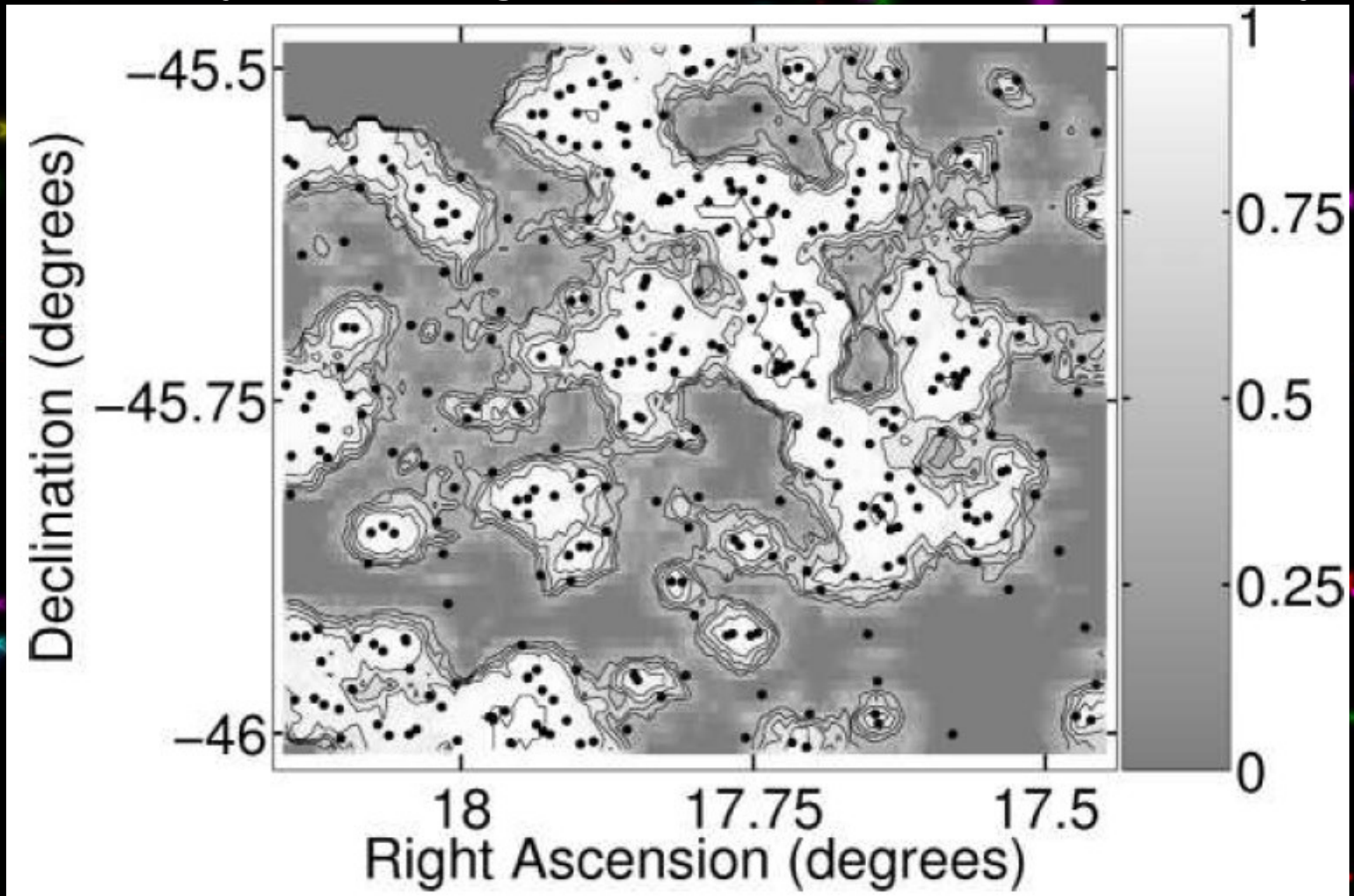


# Testing a Hypothesis at Each Location on the Sky

- Null hypothesis: there is no cluster here
- Alternative hypothesis: there is one
- At each location, we ask: what is the probability of obtaining the observed density of galaxies or more, given the mean density?

# Cluster Finding with Statistics

- Probability mapping: a tool to tell us where they are



# Quick Example

- Suppose the mean density of galaxies is 0.2, and at a particular location we find a galaxy density of 0.4
- This density at this location is determined using 6 galaxies
- The standard deviation of the galaxy density is 0.2
- Let's assume that the galaxy density from location to location is normally-distributed
- At 10% level, have we found a cluster here?



# Quick Solution

- Population standard deviation known so we use Z- distribution
- This is a one-tail test, so rejection region is in the upper tail and has an area of 0.1
- Converting the observed density to Z:
  - $Z = (0.4 - 0.2) / (0.2 / 6^{0.5})$
  - $Z = 2.4495$

# Quick Solution

- Population standard deviation known so we use Z- distribution
- This is a one-tail test, so rejection region is in the upper tail and has an area of 0.1
- Converting the observed density to Z:
  - $Z = (0.4 - 0.2) / (0.2 / 6^{0.5})$
  - $Z = 2.4495$

# Quick Solution

- For cumulative probability of 0.9,  $Z = 1.28$  – this is our critical value and the lower limit of the rejection region
- So our  $Z$  of 2.45 is in the rejection region and we reject the null hypothesis (that there is no cluster)
- So we have discovered a cluster!



# How We Actually Did This

- It's more complicated than that, as usual
  - [Link to my thesis](#)

# "Who Can I Be Now?"

- Statisticians don't accept results at face value
- A result might not be “real” or “statistically-significant”
- There are often competing hypotheses, each with people who believe them to be real, because people have various points of view
- But with enough data, a statistician has the means – the “sight” to determine which one is actually real
- And so the statistician helps the truth to be real to everyone else

# "Who Can I Be Now?"

*Everybody feels that everything is real*

***Anybody's point of view***

*But even in my life, I knew you found your sight*

*And nothing would be quite the same*

*Please help me*

*Who can I be now? You found me*

*Who can I be? Fell apart, you found me*

*Now can I be now? You found me*

***Now can I be real?***

- David Bowie, "Who Can I Be Now?" (Young Americans, 1975)



Questions, suggestions for next time:  
[anthony.smith@scu.edu.au](mailto:anthony.smith@scu.edu.au)

These slides available from the usual place

Any Questions About Any Maths At  
All?

# Credits

- All comics from xkcd, the webcomic of romance, sarcasm, math and language
  - <https://xkcd.com/>
- Any songs were by David Bowie