Stat310

Hypothesis tests

Hadley Wickham

1. Assessment

2. Revision

3. Testing process

Assessment

Final

~8 smaller questions

3-4 on inference (maximum likelihood estimation, confidence intervals, hypothesis testing)

The rest on everything else

Will soon have page on website with details and last years final

CHECK YOUR OWLSPACE GRADES!

Revision

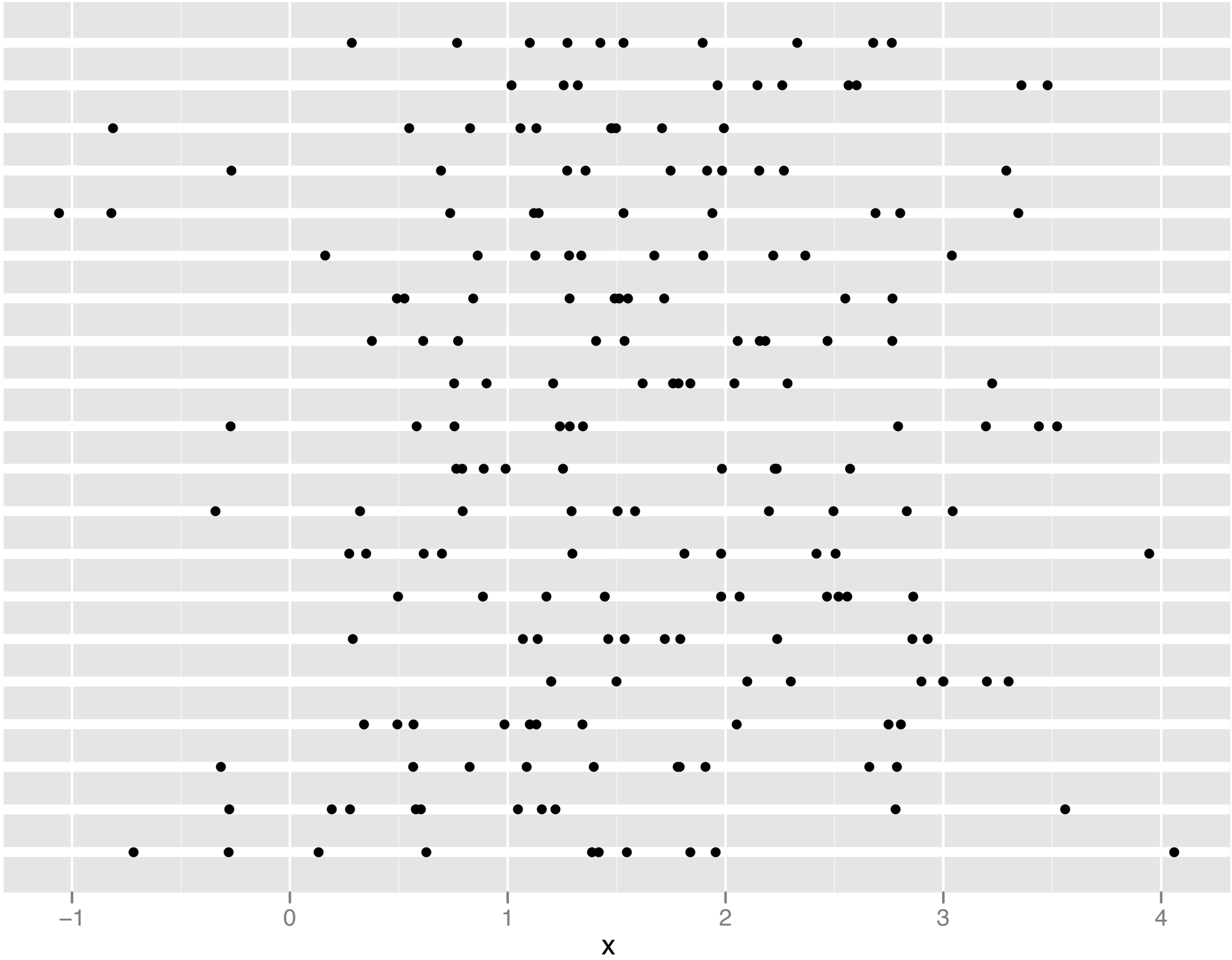
A **suspect** is accused of a **crime**. The suspect is declared guilty or not guilty based on a **trial**. Each trial has a **defence** and a **prosecution**. On the basis of how **evidence** compares to a **standard**, the judge makes a decision to **convict** or **acquit**.

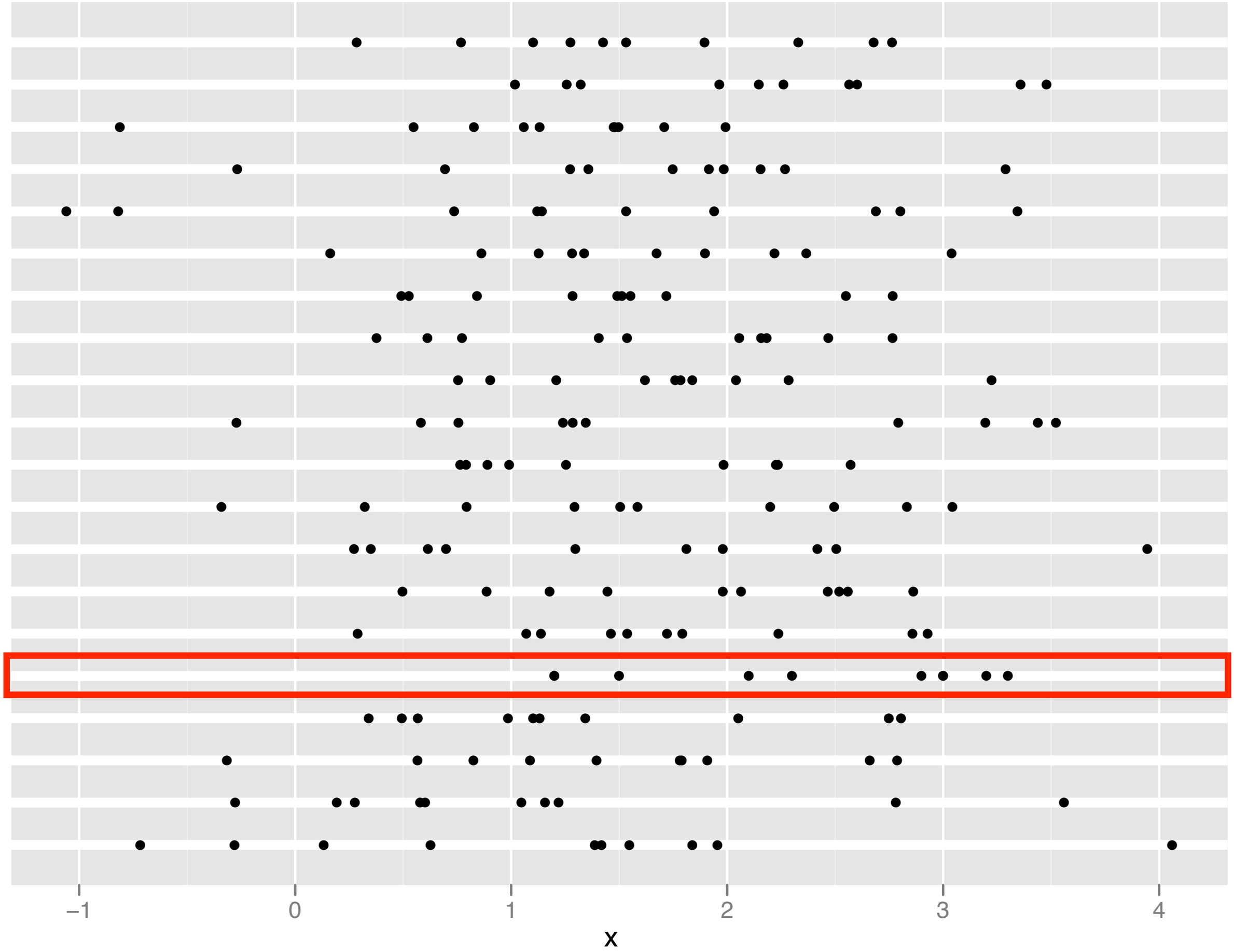
Your turn

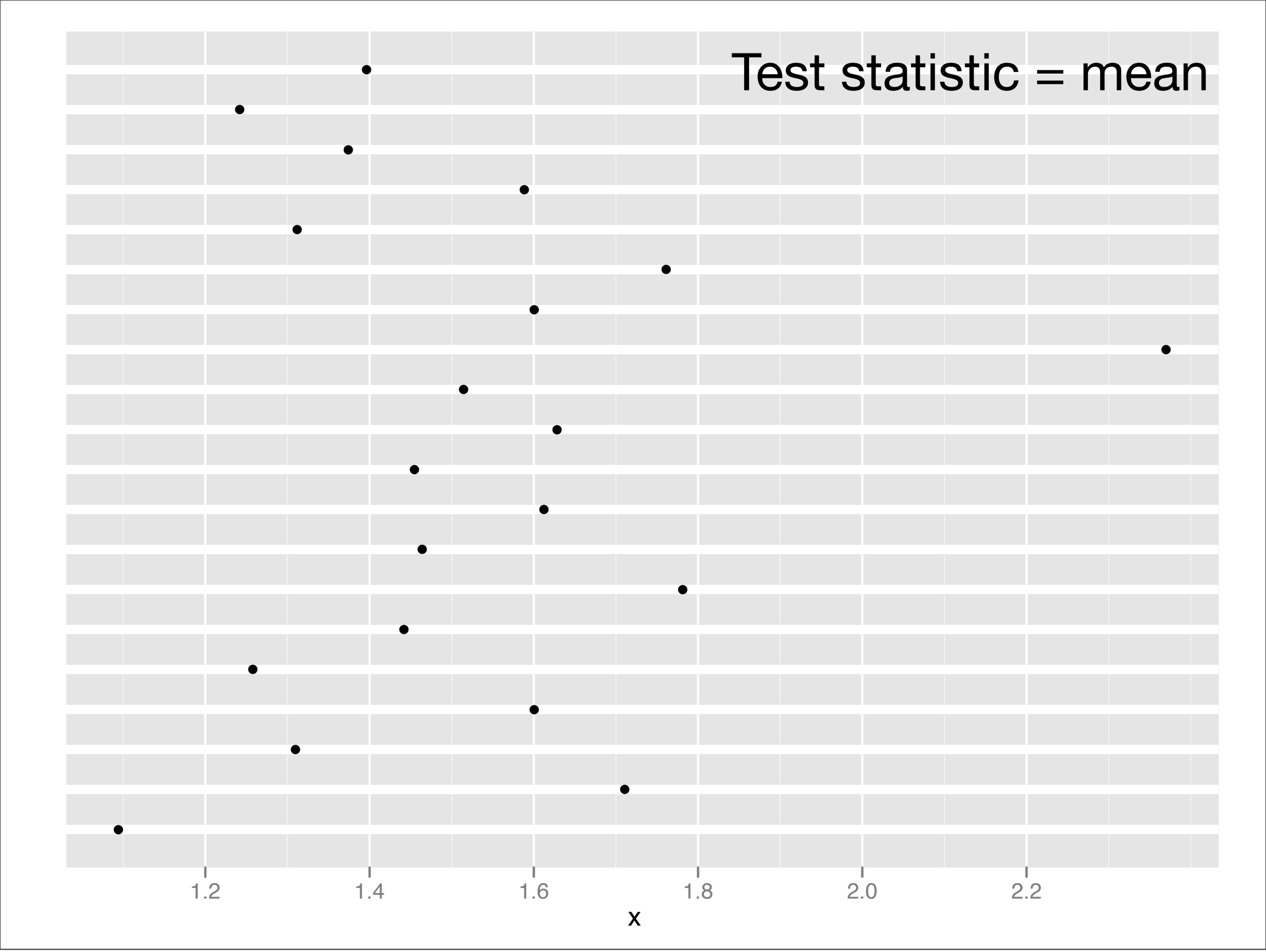
With a partner, write down the criminal equivalents to the following terms:

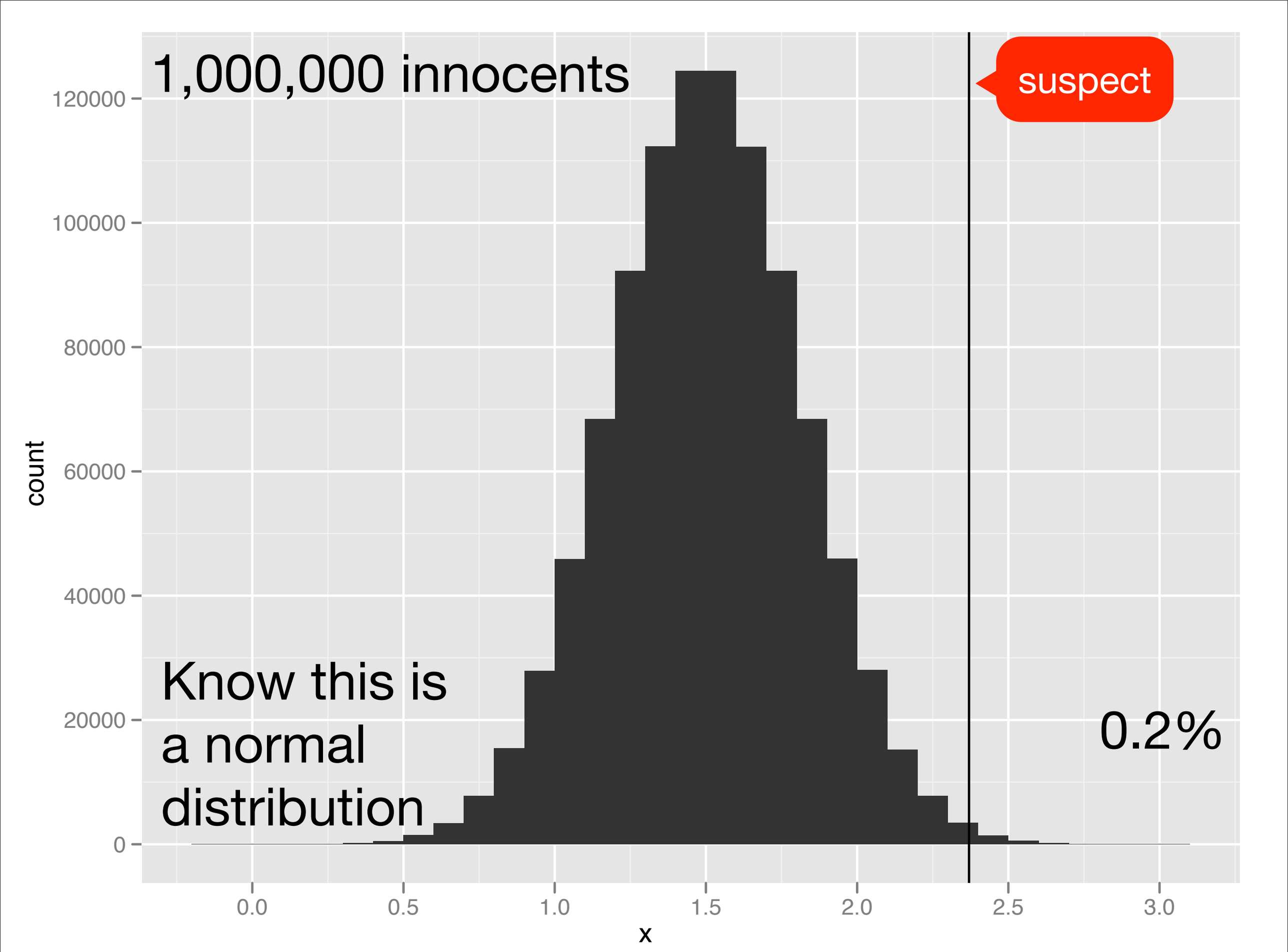
null hypothesis, alternative hypothesis, test statistic, null distribution, reject the null hypothesis, fail to reject the null hypothesis, accept the null hypothesis

Suspect	2.9 2.1 3.0 3.2 1.2 3.0 3.3 1.2 2.3 1.5
Defence	From Normal(1.5, 1)
Prosecution	Not from that distribution (mean is different)
Innocents	0.9 0.9 2.4 1.4 2.1 1.4 1.5 1.1 2.6 1.0 1.7 1.3 0.8 0.9 2.1 2.1 2.6 2.8 1.1 0.5 2.1 3.0 2.1 1.8 3.5 1.4 0.0 1.8 2.6 3.7


−1
0
1
2
3
4
x


−1
0
1
2
3
4
x


Test statistic = mean
1.2
1.4
1.6
1.8
2.0
2.2
x


1,000,000 innocents
suspect
Know this is
a normal
distribution
0.2%
count
120000
100000
80000
60000
40000
20000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x

Definitions

test statistic: numerical summary used to collapse sample into a single number

null distribution: distribution of test-statistic of innocents

p-value: probability that a true innocent would look as guilty as the suspect

vs Stat 280/AP

There you learn a set of specific tests and when they apply

Here you learn how tests work in general, so you can understand all tests work and derive new ones for new situations.

Process

1. Write down **null** and **alternative hypotheses** (positions of defence and prosecution)

2. Figure out good test statistic (what numeric summary captures useful evidence)

3. Work out null distribution (distribution of innocents)

4. Calculate p-value by comparing actual value to null distribution (what proportion of true innocents look more guilty than the suspect)

5. Reject H_o if p-value smaller than cutoff

Hypothesis

Null hypothesis = H_o

Alternative hypothesis = H_a / H_1

Hint: because we need to be able to calculate the null distribution, the null hypothesis will always be of the form:

Some parameter = some value
$\mu = 0$

Suitcase

Does this suitcase contain a radioactive bomb? Construct H_o and H_a.

Let R be the background radiation measured over a minute. R ~ Poisson(2). Let S be the radiation from the suitcase. Construct a more precise H_o and H_a.

Grade difference

I'm interested in whether or not there is a difference between this years average stat310 grade and last years. Construct H_o and H_a. (What should I do to make this experiment meaningful?)

If grades are normally distributed both years, can you rewrite the null hypothesis to be more precise? What other assumptions do you need?

Alternative

Can be **one-sided** or **two-sided**

What does more guilty mean?

For the **suitcase**: guilty means higher radiation = one-sided

For **course grades**: guilty means positive or negative difference = two-sided.

1. Write down H_o and H_a
   (positions of defence and prosecution)

2. **Figure out good test statistic**
   (what numeric summary?)

3. **Work out null distribution**
   (distribution of innocents)

4. Calculate p-value by comparing actual value to null distribution (what proportion of true innocents look more guilty than the suspect)

5. Reject H_o if p-value smaller than cutoff

X_i iid, and n large:

$$\frac{\bar{X}_n - E(\bar{X})}{sd(\bar{X})} \dot{\sim} Z$$

and even more approximately

$$\frac{\bar{X}_n - E(\bar{X})}{\hat{sd}(\bar{X})} \dot{\sim} Z$$

$$X_i \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$$

$$\frac{\bar{X}_n - E(\bar{X})}{sd(\bar{X})} \sim Z$$

but

$$\frac{\bar{X}_n - E(\bar{X})}{\hat{sd}(\bar{X})} \sim t_{n-1}$$

$$X_i \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Others

Difference of normals

Sum of poisson

Sum of binomial

Sum of exponential

...

Suitcase

What test statistic might you use?
(What experiment might you conduct?)

What is the null distribution?

Grade difference

What test statistic might you use?

What is its null distribution?

1. Write down H_o and H_a
   (positions of defence and prosecution)

2. Figure out good test statistic
   (what numeric summary?)

3. Work out null distribution
   (distribution of innocents)

4. **Calculate p-value** by comparing actual value to null distribution (what proportion of true innocents look more guilty than the suspect)

5. Reject H_o if p-value smaller than cutoff

P-value

Standardised measurement of evidence.

Low p-value = low probability of innocent looking this guilty = **reject the null**

High p-value = high probability of innocent looking this guilty = **don't reject**

Don't need to know anything else about the test!

Reject the null = **guilty**

Accept the null = innocent

Fail to reject the null = **not guilty**

Suitcase

Background radiation is Poisson(2)

If I measure the suitcase and record a 3, what's the p-value? What if I record a 5?

What's the probability it's a bomb?

$X \sim \text{Poisson}(2)$

X	P(X ≤ x)
0	0.14
1	0.41
2	0.68
3	0.86
4	0.95
5	0.98

Course grades

Assume for simplicity there were 100 students both years, and the variance of the course grade was 80.

What would the distribution of the test statistic be?

1. Write down H_o and H_a
   (positions of defence and prosecution)

2. Figure out good test statistic
   (what numeric summary?)

3. Work out null distribution
   (distribution of innocents)

4. Calculate p-value by comparing actual value to null distribution (what proportion of true innocents look more guilty than the suspect)

5. **Reject H_o if p-value smaller than cutoff**

	Say is guilty	Say is not guilty
Is guilty	Correct	**False acquittal**
Is innocent	**False conviction**	Correct

Your turn

Which type of error is more expensive/ more costly/worse in the criminal justice system?

	Reject H_O	Fail to reject H_O
H_O false	Correct	**Type II error**
H_O true	**Type I error**	Correct

Rates

For a given test,

P(false conviction) = α = significance level

P(true conviction) = β = power

What do think happens to β if you try to make α smaller?

$$\alpha\uparrow\ \beta\downarrow$$

$$\alpha\downarrow\ \beta\uparrow$$

Cut off

Choose cut-off based on rate of false convictions.

If you want a 5% rate of false convictions, reject H_o if the p-value is less than 0.05.

0.05 is standard.

Can work out power.

Connection to confidence intervals

If you construct a 90% confidence interval, and it doesn't include the parameter under the null, then the p-value must be $> 1 - 0.9 = 0.1$.

If the p-value is 0.08, then a 92% or greater confidence interval would include the null parameter, and a smaller confidence interval would not.

Next week

Tuesday: Develop specific tests

Thursday: Revision. Why you should major in statistics. Cupcakes. Feedback.

Ask and vote on review questions!

CHECK YOUR OWLSPACE GRADES!