Introduction to Tests of Significance

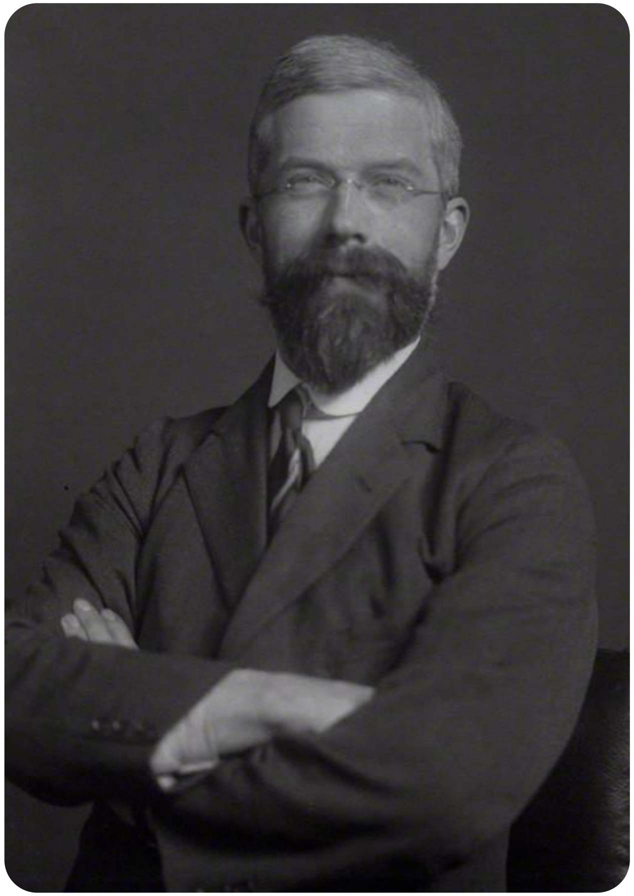
Gaston Sanchez

Tests of significance or Hypothesis Tests

(FPP chapter 26)

The Lady Tasting Tea

Background info



1920's

Cambridge, England

Afternoon tea party at the University

Ronald Aylmer Fisher

(one of the founding fathers of modern statistics)

Lady’s claim:

“I can tell whether the milk is poured first and the tea is added next, or whether the tea is poured first and the milk is added to the tea”

How to test the claim?

Try one cup (random guess)

50% chance of guessing correctly

Try two cups (random guess)

Assuming binomial process ...

Guess none = 25%

Guess one = 50%

Guess two = 25%

Try three cups (random guess)

Assuming binomial process ...

Guess none = 12.5%

Guess one = 37.5%

Guess two = 37.5%

Guess three = 12.5%

Try *n* cups (random guess)

Assuming binomial process ...

$$P(k_{\text{successes}}) = \binom{n}{k} 0.5^k (0.5)^{n-k}$$

What Fisher did …

Set up a tasting tea experiment

- Fisher proposed 8 cups
- 4 with tea first
- 4 with milk first
- Present them in random order
- Have the lady taste them and guess

Tea first

Put 4 cups here

Milk first

Put 4 cups here

Tea first

Milk first

Reality

Guess

	Milk	Tea
Milk		
Tea		

Tea first

Milk first

One possible result

Reality

Guess		Milk	Tea
	Milk	3	1
	Tea	1	3

So she was mostly correct

Reality

Guess

	Milk	Tea
Milk	0	4
Tea	4	0

All incorrect

	Milk	Tea
Milk	1	3
Tea	3	1

Mostly incorrect

Evenly?

	Milk	Tea
Milk	2	2
Tea	2	2

Mostly correct

	Milk	Tea
Milk	3	1
Tea	1	3

All correct

Lady’s result

	Milk	Tea
Milk	4	0
Tea	0	4

Fisher's approach

	Milk	Tea
Milk	a	b
Tea	c	d

$$P(\text{arrangement}) = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{N!\,a!\,b!\,c!\,d!}$$

Fisher used a hypergeometric distribution to figure out probabilities

Probabilities

0	4
4	0

p=.014

1	3
3	1

p=.229

2	2
2	2

p=.514

3	1
1	3

p=.229

4	0
0	4

p=.014

P(guessing all correct) = 0.014

It looked like the lady's claim was believable after all

Hypothesis Testing

Try one cup

50% chance of guessing correctly

Let’s say you taste a cup of tea everyday, for 100 consecutive days

Randomly guessing

In 100 trials, you should expect to guess correctly 50% of the time

47 right - 53 wrong

52 right - 48 wrong

54 right - 46 wrong

57 right - 43 wrong

60 right - 40 wrong

These results can be explained by chance

Let’s say you have some tasting tea ability

Guessing super power

In 100 trials, what should you expect to guess correctly?

65 right - 35 wrong?

75 right - 25 wrong?

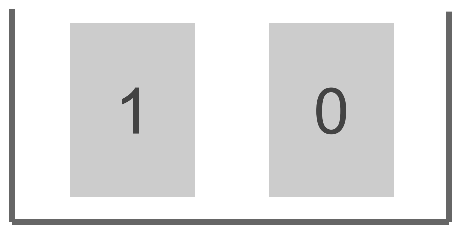
88 right - 12 wrong?

90 right - 10 wrong?

100 right - 0 wrong?

How (un)likely are these results just by randomly guessing?

Box model: EV



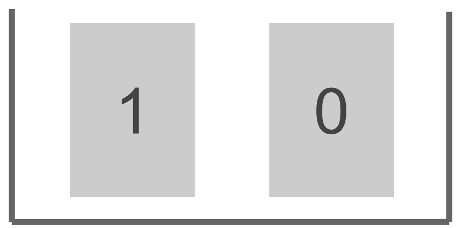
100 draws

Avg(box) = (1 + 0) / 2 = 0.5

EV(sum) = 100 (0.5) = 50

EV(percentage) = (50 / 100) x 100% = 50%

Box model: SE



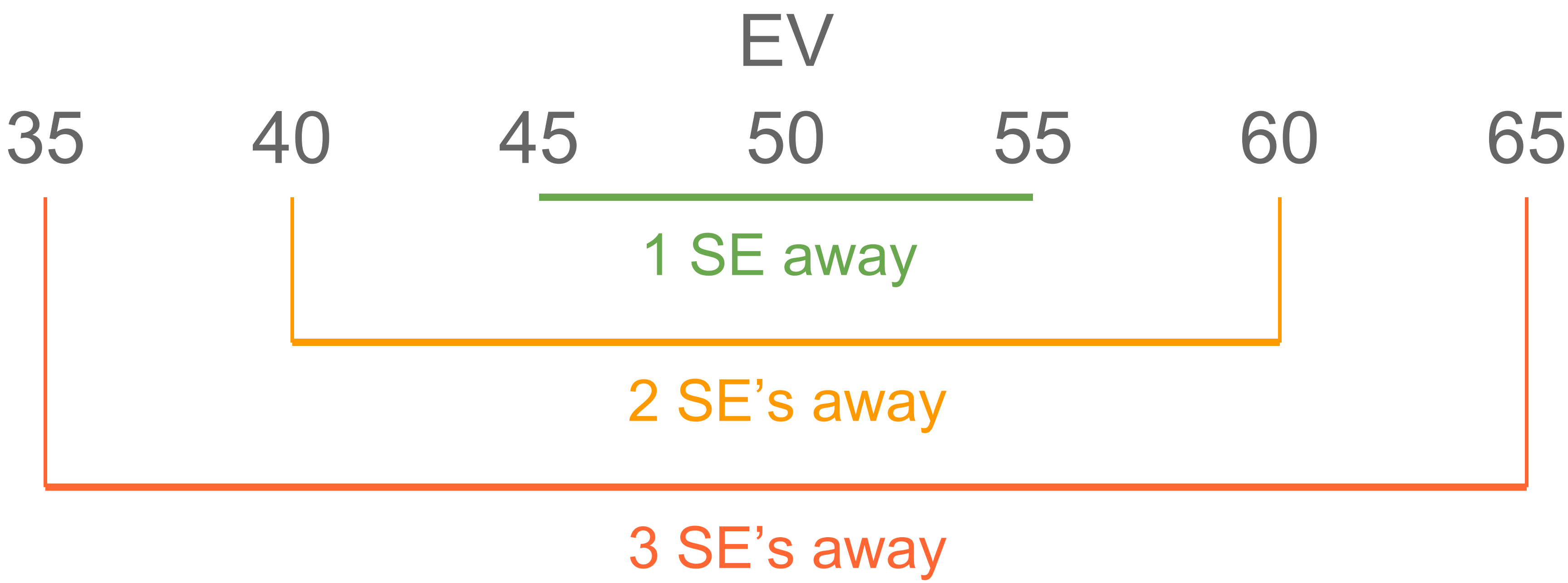
100 draws

$$SD(box) = (1 - 0)\sqrt{0.50 \times 0.50} = 0.5$$

$$SE(sum) = \sqrt{100}\ (0.5) = 5$$

$$SE(percentage) = (5 / 100) \times 100\% = 5\%$$

How far from Expected Value?



How far should the observed results be from the EV to say that results are ***unlikely****?*

Tests of Significance

Can the result be explained by chance or is another explanation necessary?

2 competing ideas

2 competing ideas (hypotheses)

Just chance

Null hypothesis

Other explanation

Alternative hypothesis

Main steps for a hypothesis test

1. State **Null** and **Alternative** hypotheses

2. Determine and calculate **Test Statistic**

3. Compute **P-value**

4. Make a **Conclusion**

Null Hypothesis

Results obtained due to random variation

Results can be explained by chance

Chance alone is a reasonable explanation

Alternative Hypothesis

Alternative to the null

Another explanation is necessary

Results can't be explained by chance

Chance alone is not a reasonable explanation

Test Statistic

Measures the difference between the data and what is expected on the null hypothesis

Quantifies the differences between observed data end expected results (under null)

P-value

P-value is the chance of getting data like we got, or more extreme, given the null hypothesis is true.

The smaller the **p-value**, the stronger the evidence against the null hypothesis.

Conclusion

Either:

Fail to reject null hypothesis (accept)

OR

Reject null hypothesis