Chapter 6:
Hypothesis testing

1

Test if new drug has a clinical effect

2

Hypothesis testing

- **Hypothesis testing** compares data to what we would expect to see if a specific null hypothesis were true. If the data are too unusual, compared to what we expect to see if the null hypothesis were true, then the null hypothesis is rejected

3

Hypotheses

- The **null hypothesis** is a **specific statement** about a population parameter made for the purposes of argument. A good null hypothesis is a statement that would be interesting to reject

- "default" hypothesis that has an interest of zero
  - e.g., no effect, no preference, no correlation, no difference

- Abbreviated as H_0

4

Hypotheses

- The **alternative hypothesis** includes all other feasible values for the population parameter besides the value stated in the null hypothesis
- Includes possibilities that are biologically interesting
  - e.g., there IS an effect, preference, correlation, difference
- Abbreviated as H_A

5

Reject H_0 or not

- The H_0 is what is being tested
- If data are consistent with H_0 then you “fail to reject” it
- If data are inconsistent with H_0 then you reject it
  - “rule out” the null hypothesis
- You do not “prove” the H_A, you can only “reject” or “fail to reject” the H_0

6

Ex 6.2: The right hand of toad

- Tested if individual toads prefer to use one forelimb over another (display "handedness")
- Data showed that individuals show preference for a particular forelimb
- Do right- and left-handed toads differ in frequency?

7

Null hypothesis

- Question: Do right- and left-handed toads differ in frequency?
- H_0: Right and left-handed toads are equally frequent
- …or H_0: $p_R = p_L$
- …or H_0: $p_R = 0.5$
- …or H_0: $p_L = 0.5$

8

Null and alternative hypotheses

- H_0: $p_R = 0.5$
- There are two alternative possibilities: $p_R < 0.5$ and $p_R > 0.5$
- In a **two-sided (or two-tailed) test,** the alternative hypothesis includes parameter values on both sides of the parameter value specified by the null hypothesis
- H_A: $p_R \neq 0.5$

9

Test statistic

- The **test statistic** is a number calculated from the data that is used to evaluate how compatible the data are with the result expected under the null hypothesis
- In study, 18 toads were sampled and 14 were observed to be right-handed
- In this case the test statistic is 14 (or p_R = 14/18 = 0.7778)

10

Null distribution

- Test statistic = 14
- Null expectation = 9 (with H_0 $p_R = 0.5$ and 18 observations)
- Samples have error… so it's possible that the null hypothesis is true and by chance you observed 14 right-handed toads
- The **null distribution** is the sampling distribution of outcomes for a test statistic under the **assumption that the null hypothesis is true**

11

Null distribution

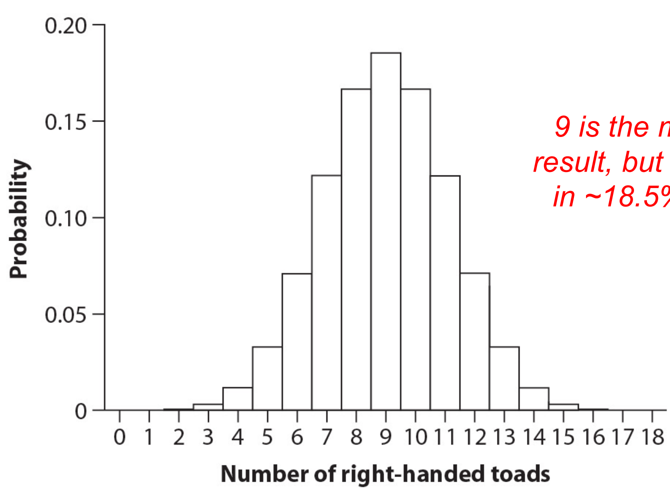

Fig 6-2.1

9 is the most common result, but only happened in ~18.5% of samples!

TABLE 6.2-1 All possible outcomes for the number of right-handed toads when 18 toads are sampled, and their probabilities under the null hypothesis.

Number of right-handed toads	Probability
0	0.000004
1	0.00007
2	0.0006
3	0.0031
4	0.0117
5	0.0327
6	0.0708
7	0.1214
8	0.1669
9	0.1855
10	0.1669
11	0.1214
12	0.0708
13	0.0327
14	0.0117
15	0.0031
16	0.0006
17	0.00007
18	0.000004
Total	1.0

12

P-value

- Is 14 different enough from 9, given the null distribution, to reject the null hypothesis?
- The ***P*-value** is the **probability of obtaining the data** (or data showing as great or greater difference from the null hypothesis) **given that the null hypothesis were true**

13

P-value

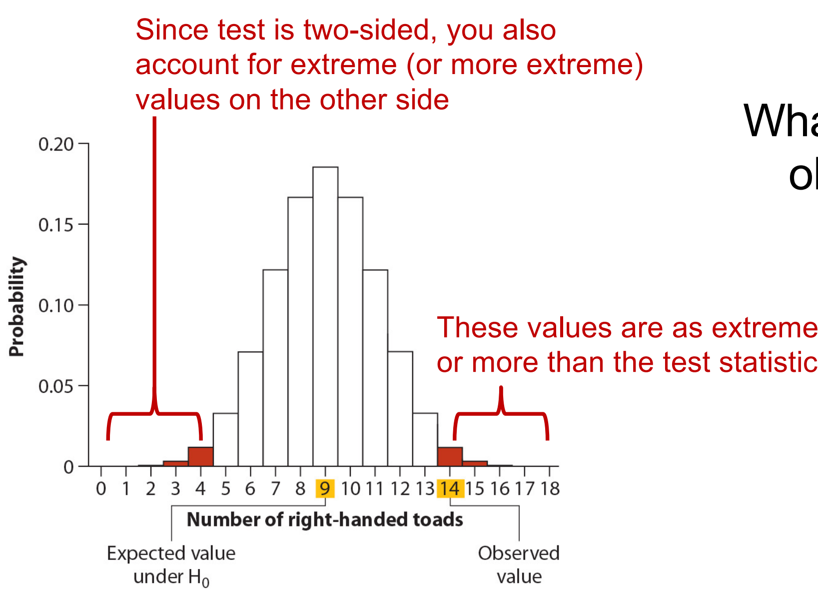

Fig 6-2.2

What is the probability of obtaining a value as extreme as 14?

14

P-value

TABLE 6.2-1 All possible outcomes for the number of right-handed toads when 18 toads are sampled, and their probabilities under the null hypothesis.

Number of right-handed toads	Probability
0	0.000004
1	0.00007
2	0.0006
3	0.0031
4	0.0117
5	0.0327
6	0.0708
7	0.1214
8	0.1669
9	0.1855
10	0.1669
11	0.1214
12	0.0708
13	0.0327
14	0.0117
15	0.0031
16	0.0006
17	0.00007
18	0.000004
Total	1.0

- Observations of 0, 1, 2, 3, etc. right-handed toads are mutually exclusive
- So **sum probabilities** of getting values as extreme as 14
  - Pr[*14 or more right-handed*] = Pr[14] + Pr[15] + Pr[16] + Pr[17] + Pr[18]
- Pr[14 or more right-handed] = 0.0155

15

P-value

TABLE 6.2-1 All possible outcomes for the number of right-handed toads when 18 toads are sampled, and their probabilities under the null hypothesis.

Number of right-handed toads	Probability
0	0.000004
1	0.00007
2	0.0006
3	0.0031
4	0.0117
5	0.0327
6	0.0708
7	0.1214
8	0.1669
9	0.1855
10	0.1669
11	0.1214
12	0.0708
13	0.0327
14	0.0117
15	0.0031
16	0.0006
17	0.00007
18	0.000004
Total	1.0

- But test is two-sided, so you need to also account for getting values as extreme in the other direction
  - i.e., values ≤ 4
- Quickest way to calculate is to multiply Pr[*14 or more right-handed*] by 2
- P = 0.0155 x 2 = 0.031

16

P-value

- $P = 0.031$
  - There is a 3.1% chance of the observed data given that the null hypothesis is true

- Is this a small enough probability to reject the null?

- In many areas of biology, a P-value < 0.05 is small enough to reject the null hypothesis

- The **significance level, α,** is a probability used as a criterion for rejecting the null hypothesis.
  - If the P-value is less than or equal to α, then the null hypothesis is *rejected*. If the P-value is greater than α, then the null hypothesis is *not rejected*.

17

18

Errors in hypothesis testing

TABLE 6.3-1 Types of error in hypothesis testing.

	Reality	
Conclusion	H_0 true	H_0 false
Reject H_0	Type I error	Correct
Do not reject H_0	Correct	Type II error

power

- **Type I error** is rejecting a true null hypothesis. The significance level α sets the probability of committing a Type I error.
- **Type II error** is failing to reject a false null hypothesis.
- The **power** of a test is the probability that a random sample will lead to rejection of a false null hypothesis

19

20

Ex 6.4: The genetics of mirror-image flowers

- Most flowering plants are hermaphrodites, with flowers that include female (style) and male (anther) reproductive organs
- To avoid self-fertilization, some species have mixture of right and left "handed" flowers
- Study of genetics of this trait
  - Produce pure strains right/left individuals
  - Crossing pure strain of left X pure strain of right produced all right-handed offspring
  - Crossed these offspring and expected 1:3 of left:right F2 under simple model of inheritance

21

Ex 6.4: The genetics of mirror-image flowers

- Study of genetics of this trait
  - Produce pure strains right/left individuals
  - Crossing pure strain of left X pure strain of right produced all right-handed offspring (F1)
  - Crossed these offspring and expected 1:3 of left:right F2 under simple model of inheritance

1st cross

	R	R
L	RL	RL
L	RL	RL

2nd cross

	R	L
R	RR	RL
L	RL	LL

22

Hypotheses and test statistic

- H_0: left and right-handed offspring occur at 1:3 ratio
  - H_0: $p_L = 0.25$
- H_A: left and right-handed offspring do not occur at 3:1
  - H_A: $p_L \neq 0.25$
- Study resulted in sample of 27 offspring with 6 that were left handed
  - Test statistic: $p_L = 6/27 = 0.222$

23

Null distribution

- Test statistic: $p_L = 6/27 = 0.222$
- Expectation with sample of 27:
  - 27 x 0.25 = 6.75

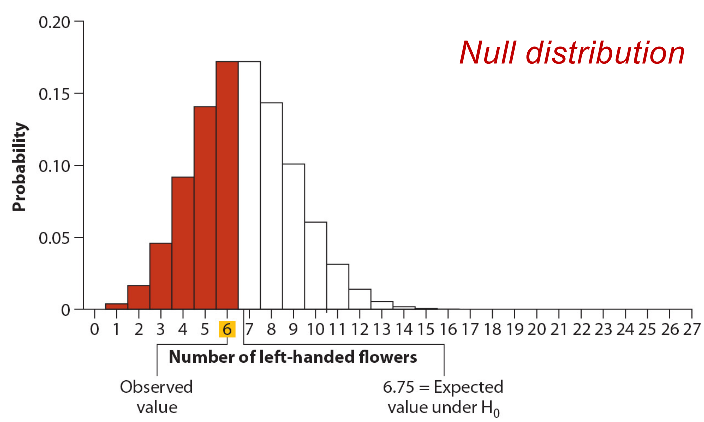

Fig 6.4-1

24

P-value

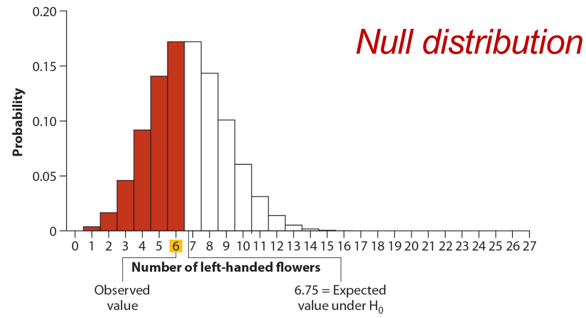

Fig 6.4-1

- Pr[*num left* ≤ *6*] = Pr[6]+Pr[5]+Pr[4]…Pr[0]
- Pr[*num left* ≤ *6*] = 0.471
- But test is two-sided
- *P* = 0.471 x 2 = 0.942

25

Interpreting a non-significant result

- *P* = 0.471 x 2 = 0.942
- There is a 94.2% probability of the observed data given that the null hypothesis is true
- Data are *compatible* or *consistent* with the null hypothesis
- “Fail to reject” the null hypothesis

26

27

One-sided tests

- In a **one-sided (or one-tailed) test,** the alternative hypothesis includes parameter values on only one side of the value specified by the null hypothesis

- H_0 is rejected only if the data depart from it in the direction stated by H_A

- In practice one-sided tests are used sparingly

- For this course, **all hypothesis testing will be done using a two-tailed test**

28

Hypothesis testing vs confidence intervals

- Recall that 95% confidence interval puts bounds on the most plausible population parameter based on your random sample
- Almost always, the 95% confidence interval and a hypothesis test give the same answer
- Example: null that $p_R = 0.5$
  - If p-value < 0.05 then you reject the null hypothesis
  - If 95% confidence interval does not include 0.5 then the data are inconsistent with the population proportion being 0.5

29

Hypothesis testing vs confidence intervals

- Magnitude:
  - Confidence interval has added benefit of giving actual magnitude (e.g., how far away is 0.5 from 95% CI)
  - P-value gives qualitative magnitude (smaller p-value means greater ability to reject the null)
- Generally, a hypothesis test is used more often, but both are good approaches

30