HOLIDAY HOMEWORK 假期作业

AP Statistics

Exercise sheets, topic by topic.

每个单元：练习卷。

For class or self-study • no answer keys included.

Contents 目录

1 Describing one-variable data – Part 1

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	中文	Write it 3 times (English)		
mean	平均数			
median	中位数			
standard deviation	标准差			

Recall

You have worked with data before:

- You have found the average (mean) of a set of numbers.
- You have drawn bar charts and other graphs.
- Some data sets are more spread out than others.

This sheet lines up how to describe one set of data. Each idea has a worked example.

New ideas

1 The center

Two measures of the ”middle” of data:

- the **mean** 平均数–the average (add up, divide by how many),
- the **median** 中位数–the middle value when the data are in order.

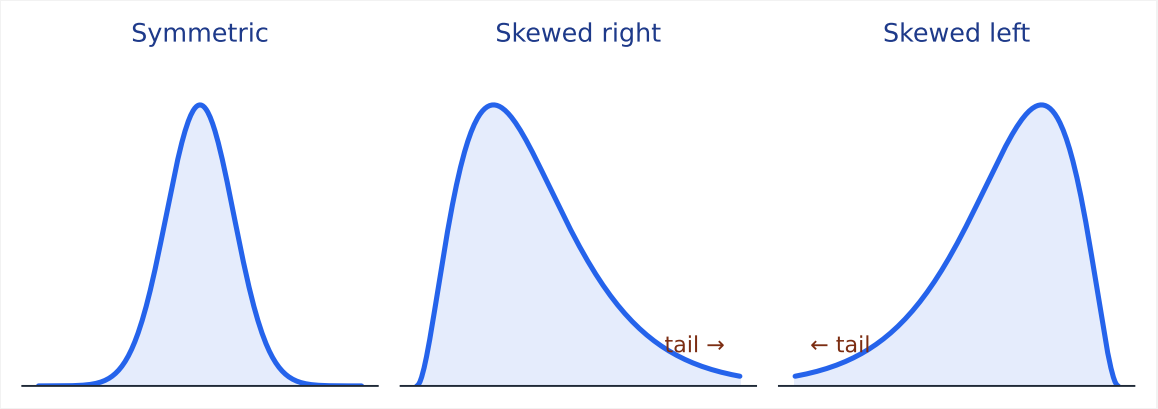
Worked example. For $2, 4, 4, 6, 9$: mean $= \dfrac{2+4+4+6+9}{5} = 5$; median $= 4$ (the middle one).

2 The spread

How spread out the data are: the **range** (largest minus smallest) and the **standard deviation** 标准差 (a typical distance from the mean).

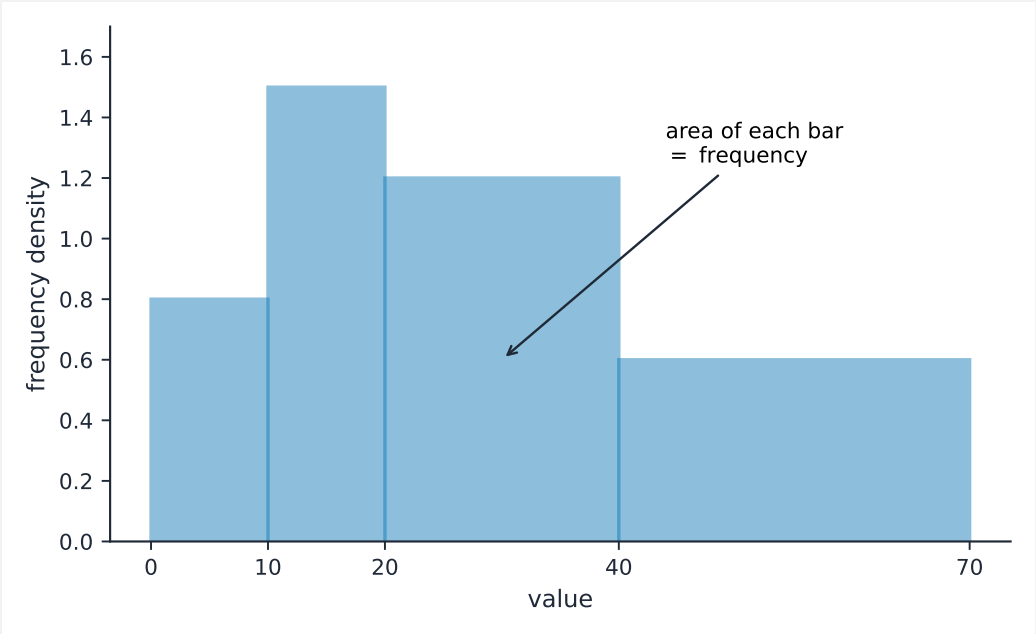
3 The shape

Data can be **symmetric** (even on both sides) or **skewed** (a long tail one way).



4 Pictures of data

A **histogram** shows the shape; a **box plot** shows the median and spread at a glance.



Watch out: the **mean** is pulled toward outliers (extreme values), but the **median** is not. For very skewed data (like incomes), the median often describes the "typical" value better.

Practice

Try these in order. The methods are all above.

2.1 State the difference between the mean and the median. [2]

2.2 Find the mean of $3, 5, 7, 9, 11$. [2]

2.3 Find the median of $3, 5, 7, 9, 11$. [1]

2.4 Find the range of $3, 5, 7, 9, 11$. [1]

2.5 A data set has one very large outlier. State whether the mean or the median is more affected by it, and why. [2]

1 The normal distribution – Part 2

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	中文	Write it 3 times (English)
normal distribution	正态分布	

Recall

In Part A you described data's center, spread, and shape. Now one very common shape:

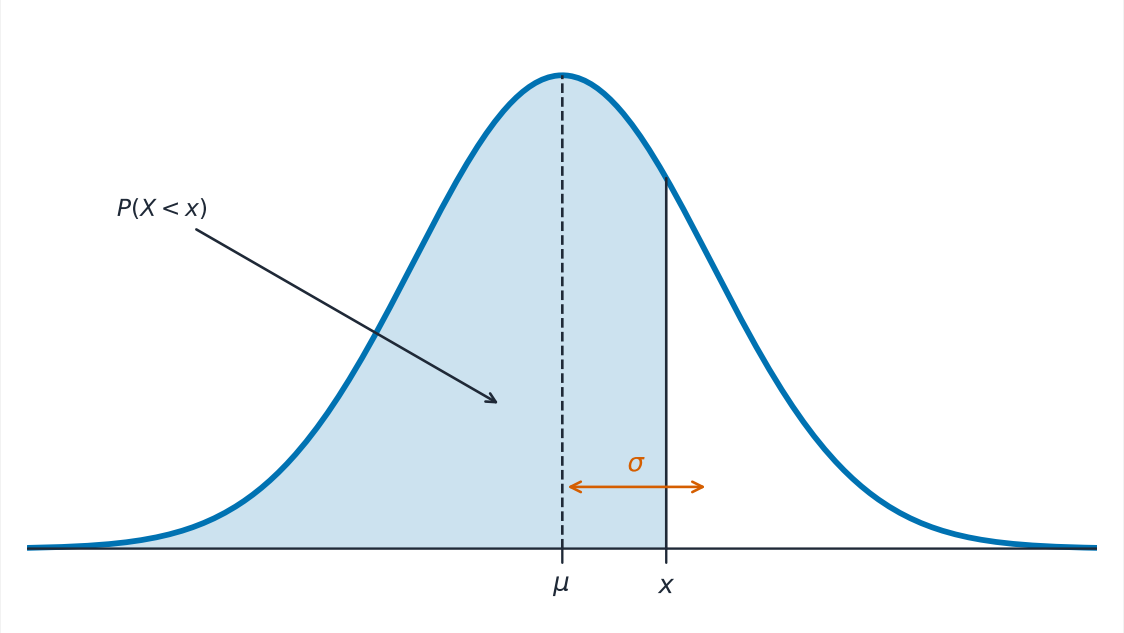
- Many things (heights, test scores) pile up in a "bell" shape.
- Most values are near the middle, with fewer far out.

Each idea has a worked example before the Practice.

New ideas

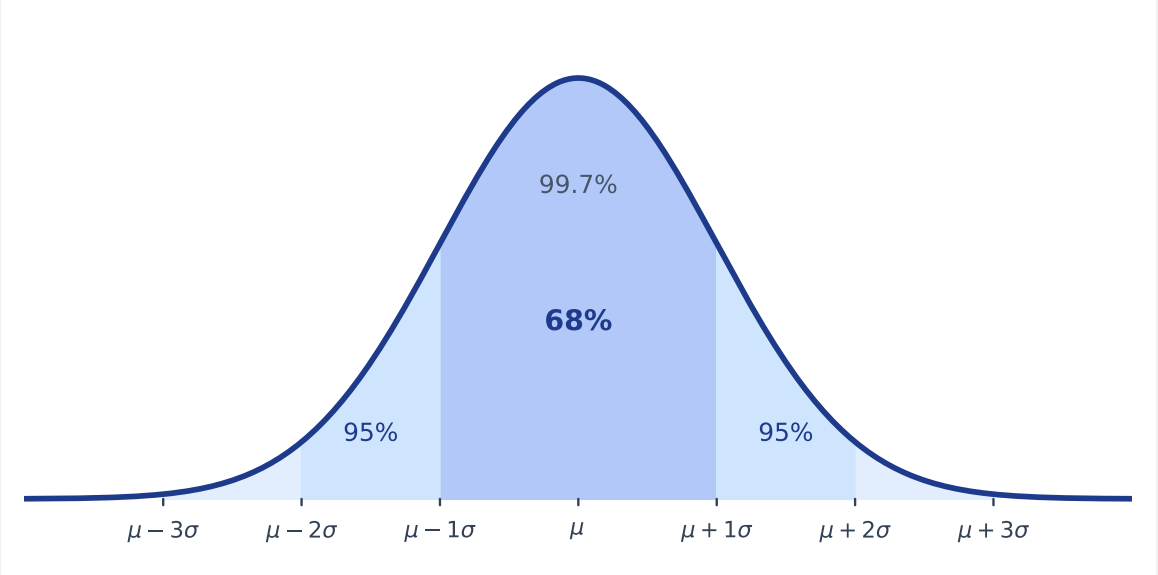
■ 1 The normal distribution

A **normal distribution** 正态分布 is the classic **bell curve** –symmetric, with most data near the mean and less as you move away.



2 The empirical rule

For a normal distribution, the **empirical rule** (the 68–95–99.7 rule) says:



- about **68%** of values lie within 1 standard deviation of the mean,
- about **95%** within 2 standard deviations,
- about **99.7%** within 3.

3 Using the rule

Worked example. Test scores are normal with mean 100 and standard deviation 10. Then about 95% of scores lie between $100 - 2(10) = 80$ and $100 + 2(10) = 120$.

4 Why it matters

The bell curve appears so often that these percentages let you quickly judge how unusual a value is.

Watch out: the 68–95–99.7 rule only applies to a **normal** (bell-shaped) distribution. For skewed data it does not hold –always check the shape first.

Practice

Try these in order. The methods are all above.

2.1 Describe the shape of a normal distribution. [2]

2.2 State the three percentages of the empirical rule. [3]

2.3 Heights are normal with mean 170 cm and standard deviation 8 cm. State the range

that holds about 68% of people. [2]

2.4 For the same heights, state the range that holds about 95%. [2]

2.5 State when the empirical rule does not apply. [1]

---

2 Scatterplots and correlation – Part 1

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	**中文**	**Write it 3 times** (English)
scatterplot	散点图	
Correlation	相关	

Recall

Often two things are linked:

- Taller people tend to weigh more.
- More study hours tend to mean higher marks.
- You can plot pairs of values to see a pattern.

This sheet lines up how to explore two variables together. Each idea has a worked example.

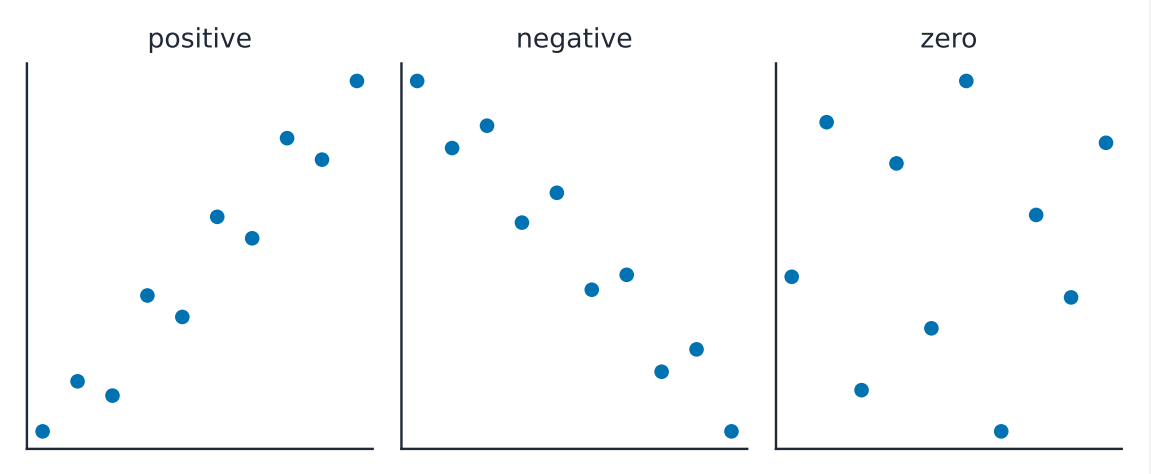
New ideas

■ 1 Scatterplots

A **scatterplot** 散点图 plots each pair of values as a dot, so you can see any pattern between the two variables.

■ 2 Correlation

Correlation 相关 describes the direction of a linear pattern:



- **positive** –as one rises, so does the other,
- **negative** –as one rises, the other falls,
- **none** –no clear pattern.

■ 3 Strength

The correlation coefficient r measures how strong and which way the pattern is: near $+1$ is a strong positive line, near -1 a strong negative one, near 0 no linear link.

■ 4 A caution

Worked example. Ice-cream sales and drownings both rise in summer –they are correlated, but neither causes the other. A third factor (hot weather) drives both.

Watch out: correlation is not causation. Two things rising together does not mean one causes the other –a hidden third factor may cause both.

Practice

Try these in order. The methods are all above.

2.1 State what a scatterplot shows. [2]

2.2 State what positive and negative correlation each mean. [2]

2.3 State what a correlation coefficient near $+1$ tells you. [2]

2.4 State what a correlation coefficient near 0 tells you. [1]

2.5 Ice-cream sales and sunburns are correlated. Explain why this does not mean ice cream causes sunburn. [2]

2 The line of best fit – Part 2

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	**中文**	**Write it 3 times** (English)
line of best fit	回归线	

Recall

In Part A you saw scatterplots and correlation. Now summarize the pattern with a line:

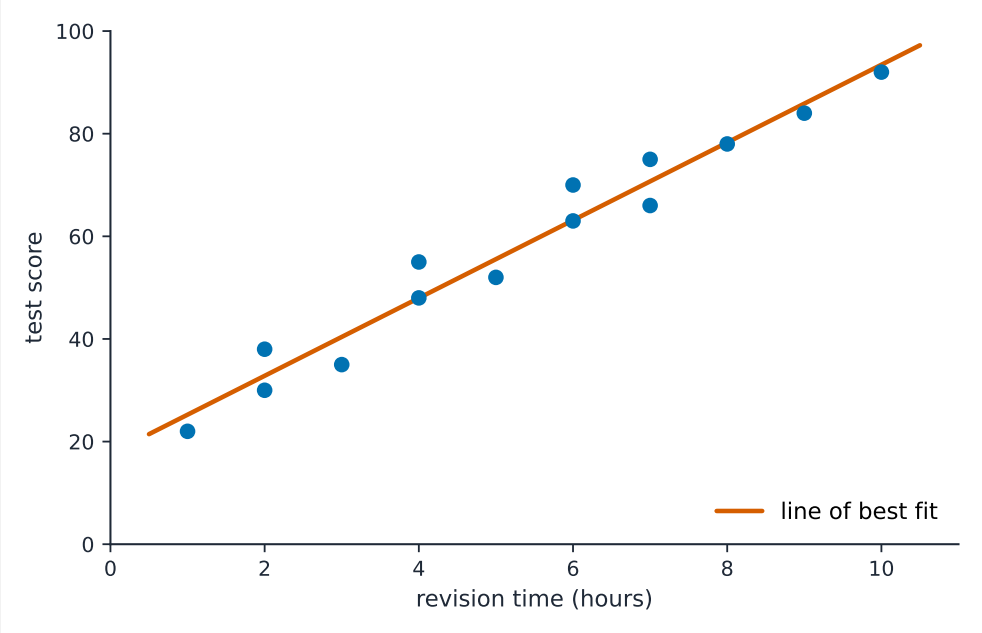
- A straight line can capture the trend in a scatterplot.
- That line lets you make predictions.

Each idea has a worked example before the Practice.

New ideas

■ 1 The line of best fit

The **line of best fit** 回归线 (also called the regression line) is the straight line that best follows the trend of the points.



2 Reading the line

The line has the form $\hat{y} = a + bx$. The **slope** b says how much y changes for each 1 unit of x; the **intercept** a is the predicted y when $x = 0$.

3 Making predictions

Worked example. If $\hat{y} = 50 + 5x$ predicts a test score from hours studied, then 3 hours predicts $\hat{y} = 50 + 5(3) = 65$.

4 Residuals

A **residual** is the gap between an actual value and the line's prediction (actual minus predicted). Small residuals mean the line fits well.

Watch out: avoid **extrapolating** –predicting far outside the range of your data. The pattern may not continue, so a line that fits well for 1–5 study hours may give nonsense for 20 hours.

Practice

Try these in order. The methods are all above.

2.1 State what the line of best fit does. [2]

2.2 For $\hat{y} = 50 + 5x$, state what the slope 5 means. [2]

2.3 Using $\hat{y} = 50 + 5x$, predict y when $x = 4$. [2]

2.4 State what a residual is. [2]

2.5 Explain why predicting far outside the data range is risky. [2]

3 Sampling – Part 1

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	中文	Write it 3 times (English)		
population	总体			
sample	样本			
random sample	随机抽样			
Bias	偏差			

Recall

To learn about a big group, you cannot ask everyone:

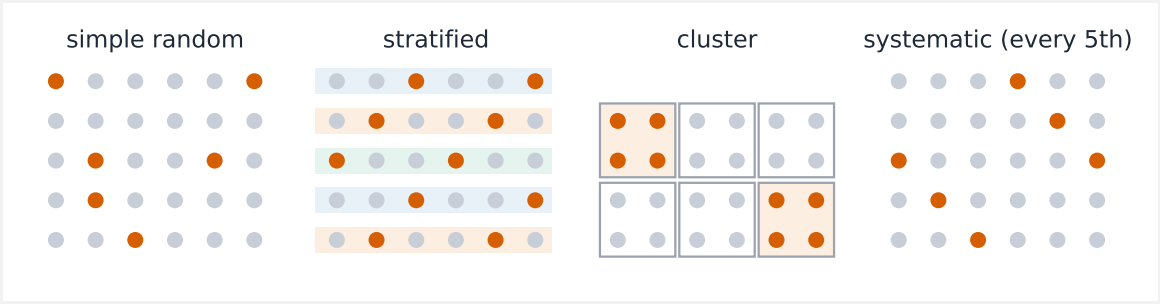
- To find the average height of a country, you cannot measure every person.
- Instead you measure a smaller group and use it to estimate.
- But the smaller group must be chosen fairly.

This sheet lines up how to collect data by sampling. Each idea has a worked example.

New ideas

■ 1 Population and sample

The **population** 总体 is the whole group you want to study; a **sample** 样本 is the smaller part you actually measure.



2 Random sampling

A **random sample** 随机抽样 gives everyone an equal chance of being chosen. This keeps the sample fair and representative.

3 Bias

Bias 偏差 is anything that makes a sample unrepresentative –for example, only surveying people online misses those without internet.

4 A worked example

Worked example. To learn students' opinions, surveying only the football team is biased (not representative); drawing names randomly from the whole school is fair.

Watch out: a **biased** sample leads to wrong conclusions no matter how **large** it is. A huge but unfair sample is worse than a smaller, properly random one.

Practice

Try these in order. The methods are all above.

2.1 State the difference between a population and a sample. [2]

2.2 State what makes a sample "random". [2]

2.3 State what "bias" means in sampling. [2]

2.4 A survey about a town is done only outside a gym. Explain why this could be

biased. [2]

2.5 State whether a large but biased sample or a smaller random sample is more trustworthy, and why. [2]

3 Experiments – Part 2

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	中文	Write it 3 times (English)
experiment	实验	
control group	对照组	
randomization	随机化	

Recall

In Part A you met sampling. Now how to test whether something actually **causes** an effect:

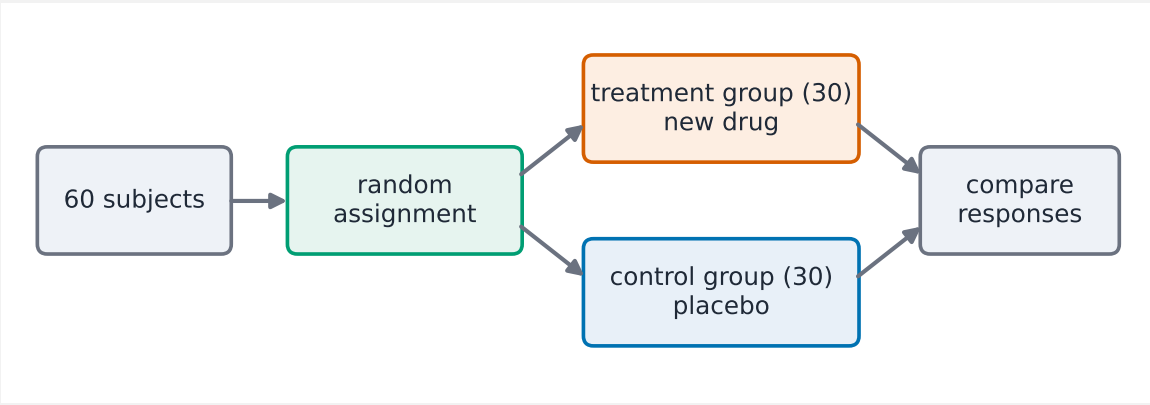
- Does a new medicine really work?
- Just observing is not enough –you need a fair test.

Each idea has a worked example before the Practice.

New ideas

1 Experiment versus observation

- An **observational study** just watches and records, without changing anything.
- An **experiment** 实验 deliberately applies a **treatment** to see its effect.



2 Control and treatment groups

An experiment compares a **treatment group** (gets the new thing) with a **control group** 对照组 (does not) –so any difference can be linked to the treatment.

3 Randomization

Subjects are put into groups by **randomization** 随机化, so the two groups are similar to begin with and the comparison is fair.

4 Only experiments show causation

Worked example. To test a medicine, randomly split patients: half get the drug, half get a placebo. Because the groups start alike, a difference in recovery can be blamed on the drug.

Watch out: only a well-designed **experiment** can show **causation**. An observational study can find a link (correlation) but cannot prove that one thing causes another.

Practice

Try these in order. The methods are all above.

2.1 State the difference between an observational study and an experiment. [2]

2.2 State the roles of the treatment group and the control group. [2]

2.3 State why subjects are randomly assigned to groups. [2]

2.4 State which kind of study can show that one thing causes another. [1]

2.5 Explain why a control group is needed to judge a new medicine. [2]

4 Basic probability – Part 1

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	中文	Write it 3 times (English)
Probability	概率	

Recall

You have a sense of chance already:

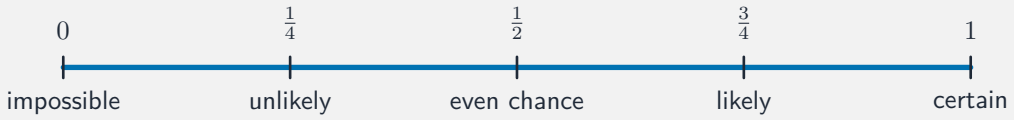
- A coin lands heads about half the time.
- A rolled die is equally likely to show 1 to 6.
- Some events are more likely than others.

This sheet lines up the basics of probability. Each idea has a worked example.

New ideas

1 The probability scale

Probability 概率 measures how likely an event is, from 0 (impossible) to 1 (certain).



2 Equally likely outcomes

When outcomes are equally likely,

$$P(\text{event}) = \frac{\text{favourable outcomes}}{\text{total outcomes}}.$$

Worked example. Rolling a 4 on a die: $P = \frac{1}{6}$. Rolling an even number: $P = \frac{3}{6} = \frac{1}{2}$.

■ 3 Complement

The chance an event does **not** happen is 1 minus the chance it does: $P(\text{not } A) = 1 - P(A)$.

Worked example. If $P(\text{rain}) = 0.3$, then $P(\text{no rain}) = 1 - 0.3 = 0.7$.

■ 4 Combining events

For two independent events (one does not affect the other), the chance **both** happen is the product of their chances.

Watch out: a probability is always between 0 and 1 (or 0% and 100%). An answer above 1 or below 0 is a sure sign of a mistake.

Practice

Try these in order. The methods are all above.

2.1 State what a probability of 0 and of 1 each mean. [2]

2.2 A die is rolled. Find the probability of rolling a 5. [2]

2.3 Find the probability of rolling an odd number on a die. [2]

2.4 If $P(\text{win}) = 0.4$, find $P(\text{not win})$. [2]

2.5 State the range that any probability must lie in. [1]

4 Random variables and distributions – Part 2

Name: _____ Class: _____ Date: _____

Recall

In Part A you found probabilities of single events. Now handle chance that varies:

- A number that depends on chance (like a dice total) is a random variable.
- Its possible values each have a probability.

Each idea has a worked example before the Practice.

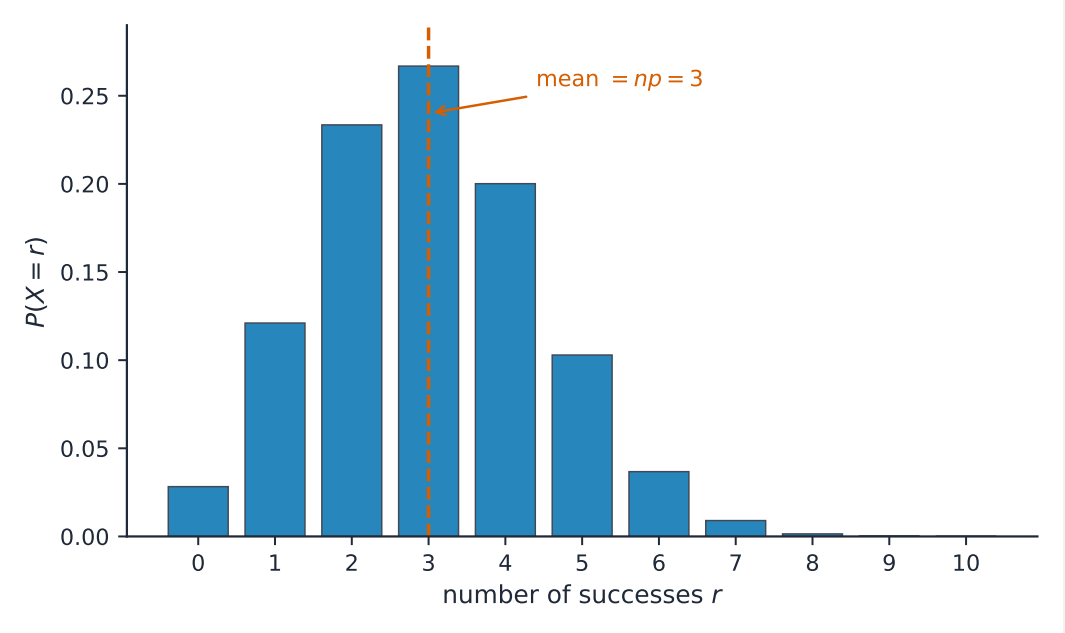
New ideas

1 Random variables

A **random variable** is a number whose value depends on chance –like the number of heads in 3 coin flips.

2 Probability distributions

A **probability distribution** lists each possible value and its probability. All the probabilities add up to 1.



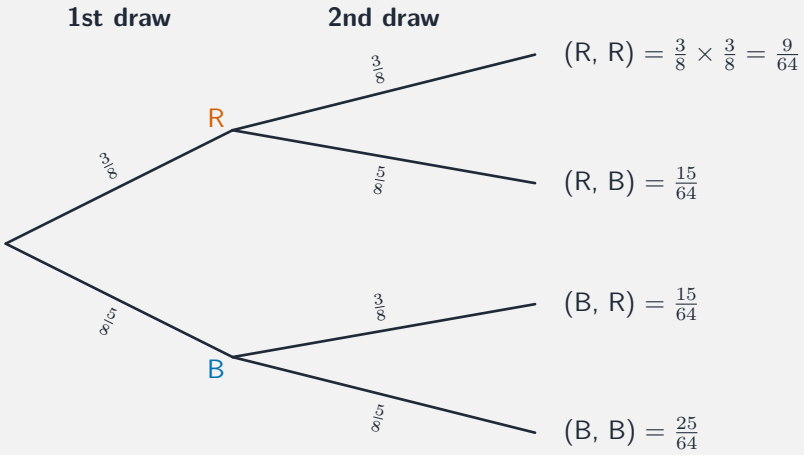
3 Expected value

The **expected value** (mean) is the long-run average: multiply each value by its probability and add.

Worked example. A game pays \$0 with probability 0.5 and \$10 with probability 0.5: expected value $= 0(0.5) + 10(0.5) = \$5$.

4 Tree diagrams

A **tree diagram** maps out the outcomes of several stages, so you can multiply probabilities along the branches.



Watch out: in any probability distribution, all the probabilities must add up to exactly 1. If your listed probabilities do not sum to 1, something is missing or wrong.

Practice

Try these in order. The methods are all above.

2.1 State what a random variable is. [2]

2.2 State what all the probabilities in a distribution must add up to. [1]

2.3 A prize is \$0 with probability 0.8 and \$20 with probability 0.2. Find the expected value. [3]

2.4 A distribution lists probabilities 0.3, 0.3, and 0.3. State why something must be wrong. [2]

2.5 State what a tree diagram is used for. [1]

5 Sampling variability – Part 1

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	中文	Write it 3 times (English)
sampling distribution	抽样分布	

Recall

You know a sample is used to estimate a whole population:

- But two different samples will give slightly different answers.
- The sample mean is not exactly the true mean –it varies.

This sheet lines up how sample results vary. Each idea has a worked example.

New ideas

1 Statistics vary

A **statistic** (like a sample mean) changes from one sample to the next, just by chance. This is **sampling variability**.

2 The sampling distribution

The **sampling distribution** 抽样分布 is the distribution of a statistic over many, many samples –it shows how much the statistic bounces around.

3 Centered on the truth

For a fair sample, the sampling distribution is **centered** on the true population value –so on average, the sample gives the right answer.

4 Bigger samples, less spread

Worked example. Estimating an average with a sample of 100 gives a much less variable result than a sample of 10 –larger samples give more reliable estimates.

Watch out: the sample mean is almost never **exactly** equal to the population mean –it varies by chance. What we can say is how **close** it is likely to be, and larger

> samples make it closer.

Practice

Try these in order. The methods are all above.

2.1 State what ”sampling variability” means. [2]

2.2 State what a sampling distribution shows. [2]

2.3 State where the sampling distribution of a fair sample is centered. [1]

2.4 State the effect of a larger sample on the variability of an estimate. [2]

2.5 Explain why two different samples from the same population can give different means. [2]

5 The central limit theorem – Part 2

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	**中文**	**Write it 3 times** (English)
central limit theorem	中心极限定理	

Recall

In Part A you saw that sample means vary. There is a beautiful pattern to how they vary:

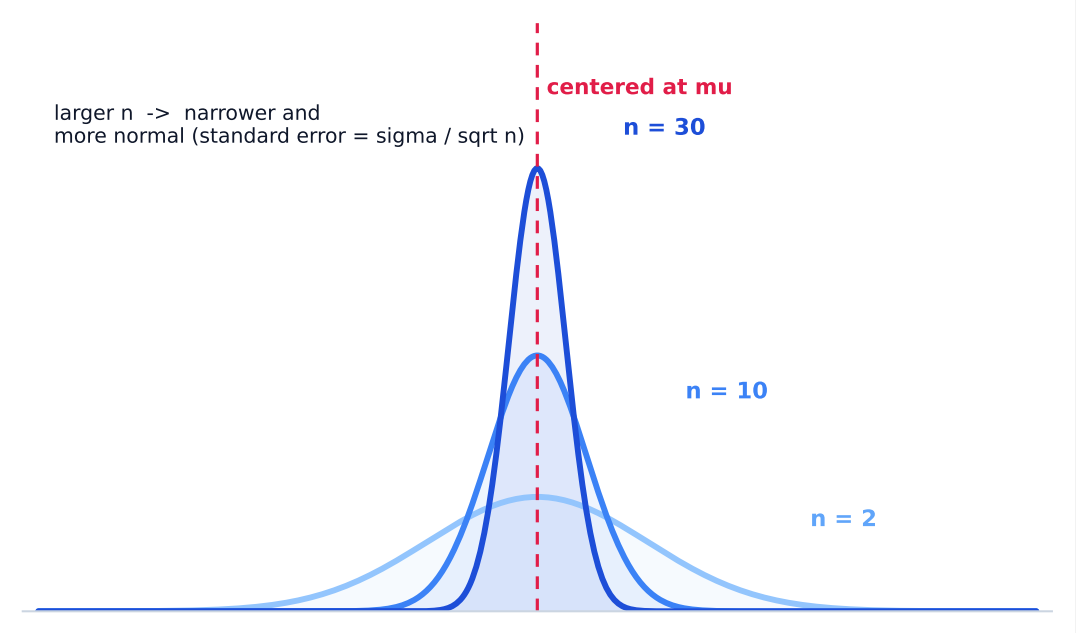
- Even if the original data are not bell-shaped, the sample means often are.
- This pattern is the key that makes statistics work.

Each idea has a worked example before the Practice.

New ideas

■ 1 The central limit theorem

The **central limit theorem** 中心极限定理 says: for a large enough sample, the distribution of the **sample mean** is approximately **normal** (bell-shaped), centered on the population mean.



2 Even from non-normal data

Remarkably, this holds **even if the original population is not normal** –the averaging smooths things into a bell shape.

3 Larger samples, tighter bell

The bigger the sample, the narrower and more normal the distribution of the sample mean becomes.

4 Why it matters

Because the sample mean is approximately normal, we can use the familiar bell curve (and the 68–95–99.7 rule) to say how confident we are in an estimate –the basis of all the inference that follows.

Watch out: the central limit theorem is about the distribution of the **sample mean**, not the individual data. The raw data can be skewed while the sample means are still bell-shaped.

Practice

Try these in order. The methods are all above.

2.1 State what the central limit theorem says about the sample mean. [2]

2.2 State the surprising part –what it does not require of the population. [2]

2.3 State what happens to the sample mean's distribution as the sample gets larger. [2]

2.4 State why the central limit theorem is so useful in statistics. [2]

2.5 State whether the theorem is about the individual data or the sample mean. [1]

6 Confidence intervals – Part 1

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	**中文**	**Write it 3 times** (English)
confidence interval	置信区间	

Recall

You use a sample to estimate something about a population:

- A poll estimates the fraction of people who support an idea.
- But the estimate is not exact –how sure can we be?

This sheet introduces confidence intervals. Each idea has a worked example.

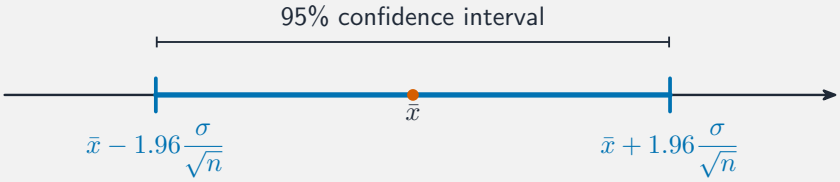
New ideas

1 A range, not a single number

Because a sample estimate varies, we give a **range** likely to contain the true value, not just one number.

2 The confidence interval

A **confidence interval** 置信区间 is an estimate plus-or-minus a margin of error: ”the true value is probably in this range”.



Worked example. A poll finds 60% support with a margin of error of ±3%, giving a confidence interval of 57% to 63%.

3 Confidence level

A 95% **confidence level** means the method captures the true value 95% of the time it is used –we are 95% confident the range is right.

4 Wider or narrower

A larger sample gives a **smaller** margin of error (a narrower, more precise interval); a higher confidence level gives a **wider** one.

Watch out: the confidence interval estimates the **population** value, not a single person. And "95% confident" refers to the method's long-run success rate, not a probability about this one interval.

Practice

Try these in order. The methods are all above.

2.1 State why we give a range instead of a single estimate. [2]

2.2 A poll finds 45% with a margin of error of ±4%. State the confidence interval. [2]

2.3 A poll finds 70% with a margin of error of ±2%. State the confidence interval. [2]

2.4 State what a larger sample does to the margin of error. [1]

2.5 State what a ”95% confidence level” means. [2]

6 Hypothesis tests – Part 2

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	中文	Write it 3 times (English)
hypothesis test	假设检验	

Recall

In Part A you estimated a value with a confidence interval. Now test a claim:

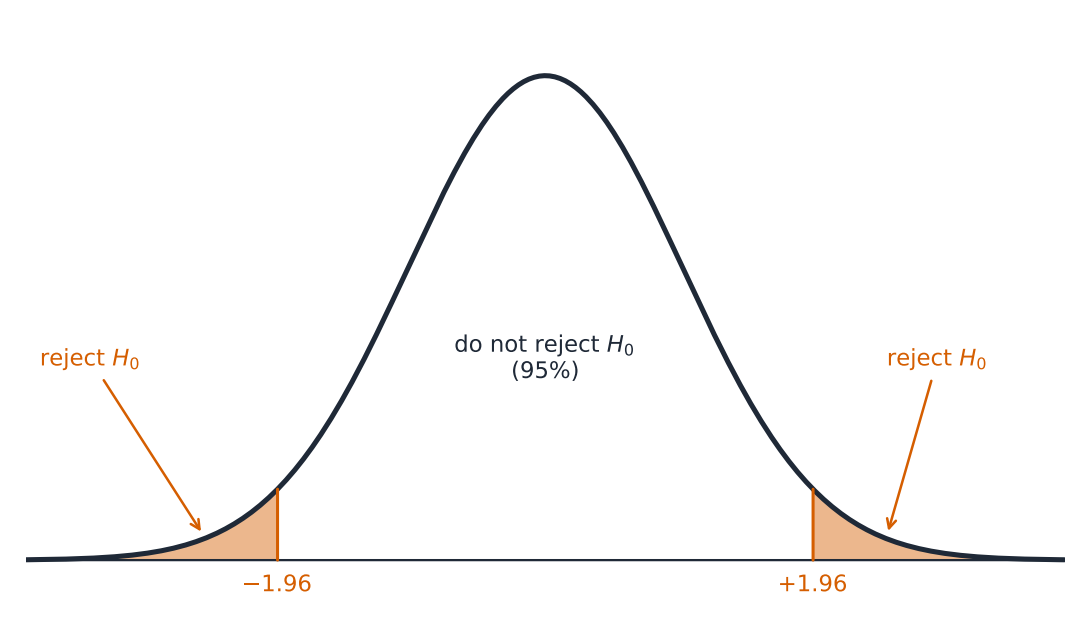
- Is a coin fair, or biased?
- Does the data give real evidence, or could it just be chance?

Each idea has a worked example before the Practice.

New ideas

■ 1 A hypothesis test

A **hypothesis test** 假设检验 weighs evidence against a claim. It starts with a **null hypothesis** –"nothing unusual is going on" (for example, the coin is fair).



2 The p-value

The **p-value** is the chance of getting data at least this extreme **if the null hypothesis were true**. A **small** p-value means the data would be surprising, which is evidence against the null.

3 The decision

If the p-value is below a chosen level (often 0.05), we **reject** the null hypothesis –the effect is ”statistically significant”. Otherwise we do not reject it.

Worked example. A coin gives 9 heads in 10 flips. If the coin were fair, that is very unlikely (small p-value), so we have evidence it is biased.

4 What it does not prove

Failing to reject the null does **not** prove it true –it just means there was not enough evidence against it.

Watch out: a small p-value means the data are **surprising if the null is true** –it does not directly give ”the probability the null is true”. And ”not significant” means ”not enough evidence”, not ”proven no effect”.

Practice

Try these in order. The methods are all above.

2.1 State what the null hypothesis usually represents. [2]

2.2 State what a p-value measures. [2]

2.3 State what a small p-value tells you. [2]

2.4 A test gives a p-value of 0.02 (below 0.05). State the decision. [1]

2.5 Explain why failing to reject the null does not prove it true. [2]

7 Estimating a mean – Part 1

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	中文	Write it 3 times (English)
mean	均值	

Recall

In an earlier sheet you estimated a **proportion** with a confidence interval. The same idea works for an **average**:

- A sample of test scores estimates the true average score.
- The estimate comes with a margin of error.

This sheet lines up estimating a mean. Each idea has a worked example.

New ideas

1 Estimating a mean

To estimate a population **mean** 均值, take a sample, find its average, and attach a margin of error to give a confidence interval.

2 The interval

The confidence interval for a mean is sample mean ± margin of error.

Worked example. A sample of heights averages 170 cm with a margin of error of ±3 cm: the interval is 167 cm to 173 cm.

3 Interpretation

We are (say) 95% confident the true average lies in this range –the method captures it 95% of the time.

4 Precision

A larger sample gives a **smaller** margin of error, so a narrower, more precise estimate of the mean.

> **Watch out:** the confidence interval estimates the **population mean**, not the value of a single individual. Individuals vary far more widely than the average does.

Practice

Try these in order. The methods are all above.

2.1 State what you attach to a sample mean to make a confidence interval. [1]

2.2 A sample averages 50 with a margin of error of ±4. State the confidence interval. [2]

2.3 A sample averages 200 with a margin of error of ±15. State the interval. [2]

2.4 State what a larger sample does to the margin of error. [1]

2.5 State what the interval estimates –the population mean or a single individual. [1]

7 The t-distribution – Part 2

Name: _____ Class: _____ Date: _____

Recall

In Part A you estimated a mean. In real life you rarely know the population's spread:

- You must estimate the spread from the sample too.
- This extra uncertainty needs a slightly different curve.

Each idea has a worked example before the Practice.

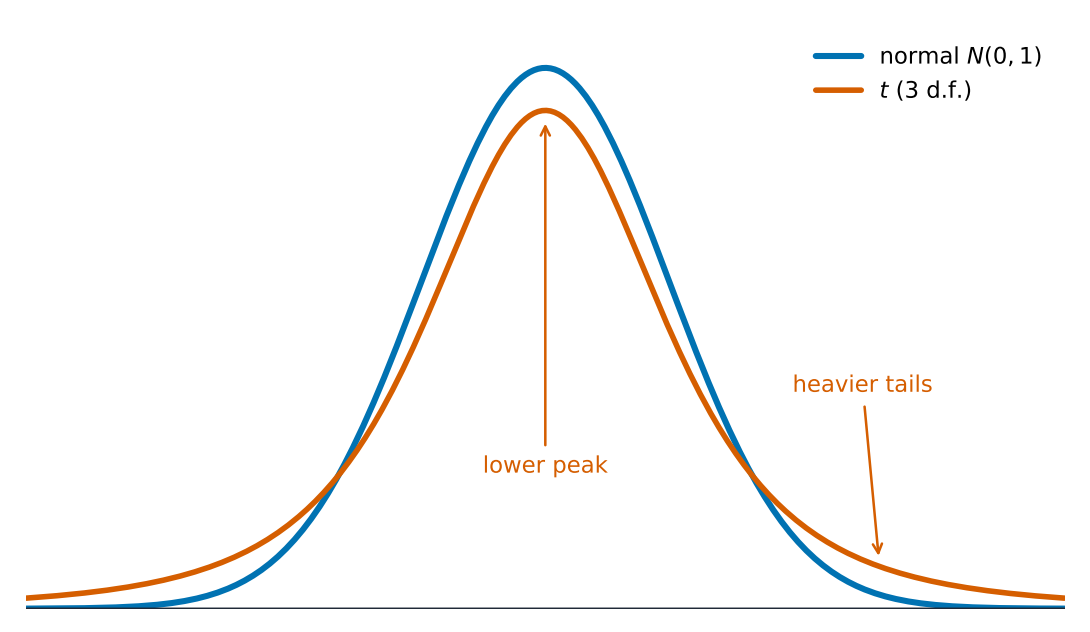
New ideas

■ 1 When the spread is unknown

When you do not know the population's standard deviation (almost always), you estimate it from the sample. This adds uncertainty, especially for small samples.

■ 2 The t-distribution

To handle that extra uncertainty, we use the **t-distribution** instead of the normal curve.



It looks like the bell curve but has **heavier tails**, allowing for more surprises with small samples.

■ **3 Sample size matters**

The smaller the sample, the heavier the tails (more uncertainty). As the sample grows, the t-distribution gets closer and closer to the normal curve.

■ **4 Same idea, adjusted**

Worked example. A confidence interval for a mean from a small sample uses a slightly wider margin (from the t-distribution) than it would with the normal curve –reflecting the extra uncertainty.

Watch out: use the **t-distribution** (not the normal) whenever you estimate the spread from the sample –which is nearly always. The two agree for large samples, but differ for small ones.

Practice

Try these in order. The methods are all above.

2.1 State why we often cannot use the population's standard deviation. [2]

2.2 State what distribution we use instead of the normal in that case. [1]

2.3 State how the t-distribution differs in shape from the normal curve. [2]

2.4 State what happens to the t-distribution as the sample size grows. [2]

2.5 State when the t-distribution should be used. [1]

8 Categorical data and expected counts – Part 1

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	中文	Write it 3 times (English)		
Categorical data	分类数据			
expected	期望			

Recall

Not all data are numbers –some are categories:

- Eye colour, favourite subject, or blood type are categories.
- You count how many fall into each category.

This sheet lines up how to compare category counts. Each idea has a worked example.

New ideas

1 Categorical data

Categorical data 分类数据 sort things into groups (categories) rather than measuring a number. You summarize them with **counts** in each category.

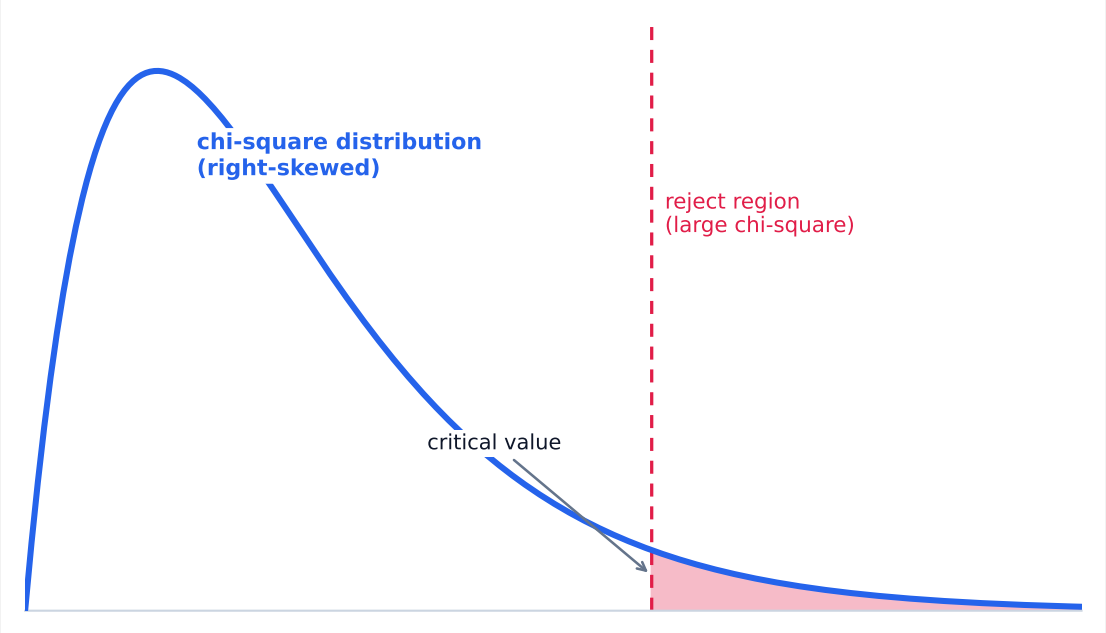
2 Observed and expected

An **observed** count is what you actually got; an **expected** 期望 count is what you would expect if a claim were true.

Worked example. Rolling a fair die 60 times, you **expect** 10 of each number. If you **observe** 15 sixes, that is more than expected.

3 Measuring the gap

To judge whether the observed counts differ from expected by more than chance, we compare each observed count with its expected count.



4 The question

The key question: are the differences between observed and expected small enough to be chance, or big enough to be real?

Watch out: the **expected** counts are what a claim (like "the die is fair") predicts – they are not the observed data. The whole test compares what you **got** against what the claim **predicts**.

Practice

Try these in order. The methods are all above.

2.1 State what categorical data are, with an example. [2]

2.2 State the difference between an observed and an expected count. [2]

2.3 A fair coin is flipped 100 times. State the expected number of heads. [2]

2.4 A fair die is rolled 120 times. State the expected count for each number. [2]

2.5 State the key question when comparing observed and expected counts. [2]

8 The chi-square test – Part 2

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	**中文**	**Write it 3 times** (English)
chi-square	卡方	

Recall

In Part A you compared observed counts with expected counts. Now measure the total gap with one number:

- We need a single value for "how far off" the observed counts are.
- That value is the chi-square statistic.

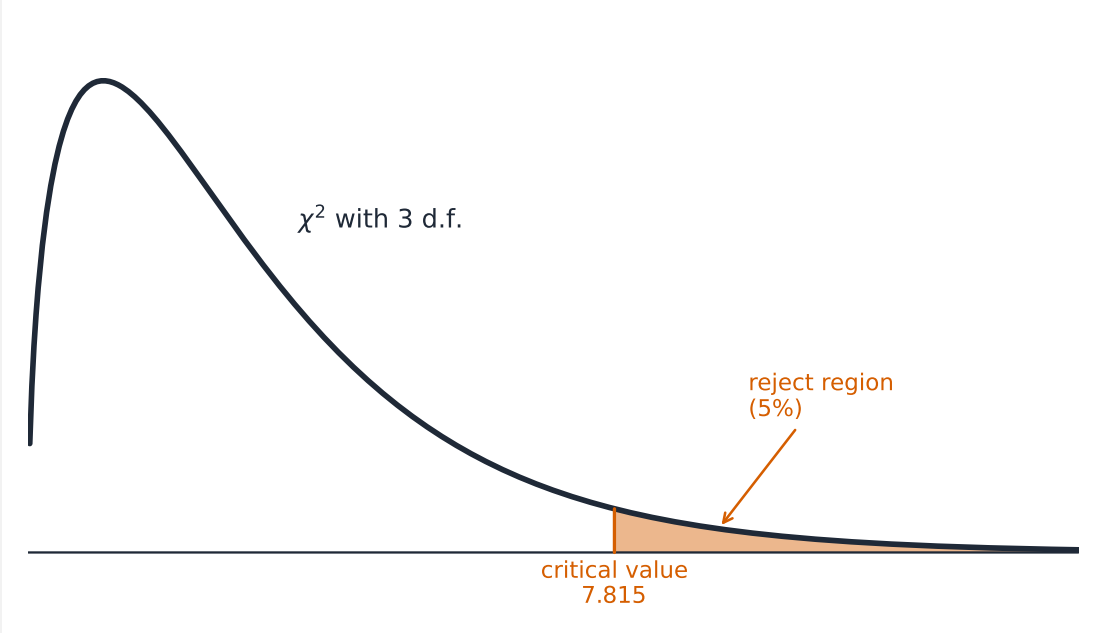
Each idea has a worked example before the Practice.

New ideas

■ 1 The chi-square statistic

The **chi-square** 卡方 statistic adds up the (observed minus expected) gaps, squared and scaled:

$$\chi^2 = \sum \frac{(O-E)^2}{E}.$$



A **bigger** χ^2 means the observed counts are further from expected.

2 A worked example

Worked example. A cross predicts 60 dominant and 20 recessive, but you observe 55 and 25:

$$\chi^2 = \frac{(55-60)^2}{60} + \frac{(25-20)^2}{20} = \frac{25}{60} + \frac{25}{20} = 0.42 + 1.25 = 1.67.$$

3 Making a decision

Compare χ^2 with a **critical value** from a table. If χ^2 is **smaller**, the differences are within chance (do not reject); if **larger**, they are a real effect (reject).

4 In the example

With 1 degree of freedom the critical value at the 0.05 level is 3.84. Since $1.67 < 3.84$, the deviation is within chance –the data fit the predicted ratio.

Watch out: always divide by the **expected** count E (not the observed) in each term. And a small χ^2 means the data **fit** the claim –it does not mean "no data".

Practice

Try these in order. The methods are all above.

2.1 Write the formula for the chi-square statistic. [2]

2.2 State what a large chi-square value tells you. [2]

2.3 For one category, observed $= 18$ and expected $= 12$. Find its term $\dfrac{(O-E)^2}{E}$. [3]

2.4 A test gives $\chi^2 = 1.2$ and the critical value is 3.84. State the decision. [2]

2.5 State what you divide by in each term of the chi-square sum. [1]

9 Inference for a regression slope – Part 1

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	**中文**	**Write it 3 times** (English)
slope	斜率	

Recall

Earlier you fitted a line of best fit to a scatterplot:

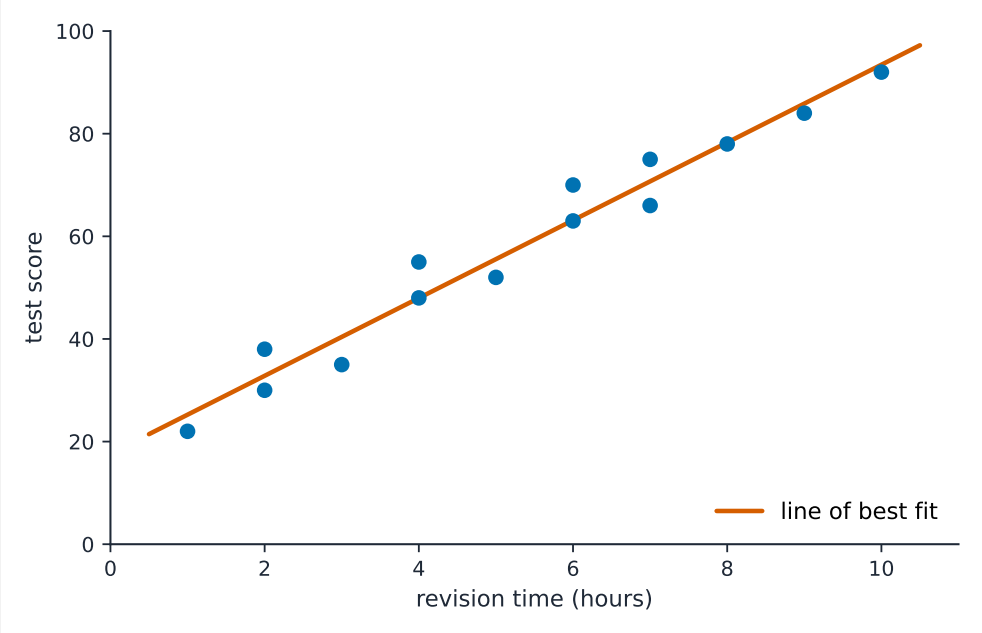
- The line's slope describes the relationship.
- But is that slope real, or could it just be chance in the sample?

This sheet lines up inference about a regression line. Each idea has a worked example.

New ideas

■ 1 The sample slope estimates the true slope

The **slope** 斜率 of the line of best fit, found from a sample, is only an **estimate** of the true relationship in the whole population.



2 Is there a real relationship?

If the true slope is 0, there is **no** linear relationship. So we test whether the slope is really different from 0.

3 A confidence interval for the slope

We can also give a confidence interval for the slope –a range likely to contain the true slope.

Worked example. If a sample slope is 5 with a margin of error of ± 2, the interval is 3 to 7. Since this does not include 0, there is evidence of a real positive relationship.

4 The idea reused

This is the same inference idea as before: a sample estimate, plus uncertainty, used to judge a claim (here, "the slope is 0").

Watch out: if a slope's confidence interval **includes** 0, you cannot conclude there is a real relationship –the true slope might be 0 (no link at all).

Practice

Try these in order. The methods are all above.

2.1 State what the sample slope is an estimate of. [2]

2.2 State what a true slope of 0 would mean. [2]

2.3 A sample slope is 8 with a margin of error of ±3. State the confidence interval. [2]

2.4 For that interval, state whether there is evidence of a real relationship, and why. [2]

2.5 State what it means if a slope's interval includes 0. [1]

9 Residuals and checking the model – Part 2

Name: _____ Class: _____ Date: _____

Vocabulary 词汇

Write each word three times. 把每个单词抄写三遍。

English	中文	Write it 3 times (English)
residual	残差	

Recall

In Part A you tested a regression slope. Now check whether a line is even the **right** model:

- A straight line is not always the best fit.
- The leftover errors (residuals) reveal whether it fits.

Each idea has a worked example before the Practice.

New ideas

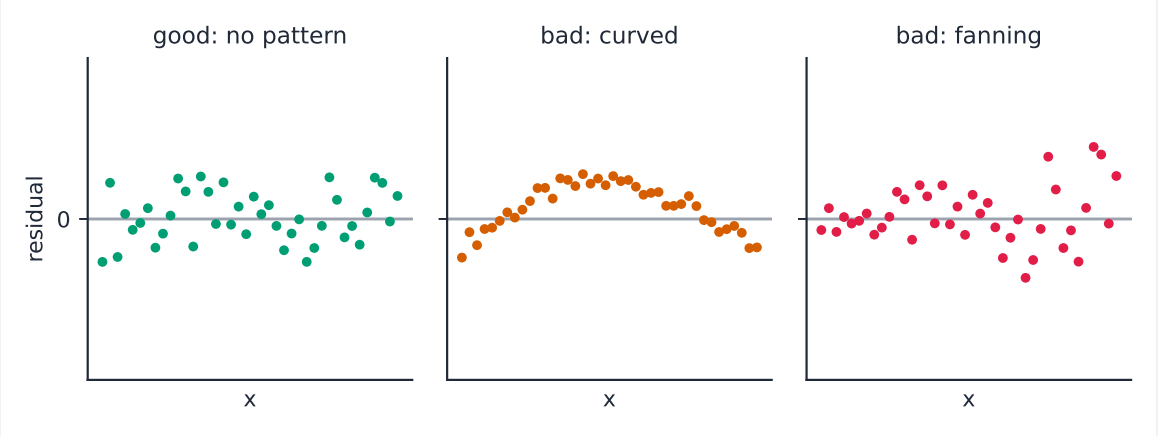
■ 1 Residuals

A **residual** 残差 is the gap between an actual value and the line's prediction: residual = actual − predicted.

Worked example. If the line predicts 70 but the actual value is 73, the residual is $73 - 70 = +3$.

■ 2 The residual plot

A **residual plot** graphs the residuals against x. It magnifies how the line misses the points.



3 No pattern is good

If the residual plot shows **no pattern** (a random scatter around zero), a straight line is a good model.

4 A pattern is bad

If the residual plot shows a **curve** or other pattern, a straight line is the **wrong** model –the data bend in a way the line cannot follow.

Watch out: a good model has a residual plot with **no pattern**. A clear pattern (like a U-shape) is a warning that a straight line is the wrong choice, even if the correlation looked high.

Practice

Try these in order. The methods are all above.

2.1 State what a residual is. [2]

2.2 A line predicts 50 but the actual value is 46. Find the residual. [2]

2.3 State what a residual plot graphs. [2]

2.4 State what a residual plot with no pattern tells you. [2]

2.5 State what a clear pattern (like a curve) in the residual plot tells you. [2]