Past-paper walk-through

Introduction to Experimental Design ▪ 3.5

eXercise

Questions in this set

- 2026 Q2
- 2024 Q3
- 2023 Q2
- 2022 Q2
- 2019 Q2
- 2017 Q6
- 2016 Q3

Reading the mark scheme

- **B1** a mark that stands on its own – the point is either made or not.
- **M1** a *method* mark: the right approach, even if the number is wrong.
- **A1** an *answer* mark, and it usually needs its M mark first.
- **C1** a working mark on the way to the answer.
- // separates answers the examiner accepts equally.
- **ecf** = error carried forward: a later part is marked on your own earlier answer.
- **owtte** = "or words to that effect" – the wording need not match.

Question 2

2026 ▪ Q2

Holly, a botanist, read an online report that stated that adding coffee grounds to the soil of rosebushes will produce more roses. Holly decides to conduct an experiment to investigate this claim. She grows 30 rosebushes in a greenhouse with a controlled environment so the conditions will be the same for all rosebushes.

When the rosebushes are a month old, she randomly assigns 15 rosebushes to have one-half cup of coffee grounds added to their soil weekly. The other 15 rosebushes will not have coffee grounds added to their soil.

After three months, she counts the number of roses on each rosebush.

botanist: a scientist who studies the biology of plants

(a)(i) Based on the information provided about Holly's experiment, identify each of the following.

Treatments

Question 2

(a)(ii) Experimental units
(a)(iii) Response variable
(b) In the context of Holly's experiment, describe how the treatments can be randomly assigned to the experimental units so that each treatment has the same number of units.
(c) After the experiment, Holly determines that her results were statistically significant at the $\alpha = 0.05$ level of significance. Explain the meaning of "statistically significant" in this context.

Question 3

2024 ▪ Q3

A car maker produces four different models of cars: A, B, C, and D. A group of researchers is investigating which model of car has the longest distance traveled per gallon of gas (mileage). Higher mileage is considered better than lower mileage. The researchers will conduct a study in which they contact several owners of each model of car and ask them to estimate their mileage.

(a) Is this an observational study or an experiment? Justify your answer in context.

Question 3

2024 ▪ Q3

A car maker produces four different models of cars: A, B, C, and D. A group of researchers is investigating which model of car has the longest distance traveled per gallon of gas (mileage). Higher mileage is considered better than lower mileage. The researchers will conduct a study in which they contact several owners of each model of car and ask them to estimate their mileage.

(b) Model D has an autopilot feature, in which the car controls its own motion with human supervision. James owns a Model D car and will investigate whether using the autopilot feature results in higher mileage than not using the autopilot. James will drive his car on 70 different days to and from work, using the same route at the same time each day. James will record the mileage each day.

James will use a completely randomized design to conduct his investigation. Describe an appropriate method James could use to randomly assign the two treatments, driving

Question 3

using the autopilot feature and driving without using the autopilot feature, to 35 days each.

Question 3

2024 ▪ Q3

A car maker produces four different models of cars: A, B, C, and D. A group of researchers is investigating which model of car has the longest distance traveled per gallon of gas (mileage). Higher mileage is considered better than lower mileage. The researchers will conduct a study in which they contact several owners of each model of car and ask them to estimate their mileage.

(c) After the investigation was completed, James verified that the conditions for inference were met and conducted a hypothesis test. He discovered the mean mileage when using the autopilot feature was significantly higher than the mean mileage when not using the autopilot feature.

James is a member of a Model D club with thousands of members who all drive Model D cars. He will give a presentation at a Model D club members' meeting later this year and would like to state that the results of his hypothesis test apply to all Model D cars in his club. Another member of the club who is a statistician tells James his findings

Question 3

do not apply to all Model D cars in the club. What change would James need to make to his original study to be able to generalize to all Model D cars in the club?

Solution 3

(a)

Mark scheme

- This is an observational study. The researchers had the car owners estimate their own mileage; the car owners were not randomly assigned a car model, so no treatment was imposed by the researchers. Full credit needs: (1) identifies "observational study" (stating it is an experiment scores Incorrect), (2) justification based on no treatment being imposed (e.g. "no random assignment of treatment"), (3) context – car models, drivers, cars, owners, or members (mileage alone is not sufficient context).

Solution 3

(b) Mark scheme

- Number the 70 days in the experiment from 1 to 70. Using a random number generator, generate 35 unique integers from 1 to 70 inclusive (no repeats). Assign the days with those 35 unique integers for James to drive the car WITH autopilot, and assign the remaining 35 days for James to drive WITHOUT autopilot. (Equivalent alternative: label 70 equally-sized slips of paper, 35 ”with autopilot” and 35 ”without autopilot,” mix them in a bag, and each day select one slip without replacement to determine that day’s driving method.) Full credit needs: (1) appropriate labels created for the units/treatments (the 70 days, and the two treatments), (2) a correctly described random process so every possible random assignment is equally likely (a coin-flip-with-stopping-rule method does NOT

Solution 3

satisfy this, since it biases the last few days), (3) the process results in exactly 35 days assigned to each treatment, using selection without replacement/without repeats.

Solution 3

(c) Mark scheme

- In order to generalize his findings to all Model D cars in his club, James would need to randomly select Model D cars from the club, then carry out a similar study using the Model D cars that were randomly sampled from the club. Full credit needs: (1) states that more cars must be sampled, (2) states that random sampling is required for generalization, (3) is in context, including sampling from the population of interest (Model D cars/owners/members in his club). Any discussion of experimental design beyond sample selection (e.g. conditions, two-sample design) is ignored in scoring.

Question 2

2023 ▪ Q2

A developer wants to know whether adding fibers to concrete used in paving driveways will reduce the severity of cracking, because any driveway with severe cracks will have to be repaired by the developer. The developer conducts a completely randomized experiment with 60 new homes that need driveways. Thirty of the driveways will be randomly assigned to receive concrete that contains fibers, and the other 30 driveways will receive concrete that does not contain fibers. After one year, the developer will record the severity of cracks in each driveway on a scale of 0 to 10, with 0 representing not cracked at all and 10 representing severely cracked.

(a) Based on the information provided about the developer's experiment, identify each of the following.

- Experimental units

Question 2

- Treatments
- Response variable

Question 2

2023 ▪ Q2

A developer wants to know whether adding fibers to concrete used in paving driveways will reduce the severity of cracking, because any driveway with severe cracks will have to be repaired by the developer. The developer conducts a completely randomized experiment with 60 new homes that need driveways. Thirty of the driveways will be randomly assigned to receive concrete that contains fibers, and the other 30 driveways will receive concrete that does not contain fibers. After one year, the developer will record the severity of cracks in each driveway on a scale of 0 to 10, with 0 representing not cracked at all and 10 representing severely cracked.

(b) Describe an appropriate method the developer could use to randomly assign concrete that contains fibers and concrete that does not contain fibers to the 60 driveways.

Question 2

2023 ▪ Q2

A developer wants to know whether adding fibers to concrete used in paving driveways will reduce the severity of cracking, because any driveway with severe cracks will have to be repaired by the developer. The developer conducts a completely randomized experiment with 60 new homes that need driveways. Thirty of the driveways will be randomly assigned to receive concrete that contains fibers, and the other 30 driveways will receive concrete that does not contain fibers. After one year, the developer will record the severity of cracks in each driveway on a scale of 0 to 10, with 0 representing not cracked at all and 10 representing severely cracked.

(c) Suppose the developer finds that there is a statistically significant reduction in the mean severity of cracks in driveways using the concrete that contains fibers compared to the driveways using concrete that does not contain fibers.

Question 2

In terms of the developer's conclusion, what is the benefit of randomly assigning the driveways to either the concrete that contains fibers or the concrete that does not contain fibers?

Solution 2

(a)

Mark scheme

- Experimental units: the 60 driveways (or 'new homes needing driveways' - the exact count of 60 is not required). Treatments: concrete that contains fibers and concrete that does not contain fibers. Response variable: the rating of the severity of the cracks after one year (reasonable synonyms such as 'rating,' 'severity,' or 'on a scale of 0 to 10' are acceptable; identifying the response variable as a mean/average does NOT satisfy this component). Full credit needs all 3 components identified correctly; partial credit for only 2 of the 3.

Solution 2

(b) Mark scheme

- Number the 60 driveways from 01 to 60. Using a random number generator, generate two-digit integers between 01 and 60, ignoring 00 and any repeat or number greater than 60, until 30 unique numbers are obtained. Assign the driveways with those 30 unique numbers to receive concrete with fibers, and the remaining 30 driveways to receive concrete without fibers. Full credit needs both: (1) a random process (RNG, slips of paper drawn without replacement, marbles drawn without replacement, or a spinner ignoring repeats) that labels all 60 driveways and makes every possible random assignment equally likely - must be clearly WITHOUT replacement / without repeats, (2) the process results in exactly 30 driveways assigned to each treatment. Partial credit if the response

Solution 2

implements a valid random process that satisfies only one of the two components (e.g. samples WITH replacement, which satisfies component 2 - an equal 30/30 split - but not component 1 since it does not guarantee every assignment equally likely). Component 1 is NOT satisfied if there is no labeling system, if it is unclear that every random assignment is equally likely, or if repeats are not explicitly excluded. If a response only assigns driveways to groups A/B without stating which group gets fibers, component 1 is not satisfied.

Solution 2

(c) Mark scheme

- Because the driveways were randomly assigned to either the concrete with fibers or the concrete without fibers, and the reduction in mean crack severity was statistically significant, there is evidence that the treatment (type of concrete) caused the response (rating of severity of cracks) - random assignment balances out confounding variables between the two groups so any systematic difference in outcome can be attributed to the treatment. Full credit needs: (1) states random assignment enables the causal conclusion that type of concrete caused the crack severity rating, (2) the explanation is in context (references the treatment - fibers/no fibers - and/or the severity rating). If the response also states an incorrect conclusion based on random assignment - such as generalization to a

Solution 2

population, representativeness, or reduction of bias/variability (those are benefits of random SAMPLING, not random ASSIGNMENT) - the score is reduced (E to P, or P to I).

Question 2

2022 ▪ Q2

A dermatologist will conduct an experiment to investigate the effectiveness of a new drug to treat acne. The dermatologist has recruited 36 pairs of identical twins. Each person in the experiment has acne and each person in the experiment will receive either the new drug or the placebo. After each person in the experiment uses either the new drug or the placebo for 2 weeks, the dermatologist will evaluate the improvement in acne severity for each person on a scale from 0 (no improvement) to 100 (complete cure).

(a) Identify the treatments, experimental units, and response variable of the experiment.

- Treatments:
- Experimental units:

Question 2

- Response variable:

Question 2

2022 ▪ Q2

A dermatologist will conduct an experiment to investigate the effectiveness of a new drug to treat acne. The dermatologist has recruited 36 pairs of identical twins. Each person in the experiment has acne and each person in the experiment will receive either the new drug or the placebo. After each person in the experiment uses either the new drug or the placebo for 2 weeks, the dermatologist will evaluate the improvement in acne severity for each person on a scale from 0 (no improvement) to 100 (complete cure).

(b) Each twin in the experiment has a severity of acne similar to that of the other twin. However, the severity of acne differs from one twin pair to another.
For the dermatologist's experiment, describe a statistical advantage of using a matched-pairs design where twins are paired rather than using a completely randomized design.

Question 2

2022 ▪ Q2

A dermatologist will conduct an experiment to investigate the effectiveness of a new drug to treat acne. The dermatologist has recruited 36 pairs of identical twins. Each person in the experiment has acne and each person in the experiment will receive either the new drug or the placebo. After each person in the experiment uses either the new drug or the placebo for 2 weeks, the dermatologist will evaluate the improvement in acne severity for each person on a scale from 0 (no improvement) to 100 (complete cure).

(c) For the dermatologist's experiment, describe how the treatments can be randomly assigned to people using a matched-pairs design in which twins are paired.

Solution 2

(a) Mark scheme

- Treatments: new drug, placebo. Experimental units: the 72 people who receive the new drug or the placebo (individual people/participants/subjects/twins – NOT the 36 pairs of twins). Response variable: improvement in acne severity. Full credit needs all three components: (1) identifies both treatments (new drug AND placebo), (2) identifies the experimental units as the 72 individual people (a response implying 36 experimental units, e.g. 'the pairs of twins', does not satisfy this), (3) identifies the response variable with the context of 'acne' and 'improvement' (e.g. 'improvement in acne severity', 'reduction', 'change', or the verbal scale description) – a binary description such as 'whether or not the acne improves' does not satisfy this.

Solution 2

(b) Mark scheme

- Improvement scores vary due to many factors, including initial acne severity, which treatment is received, and other variables such as diet and genetics. Because each pair of twins has similar initial acne severity, pairing allows the variation in improvement scores due to the treatment received to be distinguished from variation due to initial acne severity, unlike in a completely randomized design. Consequently, the matched-pairs design gives a more precise estimate of the mean difference in improvement between the new drug and the placebo, and makes it easier to find convincing evidence that the new drug is better, if it really is. Full credit needs a description of a genuine statistical advantage of matched pairs (not merely describing the pairing set-up) that satisfies all three: (1) the advantage

Solution 2

pertains to an inference made after data collection (e.g. distinguishing treatment effects, or precision of the estimate), (2) uses a comparative word (e.g. 'easier', 'clearer', 'greater') or an explicit comparison to a completely randomized design, (3) includes context (e.g. 'drug', 'improvement', 'acne', or 'twins'). Note: 'reduces bias' and 'reduces variability in the estimates of the individual treatment means' do NOT satisfy component 1, since neither is true of a matched-pairs design specifically.

Solution 2

(c) Mark scheme

- For each pair of twins, label one twin A and the other twin B. Toss a coin for each pair: if heads, twin A gets the placebo and twin B gets the active drug; if tails, twin A gets the active drug and twin B gets the placebo. (Equally acceptable: use a random-number generator to select an integer 1 or 2 per pair to assign the drug, or shuffle two labeled notecards 'A'/'B' per pair and give one to each twin, with a specified card meaning 'drug'.) Repeat for all pairs of twins. Full credit needs a response that randomly assigns the two treatments within pairs AND satisfies all three: (1) uses a random process (coin flip, random-number generator, shuffled cards, etc.) giving each twin in a pair a 50% chance of the drug and 50% chance of the placebo, (2) describes how the random outcome assigns one specific twin in

Solution 2

each pair to the drug and the other to the placebo (must cover every possible outcome of the random process, not just one branch), (3) indicates the random assignment process is completed for every pair of twins.

Question 2

2019 ▪ Q2

Researchers are investigating the effectiveness of using a fungus to control the spread of an insect that destroys trees. The researchers will create four different concentrations of fungus mixtures: 0 milliliters per liter (ml/L), 1.25 ml/L, 2.5 ml/L, and 3.75 ml/L. An equal number of the insects will be placed into 20 individual containers. The group of insects in each container will be sprayed with one of the four mixtures, and the researchers will record the number of insects that are still alive in each container one week after spraying.

(a) Identify the treatments, experimental units, and response variable of the experiment.

Treatments:

Experimental units:

Response variable:

Question 2

(b) Does the experiment have a control group? Explain your answer.
(c) Describe how the treatments can be randomly assigned to the experimental units so that each treatment has the same number of units.

Solution 2

(a)

Mark scheme

- Treatments: the four fungus spray concentrations (0 mL/L, 1.25 mL/L, 2.5 mL/L, and 3.75 mL/L) – must refer to the plural concentrations/mixtures/sprays, not just 'concentration' as a variable. Experimental units: the 20 containers, each containing the same number of insects. Response variable: the number of insects that are still alive in each container one week after spraying (must be stated as a per-unit variable, not a single total count).

Solution 2

(b)

Mark scheme

- Yes, the experiment has a control group. Because the 0 mL/L concentration contains no fungus, the containers that are sprayed with the 0 mL/L concentration form the control group. (Identifying the control group, or explaining there is a treatment containing no fungus, is sufficient justification; does not need to explicitly say 'yes'.)

Solution 2

(c) Mark scheme

- Label each container with a unique integer from 1 to 20. Use a random number generator to choose 15 integers from 1 to 20 without replacement: use the first five numbers to identify the five containers that receive the 0 mL/L treatment, the second five for the 1.25 mL/L treatment, the third five for the 2.5 mL/L treatment, and the remaining five containers receive the 3.75 mL/L treatment. (Alternative: label 20 equally-sized slips of paper, five each with 0, 1.25, 2.5, and 3.75 mL/L; mix them in a hat; for each container, draw a slip at random without replacement and spray that container with the treatment drawn.) Full credit requires: appropriate labels for units/treatments, a correctly described

Solution 2

random-assignment process, and the process resulting in an equal number of units (5) per treatment.

Question 6

2017 ▪ Q6

Consider an experiment in which two men and two women will be randomly assigned to either a treatment group or a control group in such a way that each group has two people. The people are identified as Man 1, Man 2, Woman 1, and Woman 2. The six possible arrangements are shown below.

Arrangement A	Treatment	Control
	Man 1	Woman 1
	Man 2	Woman 2

Question 6

Arrangement B	Treatment	Control
	Man 1 Woman 1	Man 2 Woman 2

Arrangement C	Treatment	Control
	Man 1 Woman 2	Man 2 Woman 1

Question 6

Arrangement D	Treatment	Control
	Woman 1 Woman 2	Man 1 Man 2

Arrangement E	Treatment	Control
	Man 2 Woman 2	Man 1 Woman 1

Question 6

Arrangement F	Treatment	Control
	Man 2	Man 1
	Woman 1	Woman 2

Two possible methods of assignment are being considered: the sequential coin flip method, as described in part (a), and the chip method, as described in part (b). For each method, the order of the assignment will be Man 1, Man 2, Woman 1, Woman 2.

Question 6

(a)(i) For the sequential coin flip method, a fair coin is flipped until one group has two people. An outcome of tails assigns the person to the treatment group, and an outcome of heads assigns the person to the control group. As soon as one group has two people, the remaining people are automatically assigned to the other group.
Complete the table below by calculating the probability of each arrangement occurring if the sequential coin flip method is used.

Arrangement	**A**	**B**	**C**	**D**	**E**	**F**
Probability						

Question 6

(a)(ii) For the sequential coin flip method, what is the probability that Man 1 and Man 2 are assigned to the same group?

Question 6

Arrangement A	
Treatment	Control
Man 1	Woman 1
Man 2	Woman 2

Arrangement B	
Treatment	Control
Man 1	Man 2
Woman 1	Woman 2

Arrangement C	
Treatment	Control
Man 1	Man 2
Woman 2	Woman 1

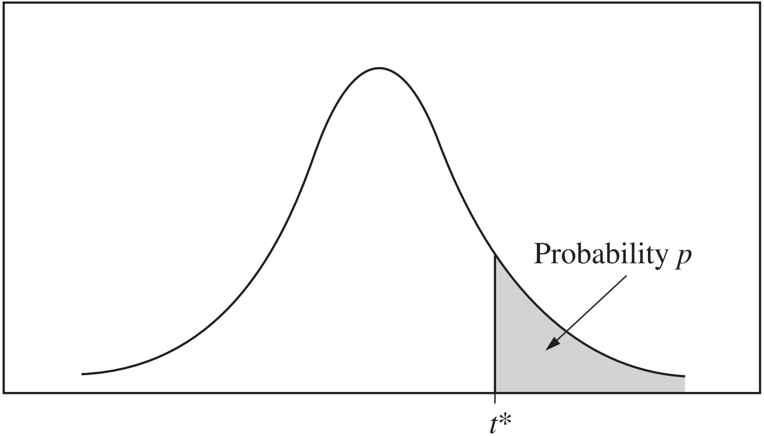
Arrangement D	
Treatment	Control
Woman 1	Man 1
Woman 2	Man 2

Arrangement E	
Treatment	Control
Man 2	Man 1
Woman 2	Woman 1

Arrangement F	
Treatment	Control
Man 2	Man 1
Woman 1	Woman 2

Question 6

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^*.



Question 6

Table entry for z is the probability lying below z.