Assessment Schedule 2024

Mathematics and Statistics (Statistics): Evaluate statistically based reports (91584)

Evidence Statement

Q	Expected Coverage	Achievement (c)	Achievement with Merit (j)	Achievement with Excellence (i)
ONE (a)	Responders could have been given a diagram of the body and the names of the body parts and asked to place the body parts in the correct place. OR Responders could have been asked "For each of the body parts below, can you confidently state where in the body it is located?", and then provided with a yes / no option for each body part. *Other responses may be valid.*	• A sensible suggestion given, with enough detail so the method could be enacted.		
(b)	$n = 2000$ $\text{MoE} = \dfrac{1}{\sqrt{2000}} = 0.022 = 2.2\%$ CI for proportion: $55\% \pm 2.2\% \rightarrow [52.8\%, 57.2\%]$ We can be fairly sure that the proportion of all UK men and women who can confidently state where the rectum is located is somewhere between 52.8% and 57.2%. *Note: Must be some form of doubt – "95% confident", "Fairly sure" or similar*	• Confidence interval **correctly** calculated.	• Confidence interval **correctly** calculated. AND Interpretation in **context** given including population of "UK men and women" or "UK adults".	
(c)	Pall Mall Medical, a private healthcare clinic, funded the study. They have a commercial interest in the healthcare of people in the UK and are looking to entice customers to go to their clinic. The report included a quote from Dr Tang linking good knowledge of body parts to people being more in control of their health, thus making the potential link back to their clinic. OR Since Pall Mall Medical may benefit from highlighting a lack of public knowledge about anatomy, this could create a bias. The clinic might use the results to promote its services, emphasizing the need for professional health guidance and possibly driving more people to seek medical consultations or health education from private providers like themselves. Thus, the source of	• Identifies that the company that funded the survey is a private healthcare clinic. AND Describes one way the survey could be used to the company's advantage.	Explains why the findings of the survey could be used to the clinic's advantage with reference to the survey results reported AND Describes how or why this issue could impact the study. (e.g., exaggerated reports)	

	funding may have influenced the study's focus (and perhaps exaggerated information) and presentation to support Pall Mall Medical's interests. *NOTE: accept counter-discussion that there may not be an issue with who funded this study if clearly articulated with WHAT and WHY not.* *E.g. Pall Mall has a vested interest in promoting the view that people are not competent in their understanding of their bodies, and so must rely on medical professionals.*			
(d)	Comparison within a group Poll percentage difference 55% – 52% = 3% MoE is $2 \times \dfrac{1}{\sqrt{2000}} = 0.045 = 4.5\%$ CI for difference: 3% ± 4.5% [–1.5%, 7.5%] *Accept 3% ±4.4 % if using rounded MOE from 1(b).* We can be fairly sure that the proportion of UK men and women who can confidently locate the rectum is somewhere between 1.5% less and 7.5% more than the proportion of UK men and women who can confidently locate the reproductive organs. As this confidence interval is both negative and positive, there is not sufficient evidence to support the claim that *a higher proportion of UK men and women can confidently locate their rectum that their reproductive organs.*	• Confidence interval correctly calculated. *If candidates have incorrect CI, but correct interpretation and claim as per incorrect CI, then c.*	• Confidence interval correctly calculated. AND EITHER CI interpreted correctly in context, including population of "UK men and women" or "UK adults". OR Claim evaluated in context with justification. *Accept other expressions of some uncertainty with the confidence interval, e.g. it's a fairly safe bet, or with 95% confidence.*	• Confidence interval correctly calculated. AND CI interpreted correctly in context, including population of of "UK men and women" or "UK adults". AND Claim evaluated in context with justification.

| (e) | The target population is all UK men and women.

The sampling frame is the OnePoll Panel and these individuals are not randomly selected from all people in the UK.

Alternative answer 1:

The target population and the sampling frame may not match, e.g. the people doing OnePoll panels might be lower socio-economic, and not earning as much income.

This could affect the results, as people not earning as much income could have less education, meaning that they are less able to identify the location of organs.

Alternative answer 2:

People who don't join the panel might not like using the computer, and these people might be less educated and know less about technical language for body parts than the people who do the survey, so these percentages could be an <u>overestimate</u> for the true population percentage.

Alternative answer 3:

One potential non-sampling error is response bias, which occurs if respondents do not provide truthful answers or if their answers are influenced by the survey's online format, or if they are offered incentives like being paid. For example, since they are being paid to participate, some panellists may rush through questions or select answers without fully considering them, which could affect the accuracy of the data. Additionally, the knowledge gap measured in an online survey may not reflect the knowledge of those uncomfortable with or unrepresented in online platforms. | • Correctly identifies:
Target population.
OR
Sampling frame. | • Achievement
AND
Explanation regarding the likely representativeness of the sample. | • Merit
AND
Discussion of non-random selection of the sample with specific evidence of a non-sampling error.
AND
Describes how / why this issue affects the finding of the study. |

NØ	N1	N2	A3	A4	M5	M6	E7	E8
No response; no relevant evidence.	Attempts at least one part of the question.	1 of c	2 of c OR 1 of j	3 of c	1 of j AND 1 of c	2 of j	1 of i	2 of i

Q	Expected Coverage	Achievement (c)	Achievement with Merit (j)	Achievement with Excellence (i)
TWO (a)	The researchers are interested in whether the socks help the descent of participants down the slippery slope, therefore the most relevant variable to measure this outcome is the participants' perception of how slippery the slope is. **Alternative answer 1:** Participants could be expected to be biased because the treatments were not blinded (placebo effect). Consequently, there could be less bias in the observer rating of slipperiness. The researchers may have chosen self-reported slipperiness as the primary outcome because that was the measure that had the significant difference. **Alternative answer 2:** Participants' self-perception of slipperiness directly reflects how safe or confident they feel, which is crucial in real-life situations where perceived safety influences behaviour. **Alternative answer 3:** Time taken to descend may not always correlate with slipperiness, as participants could walk cautiously regardless of grip. Self-rated slipperiness is more likely to capture subtle changes in traction. **Alternative answer 4:** The time to descend could be influenced by other factors like walking speed or risk-taking behaviour. For example, a participant might descend more slowly due to caution rather than experiencing slipperiness. Self-rated slipperiness focuses purely on how slippery the conditions feel.	• Identifies self-rated slipperiness as the study goal.	• Discussion linking self-rated slipperiness with study goal.	
(b)	Explanatory variable: footwear – socks over the top of footwear or no socks. Response variables: Primary outcome – self-reported slipperiness on a five-point scale. Secondary outcomes – falls, observer-rated slipperiness, and time taken to descend the study slope. *Only one of the possible response variables required*	• Explanatory variable described in context. AND Response variables are described in context. *NOTE: Primary outcome should be identified – do NOT accept "difference in mean" or "mean", must have "<u>self-reported</u>" or equivalent, "primary outcome" not needed.*		

(c)(i)	Blinding was not possible in this study, as the treatment was putting socks over the participants' footwear. As soon as the participant was asked to put the socks on, they and the outcome assessor would know if they were in the socks-treatment group.	• Clearly explains why blinding is not possible in **this** study.	• Clearly explains why blinding is not possible in this study. AND Identifies ONE measure and how it may have helped minimise the impact of not blinding.	• Clearly explains why blinding is not possible in this study. AND Identifies TWO measures and how each may have helped minimise the impact of not blinding.
(ii)	**Alternative answer 1:** All recruiters and outcome assessors were told to wear unmodified footwear. The intent of this was to ensure no implication that socks were superior. **Alternative answer 2:** Participants and outcome assessors were simply told that the performance of different types of footwear and different types of socks being worn over the top were being assessed. This would have meant that participants and outcome assessors thought there were many different categories of treatment and were not aware of which specific group they were allocated into. **Alternative answer 3:** Participants' footwear was photographed for later analysis. Photographing the participants makes them think the experiment is about differences in footwear, or some other explanatory variable other than socks on footwear. **Alternative answer 4:** Participants and assessors were not informed that the study aimed to evaluate the effectiveness of socks over shoes. By withholding the true purpose of the study, participants and assessors were less likely to consciously or unconsciously alter their behaviour or ratings to align with expected results. This reduces bias. *Accept any measure identified in the report (all wearing unmodified footwear, told looking at different types of footwear / socks).*			

| (d) | A potential confounding variable could be how often the participants have walked this route. 14% of the intervention group had been walking this route for more than 12 months, compared to 27% of the control group. Participants who have been walking the route more often may have more confidence with dealing with the icy conditions, therefore reporting a <u>lower</u> slipperiness rating or meaning that their descent time was faster.

A potential confounding variable could be the number of previous falls on the ice. 57% of the intervention group had previous falls, compared to 73% of the control group. Participants who have had more falls may be more cautious walking the route, therefore reporting a <u>higher</u> slipperiness rating or meaning that their descent time was slower.

A potential confounding variable could be gender. Footwear for women is often designed differently and may differ in slipperiness. Also, women often have smaller feet and this may affect how they slip on ice, therefore reporting a <u>higher</u> slipperiness rating or meaning that their descent time was <u>slower</u>. 50% of the intervention group were women, compared to 33% of the control group.

Accept any confounding variable discussed in the report where there is difference between the two groups (gender, previous falls, injury from the fall this winter, experience in walking the route, age of the participant).

These variables are not well controlled in the study design (with fairly different percentages between the intervention group and the control group). | • Describes ONE potential confounding variable.

AND

Discusses how the confounding variable may interact with the response variable(s). | Describes TWO potential confounding variables

AND

Discusses how each confounding variable may interact with the response variable(s). | • Justifies choice of TWO potential confounding variables.

AND

Discusses how each confounding variable may interact with the response variable(s).

AND

Candidates should comment on how the variable is not completely controlled by the study design as the group percentages are unequal for that variable for at least one of the variables. |

NØ	N1	N2	A3	A4	M5	M6	E7	E8
No response; no relevant evidence.	Attempts at least one part of the question.	1 of c	2 of c OR 1 of j	3 of c	1 of j AND 1 of c	2 of j	1 of i	2 of i

Q	Expected Coverage	Achievement (c)	Achievement with Merit (j)	Achievement with Excellence (i)
THREE (a)	The "very confident" (20%) and "somewhat confident" (26%) categories for males have been added to give a total of 46% of men who are "confident" they can land a plane. The headline would have rounded this figure to 50%.	• Identifies evidence and explains how 50% is reached.		
(b)	The report states that 13% of all respondents are very confident that they could safely land a passenger airplane in an emergency situation, relying only on the assistance of air traffic control. This is below 30%, so outside the 30–70% guide. The rule of thumb MoE will overestimate the size of the MoE, and hence using it to construct an approximate 95% confidence interval for the population proportion of Americans who are very confident they could safely land a passenger airplane in an emergency situation, relying only on the assistance of air traffic control, would result in a confidence interval that is wider than it needs to be. *NOTE: If underestimates the MOE mentioned max grade is c*	• Identifies one survey percentage in context and explains that it is outside the 30–70% range.	• Identifies one survey percentage in context that is outside the 30–70% range. AND Explains that the rule of thumb MOE will **overestimate** the size of the MOE (accept will result in a **wider** confidence interval).	
(c)	**Resource Three (a) – YouGov study** Study Design: • This is an observational study (survey) because the researchers just asked people's opinion, they didn't **do** anything to the participants. Inference: • A sample-to-population inference can be made. (accept a causal claim cannot be made) Assumptions: (one of the following) • Assumption that a random sample was taken. • Assumes participants provide truthful and accurate self-assessments of their confidence. • Assumes all respondents interpret the question similarly (e.g., what it means to "land a plane"). • Assumes the sample is representative of the general population, allowing for generalisation of results **Resource Three (b) – University of Waikato study** Study Design: • This is an experimental study (randomised controlled trial) because the researchers controlled whether the participants watched a video of	• For EITHER Resource Three (a) OR Report Three (b), the following are given: - Type of study design identified **with justification**. - Type of inference identified.	• For EITHER Resource Three (a) OR Report Three (b), the following are given: - Type of study design identified **with justification.** - Type of inference identified. - Assumption correctly identified.	• For BOTH Report Three (a) AND Report Three (b), the following are given: - Type of study design identified with justification. - Type of inference identified. - Assumption correctly identified.

pilots landing a plane or not. Inference: • A causal inference (causal claim) can be made Assumptions: (one of the following) • Assumption the participants were randomly allocated to the two treatment groups. • Assumes the video alone influenced confidence without additional factors, (not due to chance alone) despite being described as "100 percent useless" by an expert.			

			• One feature in context, with evidence.	• TWO features in context, with evidence for each, **one for each** graph.
(d)	**Possible Comments:** **Comment 1:** Participants who watched the video were more confident they could land the plane without dying than those who had not. The median confidence for the participants who watched the video was higher (around 40%) than for the group who didn't watch the video (around 30%). **Comment 2:** Participants who watched the video were more confident they could land the plane as well as a pilot could than those who had not. The confidence rating (of being able to land a plane) for the video group has a narrower bulge at around 10% confidence rating, with the rest of the ratings tapering up to about 70% confidence. In contrast, the no video group showed a much wider bulge at a lower confidence level (around 5%), tapering to a narrow neck at about 50% confidence. **Comment 3:** In the "without dying" scenario, the median confidence level in the video group is significantly higher (around 75%) compared to the no-video group (around 25%). This demonstrates that participants who watched the video were substantially more confident in their ability to land a plane without dying. **Comment 4:** In the "as well as a pilot could" scenario, the no-video group exhibits very low confidence levels, with most responses near the bottom of the scale (close to 0% - twice as wide as the video group). This suggests that without the video, participants overwhelmingly recognised the improbability of performing at a pilot's level. **Comment 5:** Even in the highly unrealistic "as well as a pilot could" scenario, the video group shows a higher median confidence level (around 30%) and a broader spread of responses compared to the no-video group. This suggests that watching the video inflated confidence even in a scenario requiring expert-level performance. ***Accept at most ONE comment comparing group medians.*** *Accept any correct comparative comments relating to the given conclusion that can be made from the information in Figure 1.*				

| (e) | The researchers randomised the order of the questions to remove 'order effect'.

 Order effect occurs when participants' responses are affected by the order of the questions.

 If they select a higher response for one question, there is likely to be a lack of independence between their responses, and they are likely to select a higher response for the second question.

 No requirement to use term "order effects" but response must be in context of the question. | | • Discussion of order effects with **links to context**. | |

NØ	N1	N2	A3	A4	M5	M6	E7	E8
No response; no relevant evidence.	Attempts at least one part of the question.	1 of c	2 of c OR 1 of j	3 of c	1 of j AND 1 of c	2 of j	1 of i	2 of i

Cut Scores

Not Achieved	Achievement	Achievement with Merit	Achievement with Excellence
0 – 7	8 – 14	15 – 18	19 – 24