

Assessment Schedule 2025

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)(i)	$\frac{1}{\sqrt{n}} = \frac{1}{2000} = 0.0224 = 2.2\%$	<ul style="list-style-type: none"> • Correct calculation of MOE. AND	Correct MOE. AND	
(ii)	The margin of error is needed to account for the natural variation that always occurs from sample to sample. For the survey, every time a sample of Americans is surveyed and asked questions about the way they interact with AI, we would expect slightly different results.	Describes why the MOE is required in the survey.	Correct description of why MOE is required in context.	
(b)	$44\% \pm 2.2\%$ $[41.8\%, 46.2\%]$ We are pretty sure that the proportion of all Americans who are polite when communicating with AI is somewhere between 41.8% and 46.2%. <i>Allow minor errors in calculation if working evident, and answer is reasonable.</i>	<ul style="list-style-type: none"> • CI correctly calculated. OR Correct inference statement in context (with population AND proportion of Americans who are polite) from incorrect CI.	<ul style="list-style-type: none"> • CI correctly calculated. AND Used to write inference statement in context (with population AND variable).	

(c)		Americans	New Zealanders	<ul style="list-style-type: none"> • Average MOE calculated (or $1.5 \times$ average MOE), i.e. 2.5% or 3.8% with clear evidence of method. OR Correct confidence interval without clear working. OR Correctly calculated an incorrect CI for difference between two groups (i.e. using wrong MOE method and / or wrong sample sizes) and both claim and interpretation are in context. 	<ul style="list-style-type: none"> • Confidence interval correct. AND EITHER Used to write inference statement in context (with populations AND variable). OR Response to claim made in context (with populations AND variable). 	<ul style="list-style-type: none"> • Confidence interval correct. AND Used to write inference statement in context. AND Response to claim made in context. With populations AND variable evident in response when read as a whole.
	Poll % polite to AI	44%	$\frac{624}{1250} = 49.92\%$			
	n	2000	1250			
	$\frac{1}{\sqrt{n}}$	$0.022361 = 2.2\%$	$0.028284 = 2.8\%$			
	Average MOE	$0.025322 = 2.5\%$				
<p>Difference between groups: $0.4992 - 0.44 = 0.0592 = 5.9\%$ 1.5 times average MOE = $0.037984 = 3.8\%$ CI for difference: $5.9\% \pm 3.8\%$ [0.021216, 0.097184] [2.1%, 9.7%]</p> <p>We're pretty sure that the proportion of New Zealanders who are polite when communicating with AI is somewhere between 2.1% and 9.7% higher than that of Americans.</p> <p>Because the confidence interval is completely positive, we have sufficient evidence to support the claim that New Zealanders are more polite when communicating with AI than Americans.</p> <p><i>Allow minor errors in calculation if working evident (e.g. rounding), and answer is reasonable.</i></p>						

<p>(d)</p>	<p>SPEEDERS: These may have been excluded as they did not take the time to read the questions carefully, or think specifically about how they communicate with AI; they may have just randomly clicked on the first option in the multichoice selection. This may have influenced the results of the survey by, for example, showing people are more polite when communicating with AI than we would actually expect in the population.</p> <p>OPEN ENDS: If respondents wrote inappropriate things in the text boxes, this may have been taken as an indication that they didn't take the survey were not considering their responses carefully. These results may have impacted the survey as irrelevant information would have been included in the statistical analysis, diluting the percentage of actual opinions about communicating with AI.</p> <p>BOTS: The survey is asking about how humans communicate with AI. If bots were not excluded, the results may have been more positive than those reported, with bots trying to influence humans to be more polite when interacting with them. For example, bots might heavily favour one option or randomly choose answers, artificially altering the reported percentages (such as the 44% who say they are polite). This would make the results less trustworthy and not representative of the population.</p> <p>DUPLICATES: It is important that respondents only answer the survey once, as otherwise their views will be influencing the survey results more than others. For example, if a respondent with extreme views answered the survey multiple times, then the results reported may present a negative picture about AI communication when this isn't actually the case.</p> <p><i>Accept other sensible and statistically valid responses that are related to the context, and one of the four situations.</i></p>		<ul style="list-style-type: none"> Explains WHY the situation was used, in context of the survey. <p>OR</p> <p>Explains HOW the survey results may have been impacted.</p>	<ul style="list-style-type: none"> Explains WHY the situation was used, in context of the survey. <p>AND</p> <p>Explains HOW the survey results may have been impacted.</p>
------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

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

<p>(c)</p>	<p>An issue linked to the specific claim is described, and how this may impact extending the results is discussed in context.</p> <p>For example:</p> <ul style="list-style-type: none"> • The experiment was conducted on people recruited through Amazon Mechanical Turk. We have no idea if these participants are representative of any broader population. The participants in the experiment may have more knowledge of art and paintings, which was why they volunteered for the study, than the general population of New Zealand say, and therefore may have been more negatively influenced in their ratings when they knew a painting was AI-generated. • The claim generalises to all “AI-created art and “human art”, yet the experiment only used AI-created representational and abstract paintings. There are many other types of art, for example, printmaking or photography, where AI may be generating pieces. People’s biases against AI in the different art genres may be different than with these limited painting selections. • The study recruited participants from Amazon Mechanical Turk, a platform whose users tend to be younger, digitally literate, and accustomed to online artificial intelligence tools. These characteristics may make them more aware of—or more sceptical about—AI-generated art. Because this study only tested this specific group, their ratings may not reflect the views of wider populations, limiting the validity of extending the claim that “people are biased against AI-created art”. <p><i>Accept other statistically valid discussions related to the claim. Do not accept discussions about sample size</i></p>		<ul style="list-style-type: none"> • Describes ONE potential issue related with extending the results, in relation to the claim. 	<ul style="list-style-type: none"> • Describes ONE potential issue related with extending the results, in relation to the claim. <p>AND</p> <p>Discusses how this issue could limit extending the results by using specific feature(s) of this study.</p>
------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	-------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

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)	If the sample was randomly selected.	<ul style="list-style-type: none"> Identifies the need for a random sample. 		
(b)	<p>A clear description is given, highlighting differences between the target population and the sampling frame. For example:</p> <ul style="list-style-type: none"> The target population is for people living in only five countries. <i>The Planted Pot</i>'s social media may attract visitors from other countries who may complete the survey. The target population is all house plant owners. People may be interested in house plants, so have signed up to <i>The Planted Pot</i>'s newsletter, yet do not own any house plants. The sampling frame only includes people who follow <i>The Planted Pot</i> online or receive their newsletter, which requires having internet access. Many house plant owners—especially older adults or people in rural areas—may not use social media or visit online plant stores. These people are part of the target population of house plant owners but are missing from the sampling frame. The sampling frame includes anyone who follows <i>The Planted Pot</i>'s social media or signs up to their newsletter, even if they do not actually own house plants. Some people might follow for gardening inspiration or home décor ideas. These individuals should not be in the target population, so the sampling frame includes people who should not be sampled Many house plant owners may buy their plants or pots from garden centres, DIY stores, supermarkets, or other websites, and may never have heard of <i>The Planted Pot</i>. These people are in the target population but have no chance of being selected because they are not part of the company's online audience. <i>The Planted Pot</i> is a New Zealand business, so its social media followers and newsletter subscribers are likely to contain a disproportionate number of New Zealand customers. This means New Zealand house plant owners are over-represented in the sampling frame relative to owners in the other four countries. 		<ul style="list-style-type: none"> ONE situation described in context, where the target population and sampling frame do not match. 	

<p>(c)</p>	<p>The report states that 25% of respondents own fewer than 10 house plants. This is below 30%, so outside the 30–70% guide.</p> <p>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 house plant owners who own less than 10 house plants, would result in a confidence interval that is wider than it needs to be.</p>	<ul style="list-style-type: none"> • Clearly identifies one survey percentage in context that is outside the 30–70% range. 	<ul style="list-style-type: none"> • Clearly identifies one survey percentage in context that is outside the 30–70% range. <p>AND</p> <p>Explains that the rule of thumb MOE will overestimate the size of the MOE (accept will result in a wider confidence interval).</p>	
<p>(d)</p>	<p>Comparison within a group Poll percentage difference $43\% - 38\% = 5\%$</p> <p>Margin of error is $2 \times \frac{1}{\sqrt{n}} = 2 \times \frac{1}{\sqrt{1236}} = 2 \times 0.028444$ $= 0.056888 = 5.7\%$</p> <p>CI for difference: $5\% \pm 5.7\%$ [–0.6888%, 10.6888%] [–0.7%, 10.7%]</p> <p>We’re pretty sure that the proportion of house plant owners in New Zealand, Canada, the USA, the UK and Australia that spend less than one hour per week taking care of their house plants is somewhere between 0.7% less and 10.7% more than the those who spend between one and three hours.</p> <p>Because the confidence interval is both negative and positive, we do not have sufficient evidence to support the claim that a <i>higher proportion of house plant owners in New Zealand, Canada, the USA, the UK and Australia spend less than one hour per week taking care of their house plants than spend between one and three hours</i>.</p> <p><i>Allow minor errors in calculation if working evident, and answer is reasonable.</i></p>	<ul style="list-style-type: none"> • MOE calculated, i.e. 5.7%, with clear evidence of method. <p>OR</p> <p>Correct confidence interval without clear working.</p> <p>OR</p> <p>Correctly calculated an incorrect CI for difference between two groups (i.e. using wrong MOE method and / or wrong sample sizes) both claim and interpretation are in context.</p>	<ul style="list-style-type: none"> • Confidence interval correct. <p>AND</p> <p>EITHER</p> <p>used to write inference statement in context (with population AND variables).</p> <p>OR</p> <p>Response to claim made in context (with population AND variables).</p>	<ul style="list-style-type: none"> • Confidence interval correct. <p>AND</p> <p>Used to write inference statement in context.</p> <p>AND</p> <p>Response to claim made in context with population AND variables evident in response when read as a whole.</p>

(e)	<p>Examples:</p> <ul style="list-style-type: none"> • The report specifically asked respondents how many house plants they purchased in the last six months, so is excluding any house plants respondents have acquired for free (e.g. from propagating or as gifts). The survey results may be an <u>underestimate</u> of the number of house plants in a home because any ‘free’ house plants will not be included. • The survey only asked how many plants people purchased and did not ask how many plants they lost (through dying, pests, being thrown out, or gifting). This means someone could buy many plants while also losing many, so their total number of plants might stay the same or even decrease, making the estimate of “growth in number of plants per home” overstated. • Many plant owners grow new plants by propagation (cuttings, splitting, etc.), but the survey counted only purchased plants, ignoring propagated ones. Homes where owners propagate plants may have rapidly increasing plant numbers even if they purchased none, so relying only on purchase data underestimates true growth and gives misleading estimates therefor underestimates of plant numbers in households. • Plant purchasing varies seasonally — people buy more in spring/summer than autumn/winter — but the survey only looked at the most recent six months. If the survey happened during peak buying months, plant purchase rates would appear inflated, leading to an overestimation of annual growth and the number of plants typically found in a home. 		<ul style="list-style-type: none"> • Explains WHY asking about purchasing habits is an issue, in context of the survey. <p>OR</p> <p>Explains HOW the survey results may have been impacted.</p>	<ul style="list-style-type: none"> • Explains WHY the situation was used, in context of the survey. <p>AND</p> <p>Explains HOW the survey results may have been impacted.</p>
-----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

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 – 19	20 – 24