

MATHS AND STATISTICS PRE-ARRIVAL MODULES PRACTICE PROBLEMS: ANSWER KEY

MODULE 1

1. *What is the difference between a theory and a hypothesis?*

A hypothesis is a tentative assumption created before the study or data collection begins, which is tested during the course of the research.

A theory is supported by preexisting studies and evidence, and helps to explain findings already seen in the data. As a result, a theory is a much stronger claim than a hypothesis.

2. *Imagine you are a policymaker, and you have been given additional funding to improve education outcomes in your district. You could spend these funds in many different ways, but the head of your department suggests that hiring more teachers to reduce the size of the class will improve test scores.*

Using this idea, try to develop a research question, theory, and set of hypotheses.

Here is one way to complete this table:

RESEARCH QUESTION	What is the relationship between small class size and education outcomes such as test scores?
THEORY	Smaller classes means more attention is given to each student, which will improve their knowledge and academic performance.
HYPOTHESIS (H_A)	Children in smaller classes will have higher test scores.
NULL NYPOTHESIS (H_0)	There is no effect of class size on test scores.

3. *What is the difference between qualitative and quantitative data?*

Qualitative data is descriptive and non numerical, features a small and unrepresentative sample, and cannot be used with statistical analysis. Qualitative data aims for in depth understanding of the context or subjects in a study.

Quantitative data is numerical, features a large and representative sample, and can be used in statistical analysis. Quantitative data aims to test the hypotheses of a study, in order measure or quantify the causal effect in the research study.

4. *Given this list of data, identify which are “qualitative” and which are “quantitative.”*

- a. Number or measurements - QUANTITATIVE
- b. Open ended interviews - QUALITATIVE
- c. First hand accounts (such as diaries, or other primary sources) - QUALITATIVE
- d. Survey data - QUANTITATIVE
- e. Newspaper reports - QUALITATIVE

Note: interviews are typically considered to be qualitative because they are unstructured and typically feature a small sample, but a large scale survey of respondents (a survey is a type of interview) that asks open ended questions could be coded in a numerically meaningful way. It's also worth noting that, depending on the context, some qualitative data can in fact be quantified (or coded to be numerically meaningful).

MODULE 2

1. *What do we mean when we say “correlation does not imply causation”?*

This means that just because two events occur together, or correlate, does not mean that one has a causal effect on the other.

2. *Suppose someone supplied you with data that showed a positive correlation between the number of nurses in a hospital (X) and the number of patient deaths (Y). They then tried to argue that increasing the number of nurses would cause an increase in deaths, and therefore nurses are bad for patient health. Using the concept of omitted variables, or any other logical reasoning, and explain why this conclusion might be wrong.*

It's unlikely that more staff leads to worse outcomes for patients in a hospital, so there is most likely an omitted variable.

One such omitted variable might be the type of area in which the hospital is located — for example, maybe this area has an abnormally high rate of accidents. If this is systematically true, it also means that the hospital increased its staff in expectation of higher numbers of patients (and, the higher number of severe accidents, naturally the mortality rates would increase as well).

Another omitted variable might have to do with the type of hospital. Perhaps the hospital is responsible for caring for all the severely ill patients in the region, and these are the types of patients with high mortality rates.

Either omitted variable could both cause i) the number of nursing staff to increase, and ii) higher number of patient deaths.

For a real world example of this, see <http://www.manchester.ac.uk/discover/news/national-study-casts-doubt-on-higher-weekend-death-rate-and-proposals-for-seven-day-hospital-services>.

3. *Imagine you are a policymaker, and you are focusing on issues with low turnout in your district (as in, the number of people voting in each election has been steadily decreasing). Here, turnout is your dependent variable (Y). List as many independent variables (X) as you can that could plausibly affect turnout on the day of the election.*

Y= Voter turnout

X=

1. Adverse weather conditions
2. Dissatisfaction with candidates or parties
3. Uncompetitive elections
4. Illegal voter intimidation
5. Polling place too far away
6. Age or income of the voting population
7. And there are many more....

MODULE 3

1. Using Tables 3.1 and 3.2, write out this expression in words:

$$\forall x \in X \text{ and } \forall y \in Y, \exists z \in Z \text{ s.t. } x + y = z.$$

For every x in the set X and every y in the set Y , there is an element z in the set Z such that x plus y equals z .

2. Evaluate: $2x^2 + 3x$ for $x = 3$

$$2(3^2) + 3(3) = 0$$

$$2(9) + 9 = 0$$

$$18 + 9 = 0$$

$$27$$

3. Solve for x :

$$(x+4) - (x+2) = 6x$$

$$(x+4) - (x+2) = 6x$$

$$x+4 - x - 2 = 6x$$

$$x + 4 - x - 2 - 6x = 0$$

$$x - x - 6x = -4 + 2$$

$$-6x = -2$$

$$x = -2/6$$

$$x = 1/3$$

Carry the negative sign through

Carry the negative sign through

Subtract $6x$ from both sides

Keep "x" terms to the left, subtract 4 and add 2 to both sides

Combine like terms

Simplify

4. *Of all the countries in the world, the country of Rwanda currently has the highest proportion of female politicians serving in the national legislature.*

Term	Number of Female Legislators	Total Number of Legislators
2013	49	80
2008	45	80

Using this information, answer the following questions (you may round to one decimal point):

a. *What percentage of legislators were women in 2008?*

56.3%

b. *What percentage of legislators were women in 2013?*

61.3%

c. *The number of female legislators change by how many percentage points from 2008 to 2013?*

5 percentage points

d. *What was the percentage increase in female legislators from 2008 to 2013?*

8.9%

MODULE 4

1. Levels of measurement are usually placed in a sequence, from weakest measurement to strongest measurement (or said in another way, from the least informative to the most informative).

Place these in order, from the least informative to most informative: ratio, nominal, interval, ordinal.

Answer:

Nominal

Ordinal

Interval

Ratio

2. For each variable, provide the type of measurement:

a. The number of citizens in a town

RATIO

b. A set of categories measuring the respondent's country (1=UK, 2=Singapore, 3=India, 4=Mexico)

NOMINAL

c. A "feeling thermometer", or a survey question that asks where a respondent likes a policy idea (with 1 being don't like it at all, and 10 being like it very much)

ORDINAL

3. What does the abbreviation SD stand for?

Standard Deviation

4. *For each of these terms, list whether they are a measure of central tendency or dispersion.*

Central tendency: mean, mode

Dispersion: SD, variance, range

5. *Fill in the blank, with "large" or "small":*

If the data are spread out far from the mean, the standard deviation will be large.

If the data are bunched tightly together around the mean, the standard deviation will be small.