Stats 101/101G/108 Workshop

Proportions and Proportional Reasoning [PPR]

2019

by Leila Boyle

Stats 101/101G/108 Workshops

The Statistics Department offers workshops and one-to-one/small group assistance for Stats 101/101G/108 students wanting to improve their statistics skills and understanding of core concepts and topics.

Leila's website for Stats 101/101G/108 workshop hand-outs and information is here: **www.tinyURL.com/stats-10x**

Resources for this workshop, including pdfs of this hand-out and Leila's scanned slides showing her working for each problem are available here: **www.tinyURL.com/stats-PPR**

Leila Boyle
Undergraduate Statistics Assistance, Department of Statistics
Room 303.320 (third floor of the Science Centre, Building 303)
l.boyle@auckland.ac.nz; (09) 923-9045; 021 447-018

Want help with Stats?

Stats 101/101G/108 appointments

Book your preferred time with Leila here: **www.tinyURL.com/appt-stats**, or contact her directly (see above for her contact details).

Stats 101/101G/108 Workshops

Workshops are run in a relaxed environment, and allow plenty of time for questions. In fact, this is encouraged ☺

Please make sure you bring your calculator with you to all of these workshops!

- **Preparation at the beginning of the semester:**

  Multiple identical sessions of a preparation workshop are run at the beginning of the semester to get students off to a good start – come along to whichever one suits your schedule!

  - Basic Maths and Calculator skills for Statistics **www.tinyURL.com/stats-BM**

- **First half of the semester**

  Five theory workshops are held during the first half of the semester:

  - Exploratory Data Analysis **www.tinyURL.com/stats-EDA**
  - Proportions and Proportional Reasoning **www.tinyURL.com/stats-PPR**
  - Observational Studies, Experiments, Polls and Surveys **www.tinyURL.com/stats-OSE**
  - Confidence Intervals: *Means* **www.tinyURL.com/stats-CIM**
  - Confidence Intervals: *Proportions* **www.tinyURL.com/stats-CIP**

- **Second half of the semester**

  Five theory workshops and one computing workshop are held during the second half of the semester:

- **Statistics Theory Workshops**

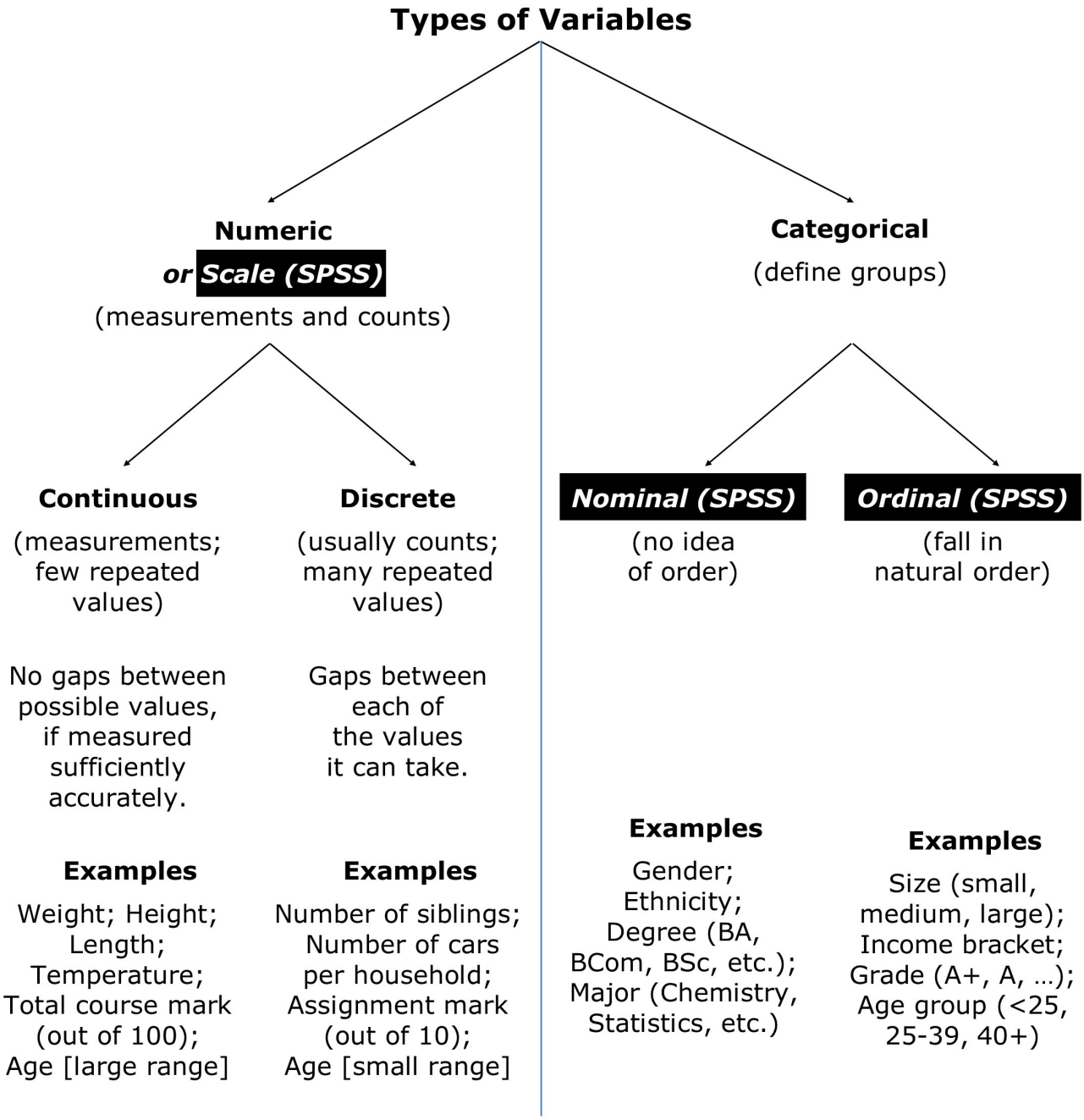
  - Hypothesis Tests: *Proportions* **www.tinyURL.com/stats-HTP**
  - Hypothesis Tests: *Means (part 1)* **www.tinyURL.com/stats-HTM**
  - Hypothesis Tests: *Means (part 2)* **www.tinyURL.com/stats-HTM**
  - Chi-Square Tests **www.tinyURL.com/stats-CST**
  - Regression and Correlation **www.tinyURL.com/stats-RC**

- **Computer Workshop:** Hypothesis Tests in *SPSS* **www.tinyURL.com/stats-HTS**

- **Useful Computer Resource:**

  If you haven't used SPSS before, you may find it useful to work your way through this self-paced workshop: **www.tinyURL.com/stats-IS**

Proportions and Proportional Reasoning



📖**Useful reference:** Chance Encounters, pages 40 – 42

Exploring relationships between two variables

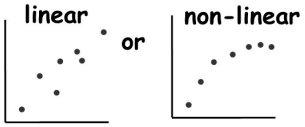
Two Numeric

Exploratory tools to use to explore relationship/s

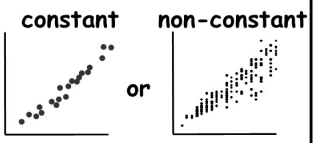
Scatter plot

Features to look for:

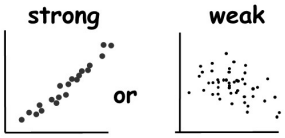
- Trend



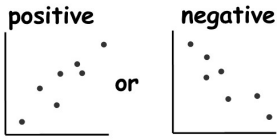
- Scatter



- Strength of relationship

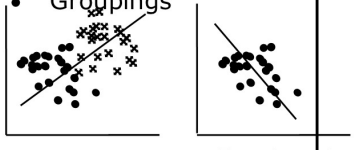


- Association



- Outliers

- Groupings



Numeric versus Categorical

Side-by-side plots on the same scale:

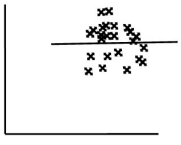
- any n: Dot plots
- $n \geq 20$: Stem-and-leaf plots, Box plots
- $n \geq 50$: Histograms

Features to look for:

Compare the groups by looking at:

- Any group differences:
  - averages (centres): medians, means
  - variability (spread): IQRs, ranges
  - shapes
    - ♦ symmetric/skewed: left/negative, right/positive
    - ♦ modes: unimodal, bimodal, trimodal (not box!)
- Details of individual groups:
  - outliers, gaps, clusters, groupings

Think about reasons *why* these differences, similarities and features are seen



Two Categorical

Two-way table of counts

and/or

Bar graphs of proportions (on rows/columns)

Features to look for:

- Most common and least common combinations
- Differences in distributions (e.g. row and/or column bar graphs)

Useful tips and tricks for dealing with fractions, decimals, proportions and percentages:

- A **proportion** is a number between 0 and 1 that estimates the likelihood of an event occurring.
- Proportions can be presented as **fractions**, **decimals** or **percentages**.
- When dealing with fractions, decimals and percentages, there are a few things you need to keep in mind:
    - 'Per cent' means one out of a hundred. We use **percentages** to describe parts of a whole – the whole being made up of a hundred equal parts. Percentages are used frequently in all walks of life.

      For example: *10% off all shoes*
    - Decimals are usually numbers that use a decimal point followed by digits that show a value smaller than one. We can also use **proportion** for decimals between 0 and 1.

      For example: *0.1*
    - **Proportions** and **percentages** are just different numerical representations of the same value. I find it much easier to understand the value out of 100 compared to the same value out of 1 so I prefer percentages to decimals. However, you will need to be comfortable switching between these two numerical representations!
        - To convert from **proportion** to **percentage**, you **multiply** by 100%.

          For example: *0.1 x 100% = 10%*
        - To convert from **percentage** to **proportion**, you **divide** by 100%.

          For example: *10% ÷ 100% = 0.1*

          Note that 1 = 100% - they are just different representations of the same value!
    - When you are asked to find something "of" something else, you are being asked to **multiply** the two numbers together.

      For example: *½ of 4 = ½ x 4 = 2*
    - When you are asked to find a percentage of something, you must convert the percentage to a decimal, that is, you need to divide the percentage by 100% before doing the multiplication indicated by the "of".
    - When doing fractions on your calculator, you can choose to use the fraction button, or just read the horizontal bar of the fraction as "divided by" (÷).

- When you are dealing with decimals, you can end up with quite small numbers. Your calculator may deal with these by presenting them as "standard form", otherwise known as "scientific notation". This is just your calculator's way of representing a very small number as accurately as it can. You will need to move the decimal point yourself to get it into a decimal number.

  For example: $2.459 \times 10^{-04} = 0.0002459$ *[scientific calculator]*
  $2.459E\text{-}04 = 0.0002459$ *[graphics calculator]*
  (move the decimal point four places to the left).

A note about rounding numbers

Many students ask me about rounding numbers – how to do it and how much by.

When you do a calculation, you may end up with an answer that has many (5 or more) decimal places associated with it. People don't deal too well with numbers to this level of accuracy; rounding helps us make a number a little simpler while still keeping its value relatively close to what it was. The result is less accurate, but easier to use, interpret and understand.

How to round numbers

- Decide which is the last digit to keep
- Leave it the same if the next digit is less than 5 (this is called **rounding down**)
- Increase it by 1 if the next digit is 5 or more (this is called **rounding up**)

How many decimal places should I use?

When you are doing calculations in an assignment or test/exam context, you should get some guidance from the question itself. For example, a multi-choice test question may have five answers that are all rounded to one decimal place (1dp for short) so just round your answer when you do the calculation to 1dp.

If you are going to use the value you come up with in a later calculation, go for more accuracy rather than less, as the more you round a number, the more it will affect the results of later calculations that use that value. My default amount of rounding tends to be 4dp – this is reasonably accurate without going over the top.

- Our main source of proportions is from data which will usually be presented in a table of counts.
- An **event** is a collection of outcomes. An event **occurs** if any outcome making up that event occurs.
- We sometimes label events, for example, let *A* be the event that … occurs.

There are three types of problems you may be asked to solve:

1. From a table of counts, find the proportion/s that … . These questions are easier because you are given all of the information you need in the table of counts: all you need to figure out is what piece/s of information to extract from the table!

2. From a short story with proportions, and/or percentages, and/or counts about two factors (categorical variables), find the proportion/s that … :
   - ☞ These are hard questions!
   - ☞ Need to *interpret* the story first, and then construct a table.
   - ☞ Once your table is constructed, use it to find 1 or 2 proportions.
   - ☞ Steps to constructing a table:

     Step 1: highlight numbers

     Step 2: highlight factors

     Step 3: define factor levels

     Step 4: label table

     Step 5: enter appropriate table total

     Step 6: enter row/column totals from story

     Step 7: enter cell numbers from story

     Step 8: enter remaining numbers by +/-

3. From an iNZight/*Excel* graphical display with proportions (or percentages), about one, two or three factors (categorical variables), find the proportion/s that … :
   - ☞ These are hard questions!
   - ☞ Need to read the story carefully first, and then interpret what the graph/s are showing.
   - ☞ Once you are clear on what is being shown in the graph/s, then you need to use it/them to find at least one proportion, and to assess whether statements are true or false.

Types of Proportion Calculations

There are only four possible types of proportion calculations and you may be asked to do two or three of them in the test/exam.

1. Proportion of <u>AN EVENT</u> (basic/simple)

- ☞ pr(*A*) → pr (an event)
- ☞ *Finding* the proportion for a single event

2. Proportion of an event <u>AND</u> another event:

- ☞ pr(*A* and *B*) → pr(one event and another event)
- ☞ *Finding* the proportion of both *A* <u>*and*</u> *B* occurring at the same time

3. Proportion of an event <u>OR</u> another event:

- ☞ pr(*A* or *B*) → pr(one event or another event)
- ☞ *Finding* the proportion of *A* only and/or *B* only and/or both *A* <u>*and*</u> *B* occurring

} Use ***GRAND TOTAL (TABLE TOTAL)***

4. <u>CONDITIONAL</u> Proportion:

- ☞ Harder to detect but will often have key words like "Given that…", "Of those…", "Among those…", "If".
- ☞ Look for language that <u>restricts you to part of the table</u> instead of the whole table.

} Use ***ROW TOTAL/S*** *OR* ***COLUMN TOTAL/S***

$$\text{pr (one event given another event)} = \text{pr}(A \text{ given } B) = \frac{\text{pr}(A \text{ and } B)}{\text{pr}(B)}$$

Examples 1 to **4** are about the following information.

Market researchers and pollsters worry that having chosen a sample, the people that they fail to contact may differ in important ways from the ones that they do contact. Research was done where interviewers telephoned designated houses up to 3 times to make contact with the residents. Residents' income levels as well as the number of calls required to contact each resident are tabulated in Table 1 below:

	Calls to respondents			
Income	First call respondents	Second call respondents	Third call respondents	**Total**
Less than $10,000	28	4	2	34
$10,000-30,000	51	11	3	65
$30,000-50,000	41	17	4	62
$50,000-70,000	88	27	11	126
$70,000 or more	96	55	36	187
Total	304	114	56	474

Table 1: Income by calls to respondents

Example 1: The proportion of residents who were contacted on the first call is approximately:

pr( ) = _____

= (4dp)

Example 2: The proportion of residents earning $50,000 or more who were contacted on the first call is approximately:

pr( ) = _____

= (4dp)

Example 3: The percentage of residents who were contacted on the first call and earning less than $10,000 is approximately:

pr( ) = _____

= (2dp)

Example 4: The proportion of residents who were earning less than $50,000 or were contacted on the second call is approximately:

pr( ) = _____

= (4dp)

Table 2 is used in **Examples 5** to **8**. It shows household income according to family type (A = Solo parent with one or more children, B = Couple with one or more children, C = Couple with no children, D = Single person, E = Other).

Household Income	**Family Type** A	B	C	D	E	**Total**
	No. of households (× 1000)					
0-9999	9	13	5	57	2	86
10,000 - 19,999	43	26	61	92	8	230
20,000 - 39,999	36	107	93	63	40	339
40,000 - 74,999	13	153	87	17	47	317
75,000 +	1	61	25	3	19	109
Total	102	360	271	232	116	1081

Table 2: Income by calls to respondents

Example 5: Given that the family is of type A [Solo parent with child(ren)], the proportion with household income of less than $20,000 is approximately:

pr( ) = _____

= (4dp)

Example 6: The proportion that a family's household income is less than $20,000 or the family is of type A is approximately:

pr( ) = _____

= (4dp)

Example 7: A family is type A and their household income is less than $20,000 with a proportion of approximately:

pr( ) = _____

= (4dp)

Example 8: Those families with household incomes less than $20,000 have a proportion of approximately:

pr( ) = _____

= (4dp)

Check you understand! Questions

Questions 1 to **3** refer to the following information.

At the beginning of the semester, stage one statistics students at the University of Auckland were surveyed. One question asked: 'How would you describe your Excel experience?'. A total of 918 students answered this question. Each answer was classified according to the response given by the student, and the stream the student attended. The results are given in the table below, where 101G, 101 and 108 refer to the various streams.

	Stream			
Response	101G	101	108	**Total**
None	15	36	102	153
Very Little	44	89	119	252
Some	74	150	200	424
Lots	9	29	51	89
Total	142	304	472	918

Table 3: Responses to question regarding Excel experience.

1. Of those students in this survey who had no Excel experience at the start of the course, the proportion who were in stream 108 is:

   (1) 0.216
   (2) 0.667
   (3) 0.167
   (4) 0.111
   (5) 0.514

2. The proportion of students in this survey who had no Excel experience at the start of the course or were in stream 108 is:

   (1) 0.514
   (2) 0.667
   (3) 0.167
   (4) 0.111
   (5) 0.570

3. The proportion of students in this survey who had no Excel experience at the start of the course is:

   (1) 0.216
   (2) 0.667
   (3) 0.167
   (4) 0.111
   (5) 0.514

Questions 4 to **6** are about the following information.

As part of its responsibilities to EEO (Equal Employment Opportunities), a New Zealand Government department sent out a survey to its employees. In one section of the survey the employees were asked to consider the statement "*Ethnic and cultural diversity is welcomed in this workplace*" and to make one selection from a given list of responses. The table below cross classifies ethnic origin of the respondent and their chosen response to the statement.

Ethnic Origin	1 Agree Strongly	2 Agree Somewhat	3 Neither Agree nor Disagree	4 Disagree Somewhat	5 Disagree Strongly	6 No Opinion	Total
Pakeha (NZ European)	32	35	11	10	10	6	104
Maori	1	7	2	2	0	0	12
Non-NZ European	1	1	0	0	0	0	2
Pacific Islander	3	4	4	1	0	0	12
Asian	4	5	2	0	0	0	11
Other	1	0	10	3	4	0	18
Total	42	52	29	16	14	6	159

Table 1: Income by calls to respondents

4. The percentage of those surveyed who **agreed strongly** with the statement is approximately:
   (1) 30.8%
   (2) 26.4%
   (3) 0.26%
   (4) 0.59%
   (5) 20.1%

5. The proportion of employees that are **Pacific Islanders** who either **agreed somewhat** or **agreed strongly** with this statement is approximately:
   (1) 0.074
   (2) 0.075
   (3) 0.044
   (4) 0.583
   (5) 0.591

6. Of those who **disagreed strongly** with the statement, the proportion who were **Pakeha** is approximately:
   (1) 0.654
   (2) 0.714
   (3) 0.088
   (4) 0.096
   (5) 0.063

Questions 7 to **9** are about the following information.

A nation-wide telephone survey was conducted on behalf of the Alcohol and Public Health Research Unit in order to assess drinking patterns and alcohol related issues in New Zealand. A random sample of 4,232 people aged between 14 to 65 years was selected using telephone numbers from throughout New Zealand. The response rate was 76%. The gender/age composition of the 'top 10%' of drinkers (those who had consumed the most alcohol in the previous 12 months) is given in the table below.

	Male	Female	Total
14-17 years	11	4	15
18-19 years	25	11	36
20-24 years	92	21	113
25-29 years	42	7	49
30-39 years	56	11	67
40-49 years	32	3	35
50-65 years	35	3	38
Total	293	60	353

Table 7 : Gender/Age Composition of the 'top 10%' of Drinkers.

7. The proportion of the 'top 10%' of drinkers who are under 20 years of age is approximately:

(1) 0.3729
(2) 0.1445
(3) 0.3201
(4) 0.4646
(5) 0.0425

8. The percentage of the 'top 10%' of drinkers who are males and are in the 20-24 year age group is approximately:

(1) 26.1%
(2) 32.0%
(3) 81.4%
(4) 31.4%
(5) 43.7%

9. Which one of the following statements is false for the above table?

(1) This two-way table of counts is an appropriate tool to use to explore the relationship between gender and age composition among the 'top 10%' of drinkers.
(2) The gender classification in this table enables a comparison to be made between the number of males and females in the 'top 10%' of drinkers.
(3) The gender of the respondent is a categorical variable.
(4) The age group in which a respondent is classified is a categorical variable.
(5) The age group classification in this table allows the number of the 'top 10%' of drinkers in their forties to be compared with the number in their fifties.

Questions 10 and **11** are about the following information.

The data in the following table was drawn from a study of 507 Swedish women which was reported in the *Journal of Consulting and Clinical Psychology*. Each woman had an aggressiveness score recorded at age 10 by their teacher and then the number of convictions for law breaking until they turned 30 was recorded 20 years later.

Number of Crimes by Age 30	**Aggressiveness Score at Age 10**			
	Low	Medium	High	Total
0	152	291	21	464
1	6	17	3	26
2 or more	3	11	3	17
Total	161	319	27	507

Table 1: Income by calls to respondents

10. The proportion of women convicted of at least one crime by the age of 30 is approximately:

(1) 0.018
(2) 0.051
(3) 0.034
(4) 0.056
(5) 0.085

11. Given that a woman had a high score for aggressiveness by age 10, the proportion that she is convicted of at least one crime by the age of 30 is approximately:

(1) 0.085
(2) 0.222
(3) 0.140
(4) 0.630
(5) 0.286

12. Yearlings are one-year-old thoroughbred horses; male yearlings are colts while females are fillies. Cambridge Stud sold 40 yearlings at a sale, of which 26 were colts. All 40 of the yearlings were sold to either Asian, Australian or New Zealand buyers. Australians bought 23 yearlings. Asians bought 4 colts and no fillies, and New Zealanders bought 8 colts and some fillies.

Of the yearlings sold by Cambridge Stud at the sale, what percentage were colts bought by Australian buyers?

(1) 35.0%
(2) 60.9%
(3) 87.5%
(4) 53.8%
(5) 88.5%

13. One year at the University of Auckland, 5,768 students enrolled in at least one Mathematics or Statistics course. Of these, 2,496 enrolled in two or more such courses and 1,723 men enrolled in exactly one course. 2,445 of the 5,768 students were women.

    Of those students enrolled in two or more Mathematics or Statistics courses, the proportion of women is approximately:

    (1) 0.366

    (2) 0.433

    (3) 0.359

    (4) 0.424

    (5) 0.155

14. The Journal of Paediatrics and Child Health published results from a Hawkes Bay study into child obesity in New Zealand. From the data collected from 894 children, 313 of the 633 children within the normal weight range were females. Of the 450 females, 99 were overweight. 33 males were obese.

    The proportion of obese children in the study who were female is approximately:

    (1) 0.0794

    (2) 0.5034

    (3) 0.5352

    (4) 0.0425

    (5) 0.0844

15. A lecturer of a course at The University of Auckland was interested in investigating the relationship between the term test and the exam performance of the 176 students enrolled in the course in the first semester. She found that 74% of the students passed the term test. Of those who passed the term test, 65% passed the exam, while those students who failed the term test had a 32% pass rate in the exam.

    The percentage of enrolled students who passed both the term test and the exam is approximately:

    (1) 73.9%

    (2) 56.4%

    (3) 48.3%

    (4) 85.0%

    (5) 65.4%

16. In June 2008 the Electoral Commission conducted a telephone survey of a random sample of 3000 adult New Zealanders to find out how well they understood the MMP voting system. A question in the survey was:

> Which of the two votes that you have in MMP is more important in deciding the share of seats each party has in parliament: the party vote or the electorate vote?

Of the females in the sample, 64% replied 'party vote' and 26% replied 'electorate vote'. Of the males in the sample, 70% replied 'party vote' and 24% replied 'electorate vote'. The remainder of both groups were unsure of the answer. Females made up 52% of the sample.

Of those who replied 'party vote', what proportion were male?

(1) 0.333

(2) 0.502

(3) 0.336

(4) 0.460

(5) 0.700

17. The hemoccult test is a screening test for colorectal cancer. This test checks for blood in faeces. Screening is usually carried out on people over 50 years of age.

For the population being screened:

- for those with colorectal cancer, the proportion of a positive test is 0.6
- for those without colorectal cancer, the proportion of a positive test is 0.04.

It is estimated that 0.3% of the population being screened has colorectal cancer.

Of those who have a positive hemoccult test, approximately what proportion have colorectal cancer?

(1) 0.960

(2) 0.002

(3) 0.004

(4) 0.043

(5) 0.600

Questions 18 and **19** refer to the following information.

A study in the UK investigated bullying amongst students at school. A random sample of pupils from one school completed a questionnaire as part of this study. (Source: *The Star Project*)

18. One question asked the pupils how often they were bullied at school. Table 5 shows the results by gender:

Gender	Never	Rarely	Often	**Total**
Boy	34	30	24	88
Girl	48	27	30	105
Total	82	57	54	193

Table 5: How often students were bullied

Which **one** of the following statements is **false**?

(1) Under 20% of the pupils are boys who had never been bullied.

(2) The majority of pupils who had been rarely bullied are boys but the majority of pupils who had been often bullied are girls.

(3) Proportionally, more boys than girls had ever been bullied.

(4) The majority of the pupils had never been bullied.

(5) Fewer boys than girls responded to this question.

19. Another question asked the 111 pupils who had been bullied (54 boys and 57 girls) to identify the gender of the pupils who bullied them: both boys and girls; only boys; only girls.

Approximately 94% of the boys and 33% of the girls had been bullied by only boys. None of the boys and 22 of the girls had been bullied by only girls.

Approximately what percentage of the pupils who had been bullied by both girls and boys were girls?

(1) 14%

(2) 84%

(3) 16%

(4) 28%

(5) 51%

Questions 20 to **22** refer to the following information.

The New Zealand Department of Labour regularly collects data on employment in New Zealand. Table 1 is based on the quarterly regional labour market summary, March 2010.

Working Age Population

Gender	Employed	Unemployed	Not in Labour Force	**Total**
Male	1,148,500	75,000	416,000	1,639,500
Female	1,015,900	66,600	657,100	1,739,600
Total	2,164,400	141,600	1,073,100	3,379,100

Table 1: Quarterly NZ employment data, March 2010

20. Of the New Zealand working age population who were male, the proportion who were employed is approximately:

(1) 0.701
(2) 0.531
(3) 0.340
(4) 0.641
(5) 0.485

21. Excluding those not in the labour force, approximately what percentage of the New Zealand working age population were female and employed?

(1) 44.1%
(2) 30.1%
(3) 96.7%
(4) 46.9%
(5) 58.4%

22. 22% of those employed in the working age population were employed part-time. Of these, 72% were female, while 62% of those employed full-time were males.

The proportion that a randomly chosen female from the employed New Zealand working age population was employed part-time is approximately:

(1) 0.484
(2) 0.720
(3) 0.348
(4) 0.158
(5) 0.455

Question 23 refers to the following information.

Myers *et al.* (2010) conducted a study with 270 customers of a Boston (USA) coffee shop.

Some variables used in the study were:

Gender	The gender of the customer – Female – Male
Coffee type	The type of coffee ordered – Fancy – Not fancy
Waiting time	The time between ordering and receiving coffee, in seconds

Assume the study involves a random sample from the population of all customers of this Boston coffee shop.

23. 51% of the customers in this study were female. 75% of orders placed by the female customers were for fancy coffees and 55% of orders placed by the male customers were for fancy coffees.

    Of the customers in this study who ordered fancy coffees, the proportion who were female is approximately:

    (1) 0.587

    (2) 0.680

    (3) 0.750

    (4) 0.383

    (5) 0.577

Questions 24 and **25** refer to the following information.

In 2013, Auckland Transport conducted a survey on cycling in which a random sample of 1048 adult Aucklanders, aged 15 years or over, were surveyed (AT, 2013). Respondents were asked questions related to their usage and attitudes towards cycling through an online survey. Some of the questions asked in the survey were:

> *Do you own or otherwise have access to a bicycle? If so, about how often do you use a bicycle for any reason?*

The responses were used to classify each respondent by **Type of Cyclist**: a non-cyclist (no use of a bicycle), an irregular cyclist (uses a bicycle, but no more than once per month, on average) or a regular cyclist (uses a bicycle more than once per month, on average).

Table 2 shows the survey data cross-classified by **Type of Cyclist** and **Gender**.

	Type of Cyclist			
Gender	Non-cyclist	Irregular cyclist	Regular cyclist	**Total**
Male	311	104	88	503
Female	432	85	28	545
Total	743	189	116	1048

Table 2: Gender and Type of cyclist

Figure 4 shows the stacked bar chart for Type of Cyclist and Gender.

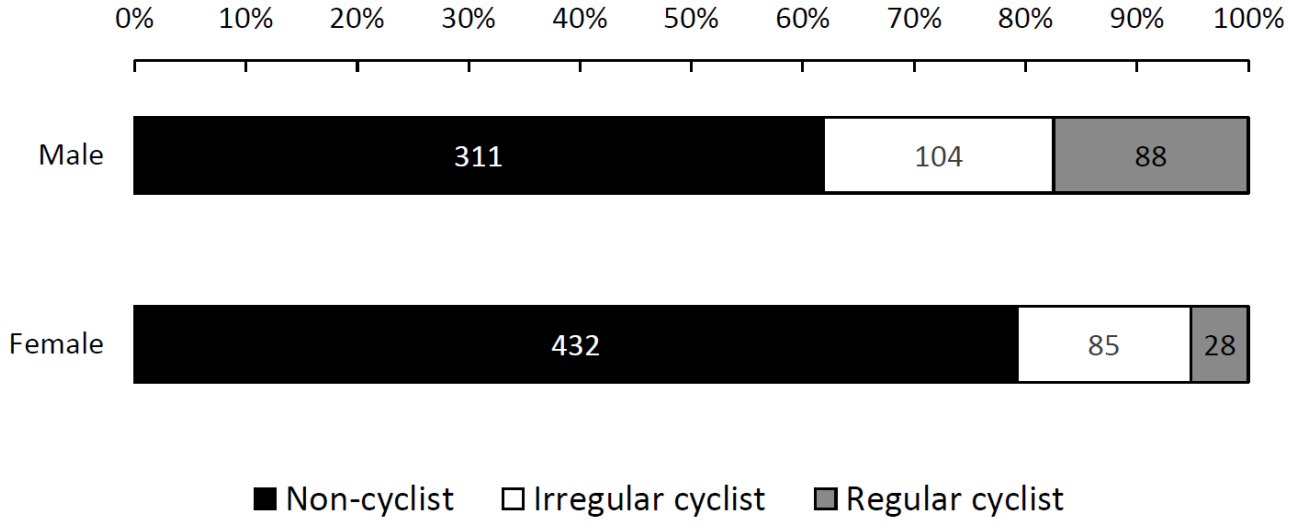


Figure 4: Type of Cyclist by Gender

24. Of the regular cyclists, approximately what percentage were male?
    (1) 24%
    (2) 76%
    (3) 8%
    (4) 88%
    (5) 18%

25. Which **one** of the following statements is **false**?
    (1) For these respondents, a non-cyclist is about 1.4 times as likely to be a female than a regular cyclist is to be a female.
    (2) Approximately 29% of these respondents use a bicycle.
    (3) For these respondents, a male is over three times as likely to be a regular cyclist than a female is to be a regular cyclist.
    (4) A female respondent to this survey is more likely to be a non-cyclist than a male respondent is to be a non-cyclist.
    (5) A respondent to this survey is more likely to be a female non-cyclist than a male irregular cyclist.

Questions 26 to 29 refer to the following information.

Statistics New Zealand collects data on the number of claims made each year for work-related injuries. The 2014 data was downloaded from `www.stats.govt.nz`.

Figure 5 shows the percentage of claims for work-related injuries by age group for both males and females. (Note that those aged 14 years and under and those who did not specify their age are excluded from this analysis.)

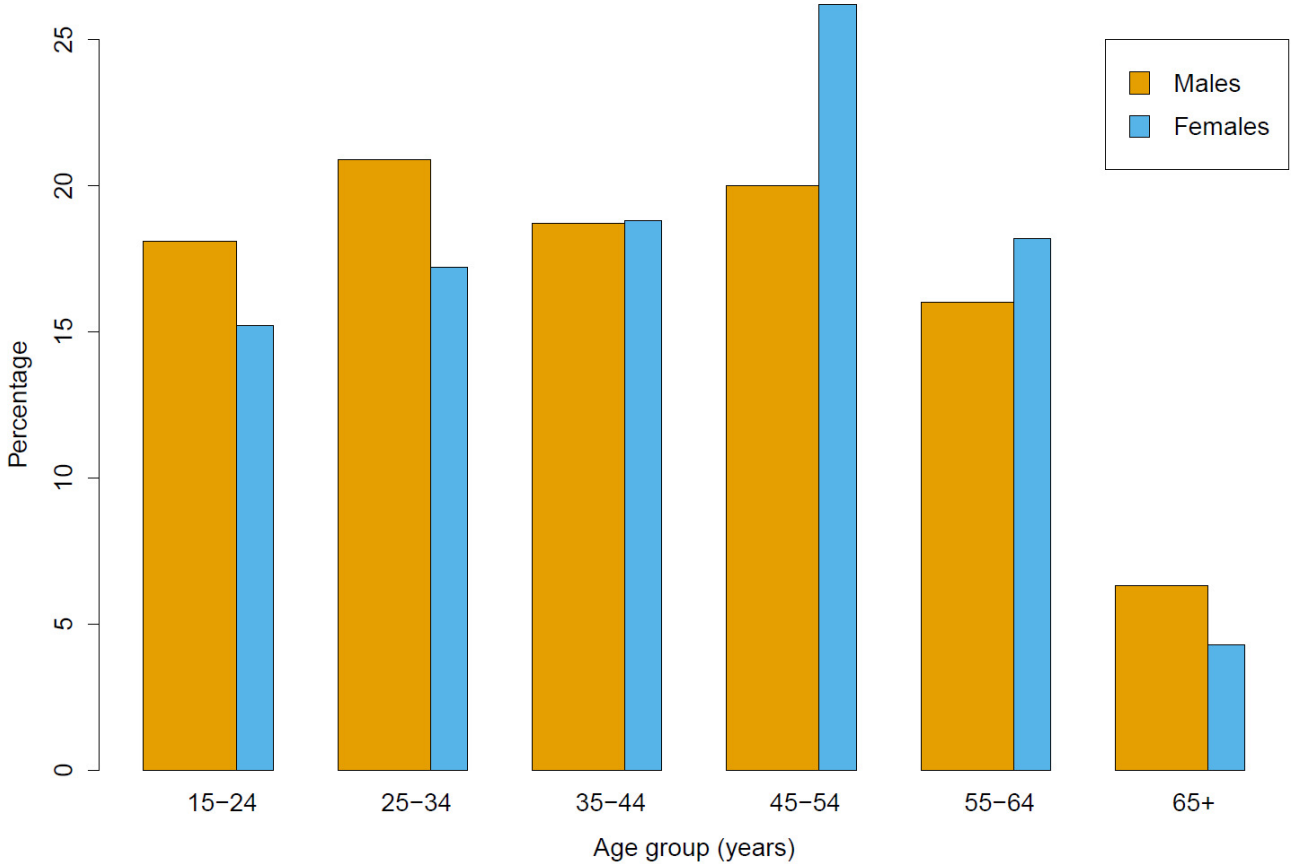


Figure 5: Distribution of claims by age group for males and females

26. Which **one** of the following statements is **false**?

(1) The distribution of claims by age group appears to depend on sex.

(2) The percentage of males aged 15-34 years who made claims was greater than the percentage of females aged 15-34 years who made claims.

(3) More males than females made claims in 2014.

(4) Across all 12 of the gender/age group combinations, females in the 45-54 age group made the highest number of claims.

(5) For both males and females, those in the 65+ age group made the smallest number of claims.

Questions 27 and **28** refer to the following additional information.

Table 1 gives the number of work-related claims in 2014 cross-classified by sex and occupation grouping.

Occupation grouping	Males (000)	Females (000)	**Total** (000)
Legislators, administrators, and managers	8.1	5.3	13.4
Professionals	6.6	9.6	16.1
Technicians and associate professionals	9.3	8.1	17.4
Clerks	3.5	4.2	7.7
Service and sales workers	12.1	15.7	27.8
Agriculture and fishery workers	25.5	7.0	32.5
Trades workers	38.9	1.0	39.9
Plant and machine operators and assemblers	28.4	3.4	31.8
Elementary occupations	19.1	5.8	24.8
Not specified	10.2	4.5	14.7
Total	161.7	64.5	226.1

Table 1: Work-related claims by sex and occupation grouping

27. Approximately what proportion of the claims made by clerks were made by females?

    (1) 0.285
    (2) 0.065
    (3) 0.545
    (4) 0.019
    (5) 0.034

28. Approximately what proportion of the claims were made by female trades workers?

    (1) 0.285
    (2) 0.016
    (3) 0.025
    (4) 0.004
    (5) 0.176

29. In 2014 there were 145,000 males aged 15 to 24 years in paid employment and 114,000 females aged 15 to 24 years in paid employment, respectively. Work-related claims were made by 29,200 males and 9,800 females, aged 15 to 24 years.

    Using this information, an estimate of the relative risk of a work-related claim for a male aged 15 to 24 years compared to a female aged 15 to 24 years is:

    (1) 2.3

    (2) 0.4

    (3) 3.0

    (4) 0.3

    (5) 1.3

Questions 30 and **31** refer to the following information.

Newhouse et al. (2010) investigated whether nicotine patches had an effect on improving the memory of non-smokers with mild declines in their thinking ability.

67 subjects previously diagnosed with mental cognitive impairment (MCI) were randomly assigned to two treatment groups; the nicotine patch group and the placebo patch group. Treatment continued for six months. The Clinical Global Impression of Change (MCI-CGIC) assessment instrument was used as the primary clinical outcome measure. Based on their MCI-CGIC each subject was categorised as either improved, no change or worsened. The results are shown in Table 5.

	Change in Memory			
Patch Type	Improved	No change	Worsened	**Total**
Nicotine	8	20	6	34
Placebo	3	21	9	33
Total	11	41	15	67

Table 5: **Change in Memory** by **Patch Type**

30. The proportion of subjects who are in the placebo patch group and whose memory worsened, to two decimal places, is:

    (1) 0.27

    (2) 0.60

    (3) 0.22

    (4) 0.13

    (5) 0.58

31. How many times as likely is it, to two decimal places, that a subject in the nicotine patch group showed an improvement in memory relative to a subject in the placebo patch group?

    (1) 1.03 times as likely

    (2) 1.43 times as likely

    (3) 0.39 times as likely

    (4) 1.38 times as likely

    (5) 2.59 times as likely

32. A random sample of 2625 US adults were asked whether they agreed or disagreed that "*there is only one true love for each person.*" In total 735 people agreed with the statement, 1812 disagreed and 78 said they didn't know. Of the males 372 agreed and 807 disagreed, while 34 said they didn't know.

    How many times as likely was a male to have agreed with the statement than a female to have agreed with the statement?

    (1) 0.98

    (2) 1.19

    (3) 1.02

    (4) 0.83

    (5) 0.51

Questions 33 to **35** refer to the following information.

Research New Zealand conducts monthly telephone surveys in which they survey a random sample of 500 adult New Zealanders, aged 18 years or over. In the July 2015 survey they included a question designed to measure the opinions of adult New Zealanders on the legalisation of euthanasia. This question had also been asked in their April 2015 survey (ResearchNZ, 2015).

The question asked was:

> *Suppose a person has a painful incurable disease. Do you think that doctors should be allowed by law to end the patient's life if the patient requests it?*

The answers to the question (**Answer**), classified by **Age Group**, for the July 2015 survey are shown in Table 1 and Figure 4.

	Age Group (years)			
Answer	18 – 34	35 – 54	55 and over	**Total**
Yes	44	136	148	328
No	14	45	63	122
Don't know	5	13	32	50
Total	63	194	243	500

Table 1: **Answer** by **Age Group**

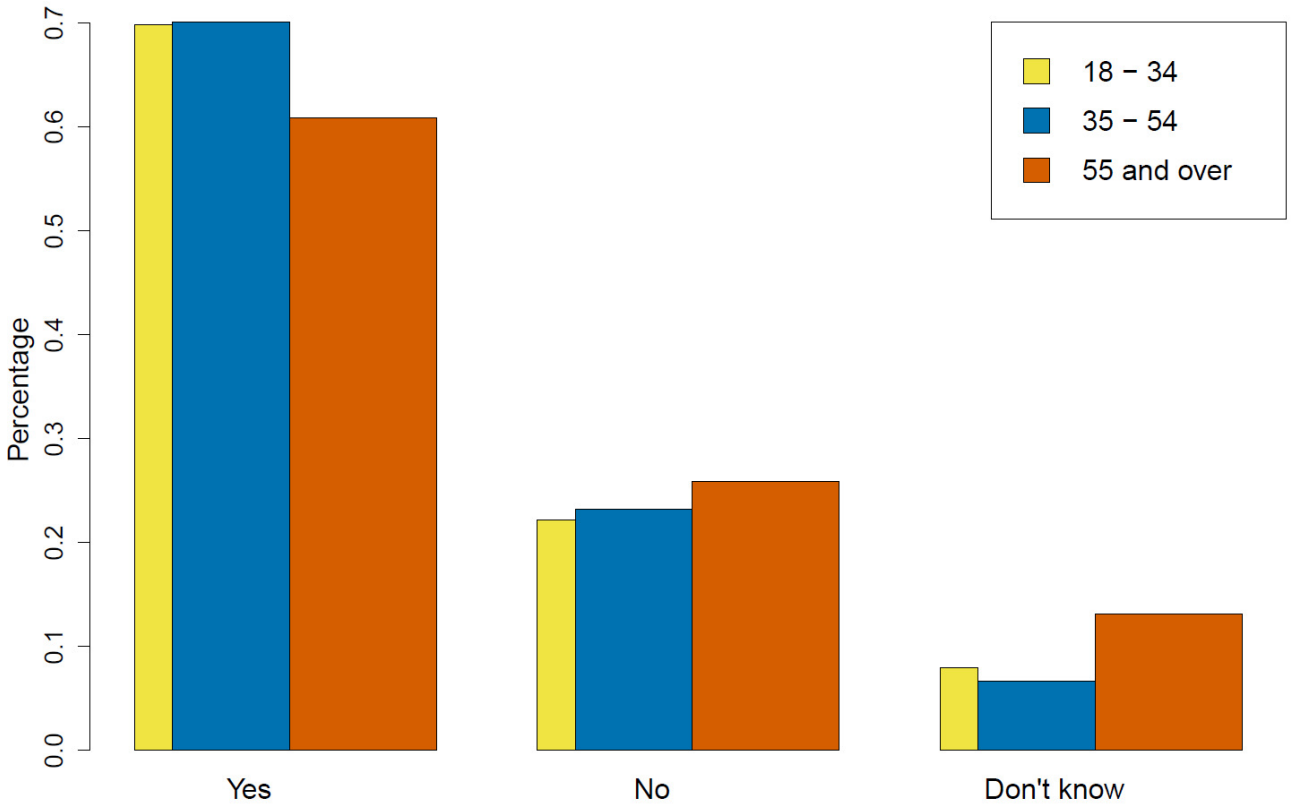


Figure 4: **Answer** by **Age Group**

33. The percentage of respondents, to the nearest whole number, who were over 34 years of age and answered 'No' is:

    (1) 89%

    (2) 22%

    (3) 25%

    (4) 87%

    (5) 9%

34. Of the respondents who answered 'Don't know', the proportion who were aged 55 and over, to 2 decimal places, is:

    (1) 0.13

    (2) 0.49

    (3) 0.10

    (4) 0.06

    (5) 0.64

35. Based on Table 1 and/or Figure 4, page 27, which one of the following statements is true?

    A respondent aged aged 35 – 54 years, compared to a respondent aged 55 years or over, is approximately:

    (1) 1.50 times as likely to have answered 'Yes'.

    (2) 1.08 times as likely to have answered 'Yes'.

    (3) 1.15 times as likely to have answered 'Yes'.

    (4) 2.33 times as likely to have answered 'Yes'.

    (5) 0.87 times as likely to have answered 'Yes'.

Questions 36 to **38** refer to the following information.

The General Social Surveys are conducted in the USA by the National Opinion Research Center every two years. In the 2014 survey, 2503 adult Americans answered both of the following two questions:

> *Compared with families in general, would you say your family income is far above average, above average, average, below average, far below average?*

and

> *Taken all together, how would you say things are these days - would you say that you are not too happy, pretty happy, very happy?*

A summary of the responses to these two questions is shown in Table 1 and the distribution of Happiness by Family Income is shown in Figure 4.

	Happiness			
Family Income	Not too happy	Pretty happy	Very happy	**Total**
Far above average	11	29	25	65
Above average	31	265	185	481
Average	101	620	394	1115
Below average	131	391	141	663
Far below average	63	83	33	179
Total	337	1388	778	2503

Source: Data from CSM, UC Berkeley

Table 1: **Happiness** by **Family Income**

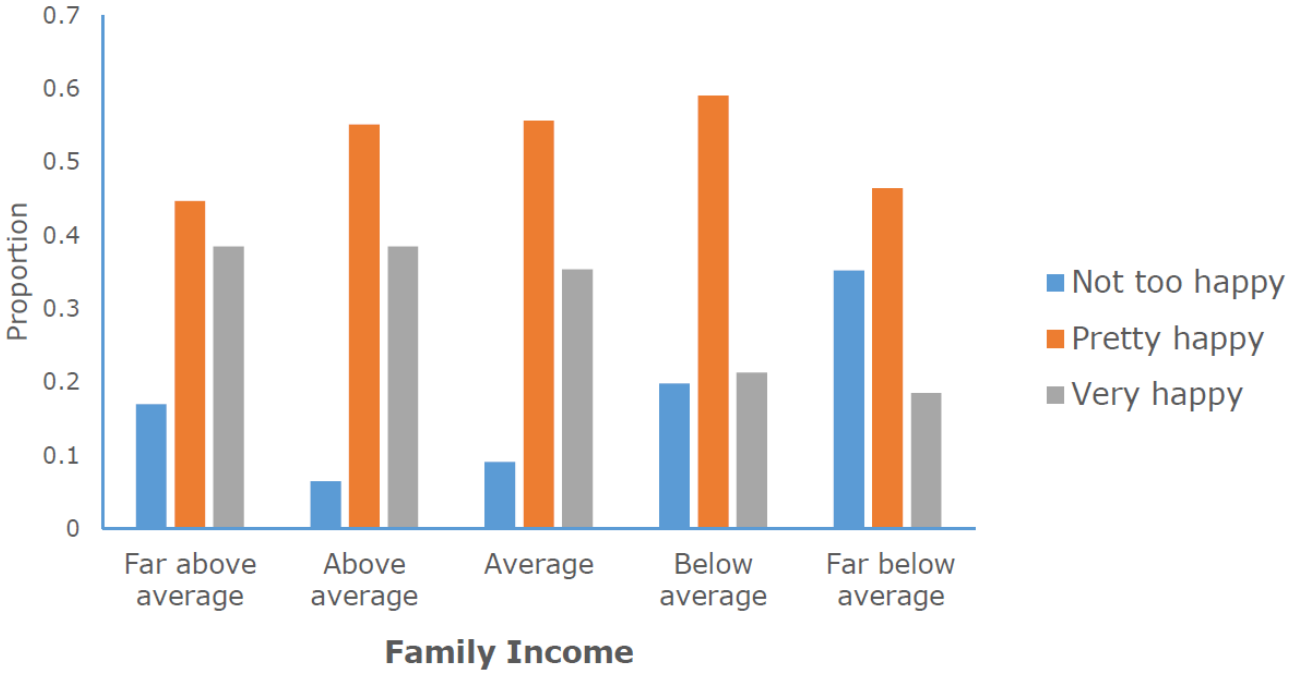


Figure 4: Distribution of **Happiness** by **Family Income**

36. As used in Table 1, page 29, which **one** of the following statements about the variables Family Income and Happiness is **true**?

    (1) **Family Income** is a discrete variable because its values are to the nearest whole dollar.

    (2) **Happiness** is a nominal variable because its levels convey a sense of order.

    (3) Both **Family Income** and **Happiness** are categorical variables.

    (4) **Family Income** is a numeric variable because it measures the number of dollars.

    (5) Both **Family Income** and **Happiness** are numeric variables.

Use Table 1 and/or Figure 4, page 29, to answer **Questions 3** and **4**.

37. Which **one** of the following statements is **false**?

    (1) Approximately 25% of the respondents reported that they had an average family income and that they were pretty happy.

    (2) The respondents who reported a far above average family income were more likely to have reported to be not too happy than the respondents who reported an average family income.

    (3) There is an approximately 1.8 percentage point difference when comparing the 'pretty happy' percentages for those who reported a far above average family income and those who reported a far below average family income,

    (4) The distribution of **Happiness** for the two groups who reported above average and below average family income are very different.

    (5) The distribution of **Happiness** for the two groups who reported above average and average family income are very different.

38. Of the respondents who reported being very happy, the percentage who also reported a family income below average or far below average is approximately:

    (1) 22.4%

    (2) 9.1%

    (3) 4.2%

    (4) 7.0%

    (5) 20.7%

39. Figure 1 shows hard drug use by gender, subset by age group, from a data set derived from the NHANES database (`www.cdc.gov/nchs/nhanes.htm`).

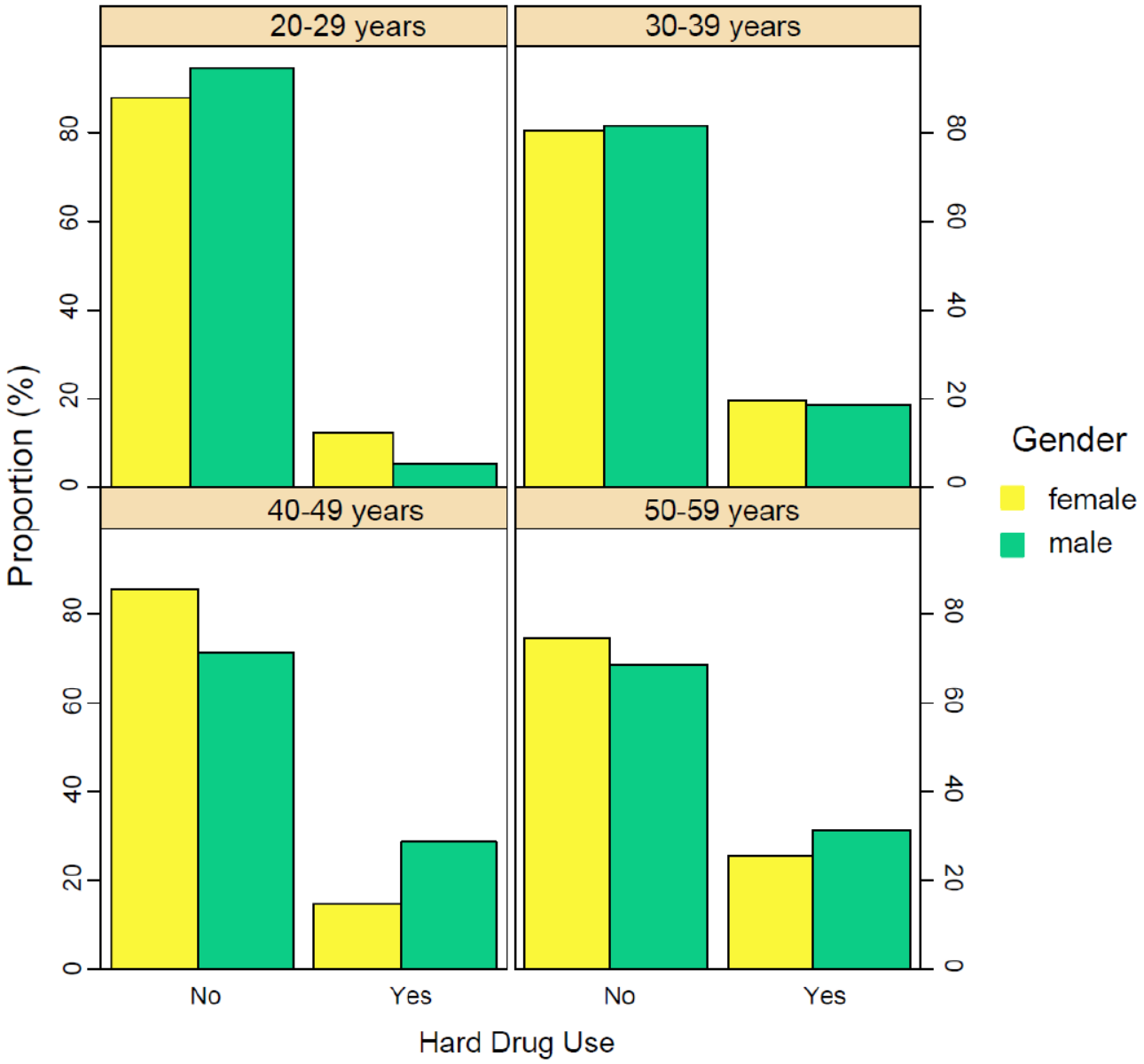


Figure 1: The distribution of hard drug use by gender, subset by age group

Which **one** of the following statements can **not** be deduced from Figure 1?

(1) In the 50-59 years group, roughly 30% of the people who said 'Yes' to hard drug use were males.

(2) In the 30-39 years group, there were more males than females.

(3) The age group in which people were least likely to have said 'Yes' to hard drug use is the 20-29 years group.

(4) In the 40-49 years group, the males were almost twice as likely as the females to have said 'Yes' to hard drug use.

(5) The age group with the biggest difference between the male and female responses about hard drug use is the 40-49 years group.

40. The NZ Herald reported the results of a two-year study at Hong Kong's Kwong Wah Hospital. The study comprised 5450 impotent men who were given Viagra at the hospital. Of these men, 3651 were smokers. Of the 926 impotent men for whom Viagra did not work, 840 were smokers.

    The percentage of men in the study who were smokers and for whom Viagra did not work was approximately:

    (1) 23.0%
    (2) 51.6%
    (3) 82.4%
    (4) 90.7%
    (5) 15.4%

ANSWERS

1.	**(2)**	2.	**(5)**	3.	**(3)**	4.	**(2)**	5.	**(3)**	6.	**(2)**
7.	**(2)**	8.	**(1)**	9.	**(5)**	10.	**(5)**	11.	**(2)**	12.	**(1)**
13.	**(3)**	14.	**(3)**	15.	**(3)**	16.	**(2)**	17.	**(4)**	18.	**(4)**
19.	**(2)**	20.	**(1)**	21.	**(1)**	22.	**(3)**	23.	**(1)**	24.	**(2)**
25.	**(1)**	26.	**(4)**	27.	**(3)**	28.	**(4)**	29.	**(1)**	30.	**(4)**
31.	**(5)**	32.	**(2)**	33.	**(2)**	34.	**(5)**	35.	**(3)**	36.	**(3)**
37.	**(5)**	38.	**(1)**	39.	**(1)**	40.	**(5)**				

WHAT SHOULD I DO NEXT?

Once you've had a go at all of the problems in the handout (check out Leila's scanned slides at www.tinyURL.com/stats-PPR for her additional handwritten notes and workings for every problem), you could:

- Go through the Chapter 1 blue pages. The blue pages relevant to the material in this workshop are:
  - the notes on page 28 (left-hand side and the top half of the right-hand side – note that the latter is **not examinable**!)
  - the glossary on pages 29 and 30
  - two true/false statements on page 31 (z. and cc.)
  - the questions on pages 34-35 (Questions 10-12) and page 38 (Questions 20-24)
  - the tutorial material (Section B on pages 42-44)
- Try Chapter 1 questions from three of the past five tests that are relevant to this workshop.
- Make sure you pick up your Assignment 1 answers from the Science Student Resource Centre and check them against the model answers on Canvas. (The material covered in this workshop was assessed in Questions 3 and 4 of Assignment 1.)