UNIVERSITY OF TORONTO SCARBOROUGH
Department of Computer and Mathematical Sciences
STAB22H3F – Statistics I
Midterm Test: June 28th, 2019
Duration: 2 Hours

Solution

Test Information & Instruction

1. Before you begin the test, complete the signature sheet, but sign it only when the invigilator collects it. The signature sheet shows that you were present at the test. Please have your student ID visible for the invigilator to verify your information.

2. On the Scantron answer sheet ensure that you enter your last name, first name (as much of it as fits), and your student number (in "Identification"). The Scantron answer sheet is the only paper that you will return to the test invigilator after you finish the test.

3. Aids allowed for the test:

   - One handwritten letter-sized sheet (both sides) of notes, which is prepared by you.
   - A non-programmable, non-communicable calculator.

4. This test is based on multiple-choice questions. There are 35 questions. All questions carry equal weight. Mark in each case the best answer out of the alternatives given (**which means the numerically closest answer if the answer is a number and the answer you obtained is not given**). Fill your answer for each question on the Scantron answer sheet.

5. Please check that you have the following:

   - The signature sheet is attached to the front of this test-booklet.
   - Standard normal distribution table (Z table) is attached to the end of this test-booklet.
   - The test-booklet has 20 pages including this page (but not including the signature sheet page, and the Z table).
   - A Scantron sheet is provided separately with this test-booklet.

Good Luck ☺

Test Version: **B: Yellow Scantron**

1. Which of the following variables is NOT a quantitative variable?

   (a) Number of children in family
   (b) Time (in minutes) spent in the previous month browsing the World Wide Web
   (c) Distance (in kilometers) commuted to UTSC
   (d) Employment status ("Part-time", "Full-time", "Unemployed")
   (e) Number of residents in the province of Ontario

2. An item on the Multiculturalism Attitudes Survey (1991) asked Canadian participants to indicate their agreement with the statement, "Ensuring that Organizations and Institutions reflect and respect the cultural and racial diversity of Canadians" on a 7-point Likert scale (1 = "Strongly Disagree" to 4 = "Neither Disagree nor Agree" to 7 = "Strongly Agree").

   What type of a variable is this?

   (a) A categorical variable measured on an ordinal scale.
   (b) A quantitative variable measured on an interval scale.
   (c) A categorical variable measured on a nominal scale.
   (d) A categorical variable measured on an interval scale.
   (e) A quantitative variable measured on a nominal scale.

3. A researcher plans to investigate the relationship between adult Canadians' income (measured in thousands of Canadian dollars per year) and their level of education (e.g., Some High School, High School Diploma, Some Post-secondary, Bachelor's Degree). What graphical display should the researcher choose for investigating the relationship between these two variables (income and education-level)?

   (a) Histogram
   (b) Boxplot
   (c) Bar Chart
   (d) Scatterplot
   (e) Side-by-side boxplots

4. Use this information for this question and the next question. An item on the Canadian Adult Literacy and Life Skills (ALLS) survey asked 45,900 participants to indicate their agreement with the statement "I have math skills to do main jobs well". Below is the frequency distribution table for the responses to the ALLS statement.

Frequency table results for Responses:
Count = 45900

Response	Frequency	Percent of Total
1: Strongly Disagree	397	0.9
2: Disagree	1571	3.4
3: Agree	16761	36.5
4: Strongly Agree	27171	59.2

What is the mode response to the ALLS statement?

(a) 1
(b) 59.2
(c) 27171
(d) 45900
(e) 4
Solution: Response 4 "as data value": Strongly Agree is the most frequent (59.2%) response.

5. Refer to the information in questions 4. What is the Q1-value (first quartile, also referred to as lower quartile value) for the distribution of responses to the ALLS statement?

(a) 1.5
(b) 3
(c) 36.5
(d) 11475
(e) 16761
Solution: Q1-value is the 25^{th} percentile: 25% of the observations are below it and 25% are above it. Add percentages to get 25%: 0.9% + 3.4% = 4.3% < 25%. Need 25% - 4.3% = 20.7% to get to the Q1-value; the next bin has 36.5% data values. Thus, Q1 is 3.

6. Use this information for this question and the next question. The Canadian Alcohol and Drug Use Monitoring Survey (CADUMS, 2012) asked 11,069 participants to indicate their level of mental health on a 5-point Likert scale (1 = "Poor", to 3 = "Good", to 5 = "Excellent"). Below is the frequency distribution table of the responses.

Frequency table results for Responses:
Count = 11069

Response	Relative Frequency
1: Poor	131
2: Fair	460
3: Good	2305
4: Very Good	3958
5: Excellent	4215

What is the mean response to this CADUMS report?

(a) 1.50

(b) 2.00

(c) 3.00

(d) 4.05

(e) 2213.8

Solution:

$$\text{mean} = \frac{(1\ x\ 131)+(2\ x\ 460)+(3\ x\ 2305)+(4\ x\ 3958)+(5\ x\ 4215)}{11069} = \frac{44873}{11069} = 4.05$$

7. Refer to the information in question 6. If 4000 respondents change their responses from "5 = Excellent" to "4 = Very Good", how would this impact the mean and the median of the original data (the original data that is described in question 6)?

(a) The mean and median will remain the same.

(b) The mean and the median are both increased.

(c) The mean and the median are both decreased.

(d) The mean will remain the same, but the median is decreased.

(e) The median will remain the same, but the mean is decreased.

Response	Original Frequency	Frequency
1	131	131
2	460	460
3	2305	2305
4	3958	3958 + 4000 = 7958
5	4215	4215 – 4000 = 215
Total	11069	11069

Position of Median: $\frac{11069+1}{2}$ = 5535th data-value.

- Original data: 131(1's) + 460(2's) + 2305(3's) = 2896. This is not enough observations so take 5535 – 2896 = 2639 data-values from next bins(s)). The next bin has 3958 values, which is enough to contribute (the 2639 data-values) to reach the median. So, median is in this bin: Median is 4.
- New data: 131(1's) + 460(2's) + 2305(3's) = 2896. This is not enough observations so take 5535 – 2896 = 2639 data-values from next bins(s)). The next bin has 7958 values, which is enough to contribute (the 2639 data-values) to reach the median. So, median is in this bin: Median is 4.

Comparing Means:

- Original Mean = $\frac{(1\ x\ 131)+(2x460)+(3x2305)+(4x3958)+(5x4215)}{11069} = \frac{44873}{11069} = 4.05$
- New Mean = $\frac{(1\ x\ 131)+(2x460)+(3x2305)+(4x7958)+(5x215)}{11069} = \frac{40873}{11069} = 3.69$

Note that the sum of the observations shifted by 4000 from original data to new data.

So, the median remained the same, but the mean is decreased.

8. Use this information for this question and the next two questions. A survey asked 1443 married respondents, "How successful do you feel in your family life?". The responses to the question had mean 3.56 and standard deviation 0.86. Below is the frequency distribution for the responses to the question.

Frequency table results for Responses:
Count = 1443

Response	Frequency	Percent of Total
1: Not at all successful	25	1.7
2: Not very successful	95	6.6
3: Somewhat successful	550	38.1
4: Very successful	588	40.7
5: Completely successful	185	12.8

What *percentage* of responses are within 1 standard deviation of the mean?

(a) 68 percent

(b) 78.8 percent

(c) 85.4 percent

(d) 98.2 percent

(e) 91.6 percent

Solution: First we need to obtain the values for the interval: mean ± 1SD.
This interval is (mean – 1SD, mean + 1SD) = (3.56 – 1(0.86), 3.56 + 1(0.86)) = (2.7, 4.42).
Based on the data-values in this distribution, we take percentages between 3 and 4 responses.
That is, above 2 (we do not include 2) and below 5 (we do not include 5).
The percentage of data-values between 3 and 4 (as responses) is: 38.1% + 40.7% = 78.8%

9. Refer to the information in question 8. Interpret the value of standard deviation within the context of this study. Select the most appropriate interpretation from the following options.

(a) We expect that respondents' feeling of success regarding family life is 0.86.

(b) We expect that respondents' feeling of success regarding family life differ by 0.86.

(c) We expect that respondents' feeling of success regarding family life to differ from the mean by 0.86.

(d) We expect that respondents' feeling of success regarding family life deviate from mean by 0.86.

(e) We expect that respondents' feeling of success regarding family life to differ from the mean by 0.86, on average.

10. Refer to the information in question 8. Below is the same frequency distribution table as the one displayed in question 8. Recall that the responses to the question had mean 3.56 and standard deviation 0.86.

Frequency table results for Responses:
Count = 1443

Response	Frequency	Percent of Total
1: Not at all successful	25	1.7
2: Not very successful	95	6.6
3: Somewhat successful	550	38.1
4: Very successful	588	40.7
5: Completely successful	185	12.8

A standardized value (a Z-score) for a respondent's answer to the question was -1.81. What was the respondent's response to the question?

(a) 1

(b) 2

(c) 3

(d) 4

(e) 5

Solution: Recall that Z-score = $\frac{Observation - Mean}{Standard\ Deviation}$

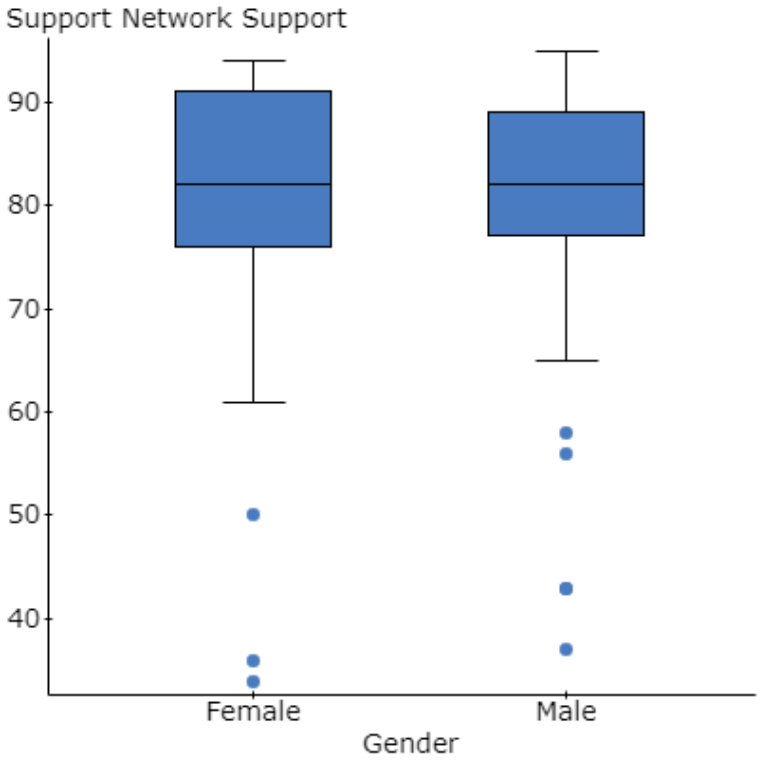
- Observation = Mean + (Z x SD)

  Observation = 3.56 + (-1.81 x 0.86)

  = 3.56 – 1.5566

  = 2.0034 ≅ 2

11. The Organisation for Economic Co-operation and Development (OECD, 2015) reported the percentage of people, aged 15 and over, who indicated that they have relatives or friends that they can count on to help them whenever they need them. This variable was denoted as support network (measured in percentage). Below are side-by-side boxplots for the percentages of male and female participants' who reported having social support network in the 35 OECD countries.



Based on the above side-by-side boxplots, which of the following statements is **NOT** correct?

(a) About 50 percent of OECD countries had approximately between 75 to 90 percent of their males and females' participants who reported having a support network.

(b) About 25 percent of OECD countries had approximately 90 percent or more of their males and females' participants who reported having a support network.

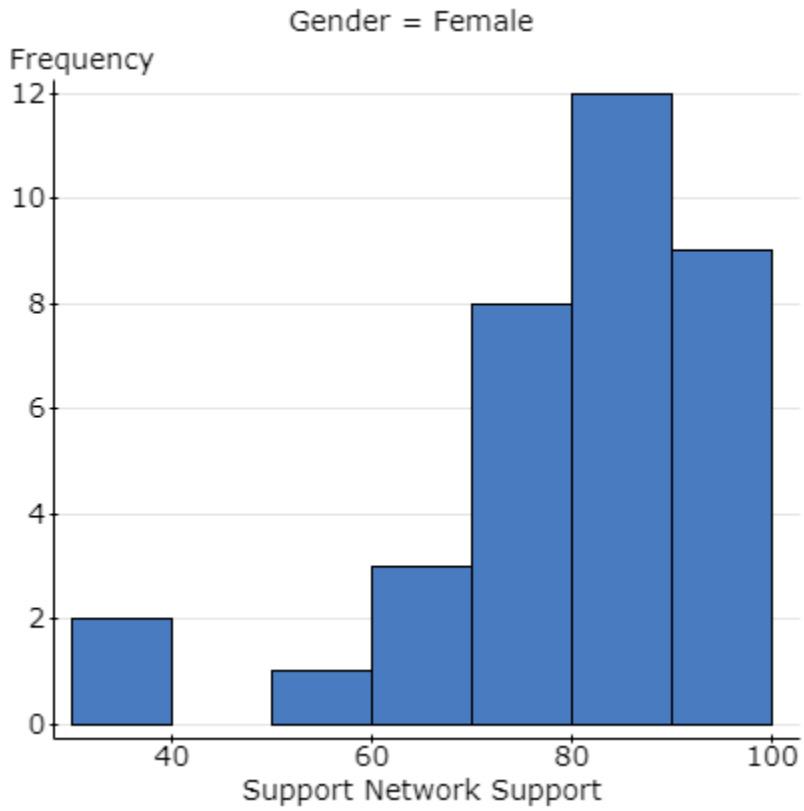
(c) There are low outliers in both distributions of support network for males and females.

(d) There are high outliers in both distributions of support network for males and females.

(e) There are no outliers in both distributions of support network for males and females.

Solution: Both options "d" and "e" are acceptable answers.

12. Use this information for this question and the next question. Below is the histogram for the distribution of females' percentages of social network support in the 35 OCED countries. Note: In order to avoid complications, you may assume that no OECD country had social network support percentage exactly at a class boundary.



What *percentage* of the 35 OECD countries had at most 80 percent of their females reported having a social network support?

(a) 23 percent

(b) 60 percent

(c) 8 percent

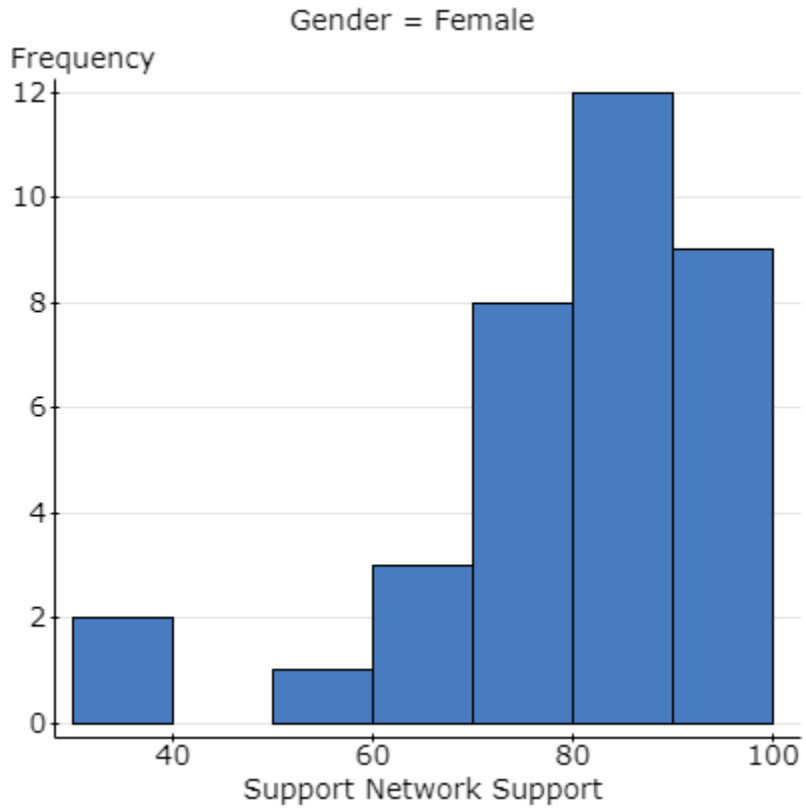
(d) 14 percent

(e) 40 percent

Solution: At most 80% means 80% or less; Add the frequencies associated with at most 80%:

$\frac{2+1+3+8}{35} = \frac{14}{35} = 0.40$ in proportion; 0.40 x 100 = 40%

13. Refer to the information in question 12. Below is the same histogram for the distribution of females' percentages of social network support in the 35 OCED countries.



What can you tell about the mean and median social network support percentage?

(a) The mean and the median values are exactly the same.

(b) The mean value is less than the median value.

(c) The mean value is more than the median value.

(d) The mean is twice as big as the median value.

(e) The median value is less than the mean value.

Solution: Mean < Median, since the distribution is skewed to the left.

14. Use this information for this question and the next question. A group of rats running a straight alley maze required the following number of trials to perform a predetermined criterion. The frequency distribution follows.

Trials to reach criterion	18	19	20	21	22	23	28
Number of rats (frequency)	1	0	4	3	3	3	1

What are the five-number summaries? Please use the method described in this course.

(a) min = 18, Q1 = 19.5, median = 21, Q3 = 22.5, max = 28

(b) min = 18, Q1 = 19, median = 21, Q3 = 23, max = 28

(c) min = 18, Q1 = 20, median = 21.5, Q3 = 23, max = 28

(d) min = 18, Q1 = 20, median = 21, Q3 = 22.5, max = 28

(e) min = 18, Q1 = 20, median = 21, Q3 = 23, max = 28

Data values: 18, 20, 20, 20, 20, 21, 21, 21, 22, 22, 22, 23, 23, 23, 28

Solution: Both versions: “d” and “e” are correct.

❖ Option “d” is based on the lecture notes.

- Min = 18
- The position of the median is (n + 1)/2 = (15 + 1)/2 = 8th ordered value: 21
- The position of the O1-value is (include median in the lower half of data): average of the 4th & 5th ordered values: (20 + 20)/2 = 20
- The position of the O3-value is (include median in the upper half of data): average of the 4th & 5th ordered values: (22 + 23)/2 = 22.5
- Max = 28

❖ Option “e” is based on the textbook.

Summary statistics:

Column	Min	Q1	Median	Q3	Max
Trials	18	20	21	23	28

15. Refer to the information in question 14. Suppose we construct a boxplot for the data described in question 14. The upper whisker in the boxplot would extend to what data value?

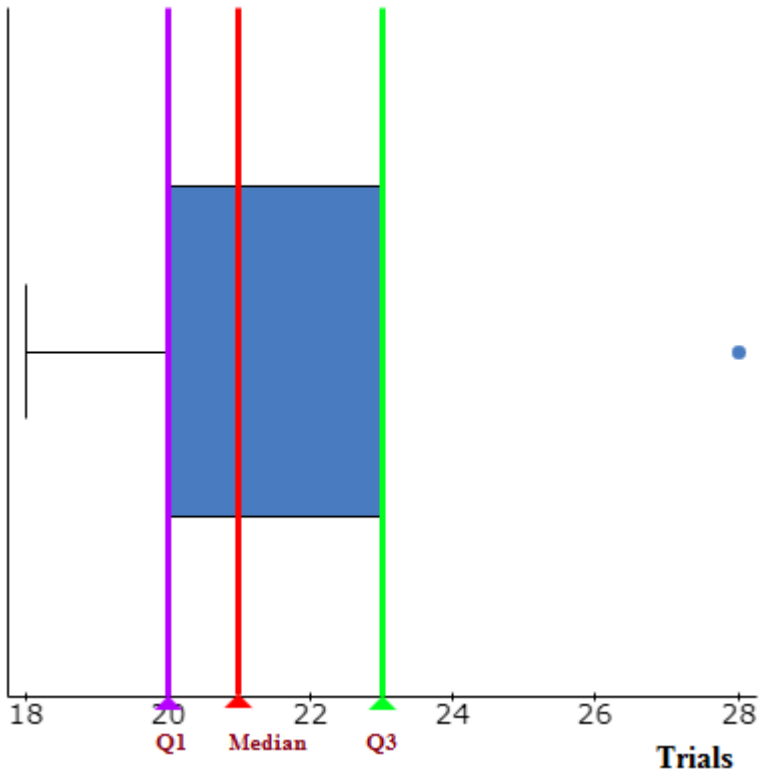
(a) 23

(b) 26.25

(c) 22

(d) 28

(e) 25.25

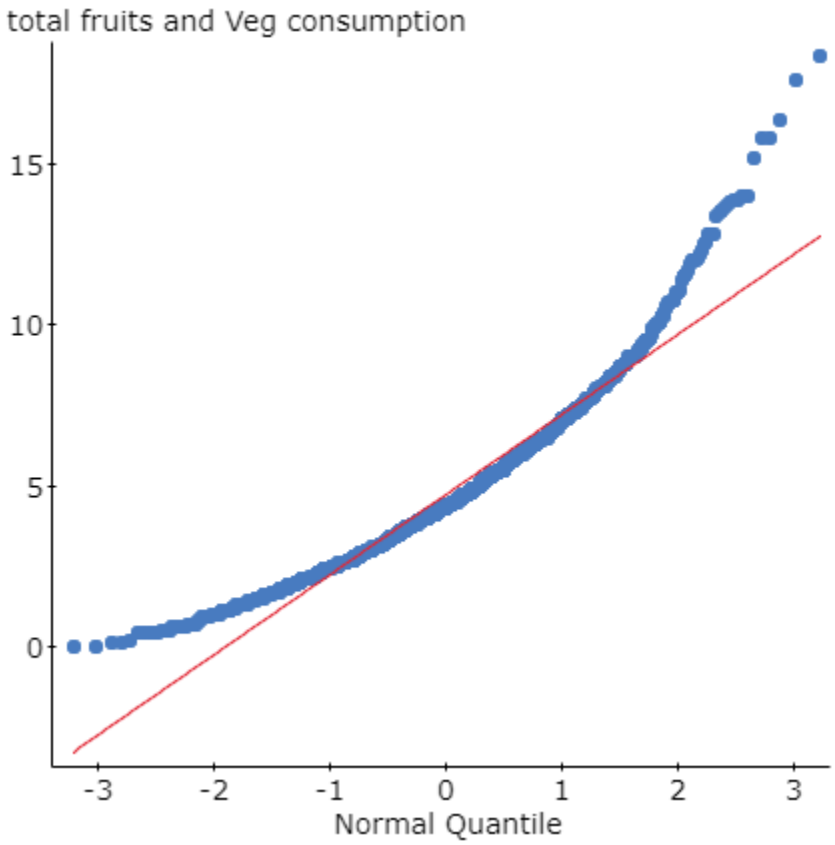


IQR = Q3 – Q1 = 22.5 – 20 = 2.5

Rule = 1.5 x IQR = 1.5 x 2.5 = 3.75

- Lower Inner Fence: Q1 – R = 20 – 3.75 = 16.25
  Draw Lower Whisker to: 18 data-value
- Upper Inner Fence: Q3 + R = 22.5 – 3.75 = 26.25
  Draw Lower Whisker to: 23 data-value
- There is one "high" outlier: 28

16. The Canadian Community Health Survey (CCHS, 2012) reported respondents' total fruits and vegetable daily consumptions. Below is the normal quantile plot of the reported information.



Based on this information what can we conclude about the distribution of total fruits and vegetable daily consumptions?

(a) The distribution is approximately normal.

(b) The distribution is left skewed.

(c) The distribution is right skewed.

(d) The distribution is symmetric but not normal.

(e) The distribution is not symmetric.

17. Use this information for this question and the next two questions. The length of pregnancy is assumed to be normally distributed with a mean of 266 days and standard deviation of 16 days. Approximately, what *percentage* of pregnancies last over 285 days?

(a) 12 percent

(b) 1.19 percent

(c) 0.12 percent

(d) 0.019 percent

(e) 88 percent

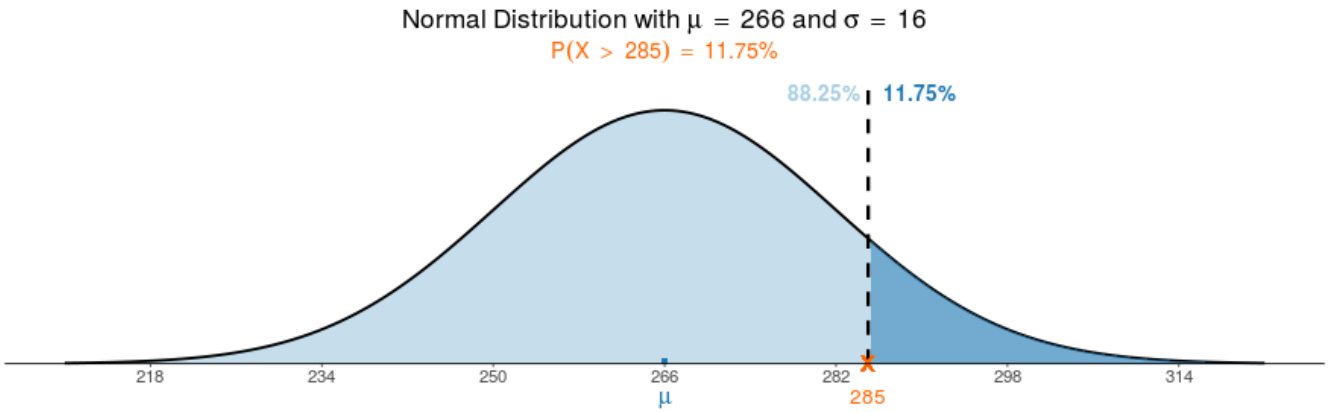
Solution:

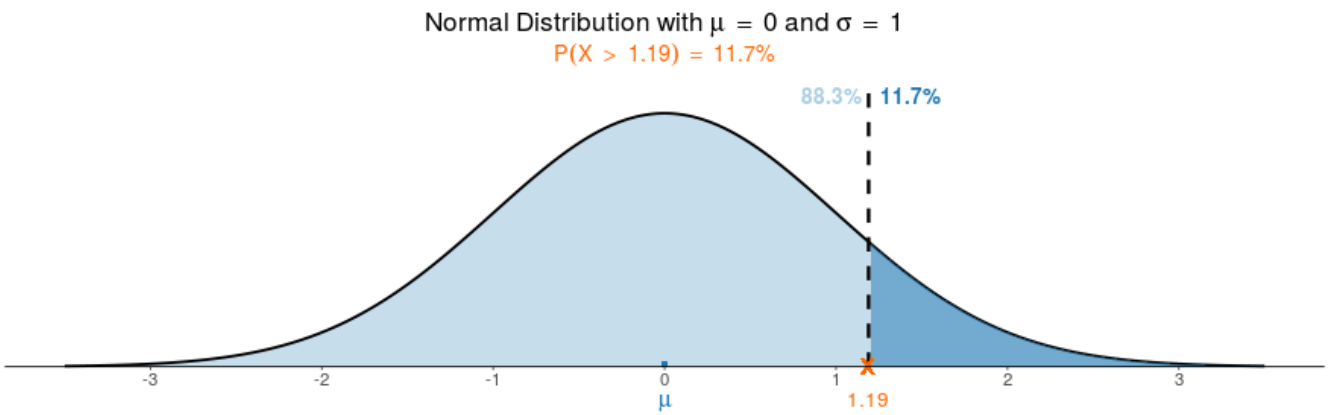
We first need to standardize the value 285.

$$Z = \frac{Obervation - Mean}{SD} = \frac{285-266}{16} = 1.19$$

By symmetry, area above Z of 1.19 is equivalent to area below Z of -1.19. From the Z-table, the area below Z of -1.19 is 0.1170 (in proportion): 0.1170 x 100 = 11.70% (approx. 12%).

Or: 1 – Area above Z of 1.19 = 1 – 0.883 = 0.117 (in proportion); 11.7%





18. Refer to the information in question 17. Using the Empirical Rule, 68-95-99.7 rule, approximately, what *percentage* of pregnancies last under 218 days?

(a) 3 percent

(b) 0.15 percent

(c) 0.0015 percent

(d) 0.003 percent

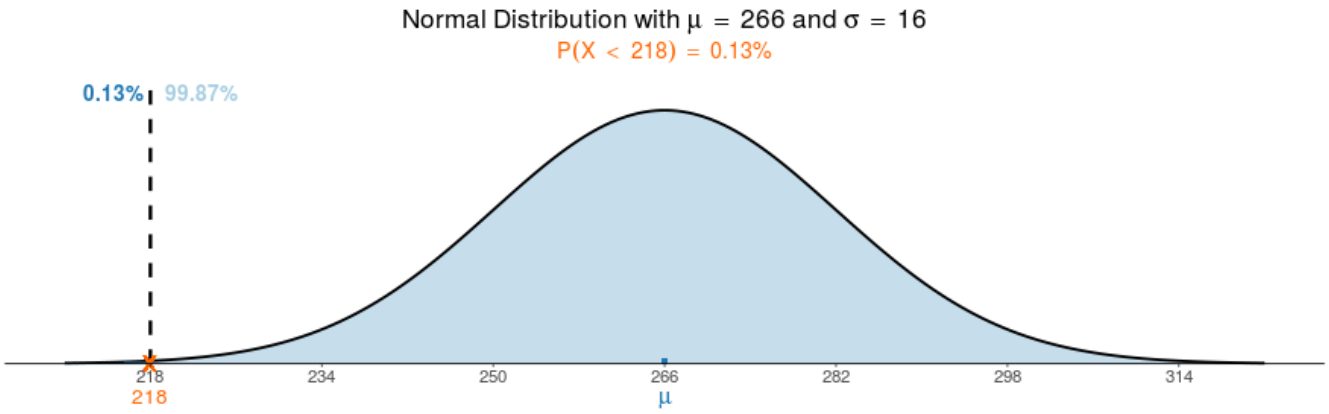
(e) 0.03 percent

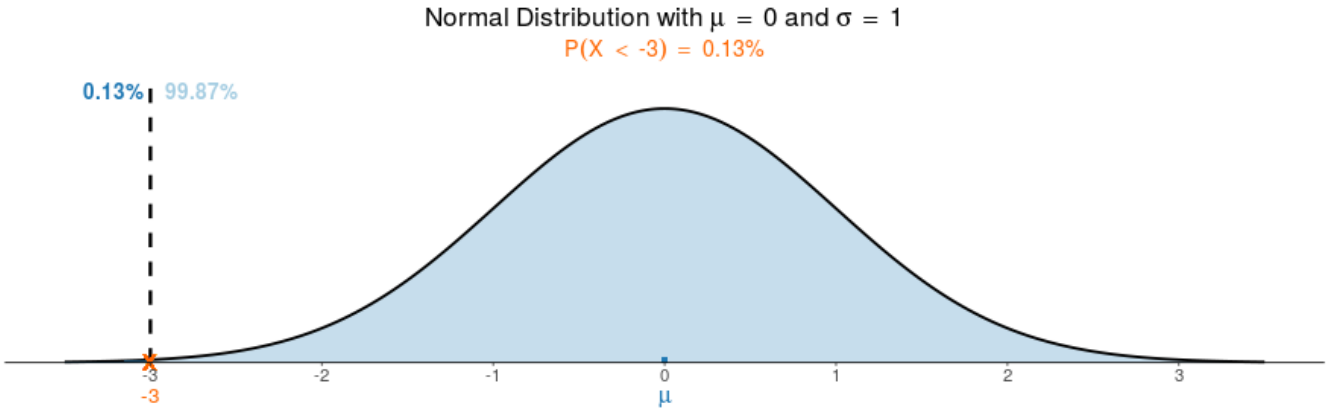
Solution:

We first need to standardize the value 218.

$$Z = \frac{Obervation - Mean}{SD} = \frac{218-266}{16} = -3$$

According to Empirical Rule, 99.7% of observations are within 3 SD of the Mean. Furthermore, approximately 0.30% of the observations are outside the 3 SD of the mean. This means that, approximately 0.30%/2 = 0.15% of the pregnancies last under 218 days.





19. Refer to the information in question 17. Approximately, what is the IQR of the length of pregnancies?

(a) 16

(b)1.33

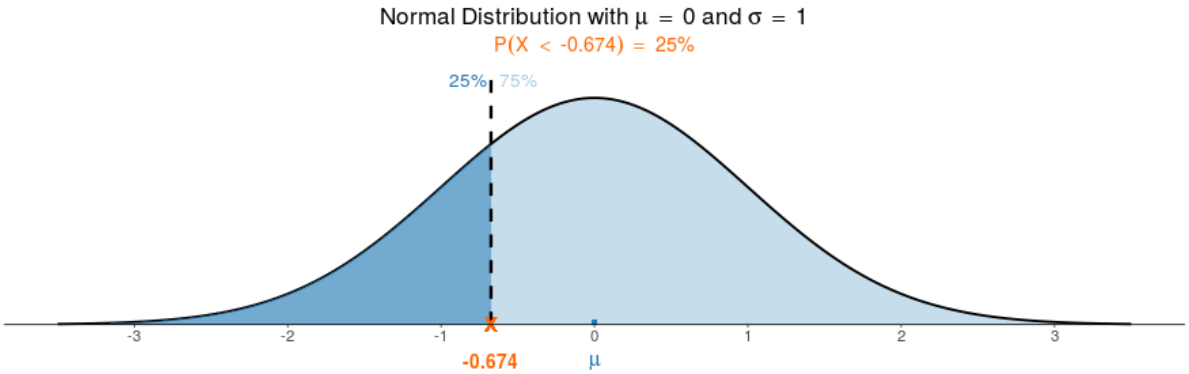
(c) 21

(d) 11

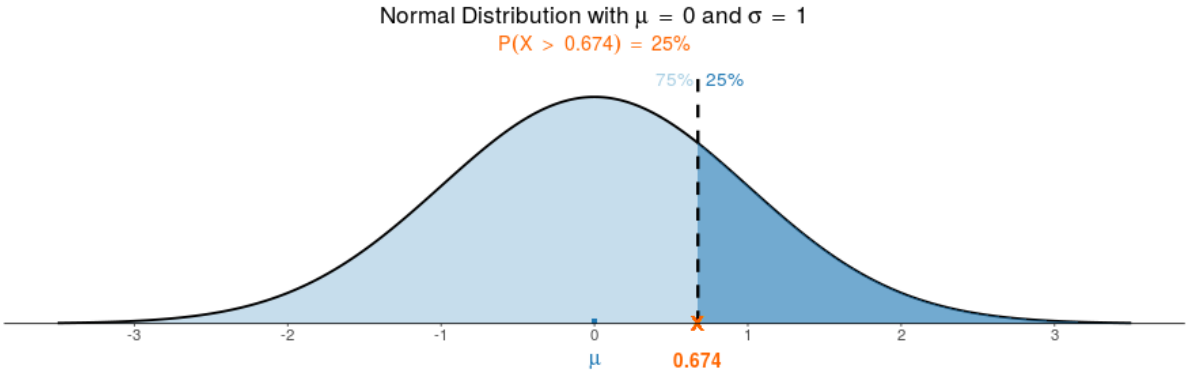
(e) 0.5

Method #1:

- IQR is the middle 50% of the distribution.
  0.25 values are below Q1, and 0.25 value are above Q3; that 0.50 values are outside IQR.
- In order to find Q1-value, find its Z-score that gives 0.25 proportion below it (area below Q1):
  In the Z-table, the Z-score with the closest 0.25 proportion (0.2514) below it, is: -0.67



- Use the Z-score formula backward:
  Observation (Q1-value) = mean + (Z x SD)
  = 266 + (-0.67 x 16) = 266 – 10.72 = 255.28
  Similarly, Observation (Q3-value) = mean + (Z x SD)
  "only the sign of Z changes since Q3 is above the mean)
  = 266 + (+0.67 x 16) = 266 + 10.72 = 276.72



- So, IQR = Q3 – Q1 = 276.72 – 255.28 = 21.44. Thus, IQR is approx.21.

Method #2 - Another way to find IQR:

Since we have a normal distribution, $\frac{IQR}{S} \cong \frac{4}{3} = 1.33$. Thus, IQR ≅ 1.33 x S = 1.33 x 16 = 21.28≅21

20. Use this information for this question and the next question. Below is a stem-and-leaf plot of salary offers (in thousands of dollars) offered to 24 MBA students.

Variable: Salary

Decimal point is 1 digit(s) to the right of the colon.
Leaf unit = 1

```
3 : 159
4 : 04589
5 : 001268899
6 : 12335
7 : 055
```

What is the shape of the data?

(a) Approximately Symmetric

(b) Symmetric and bell-shaped

(c) Skewed to the right

(d) Skewed to the left

(e) Bimodal

21. Refer to the information in question 20 (refer to the stem-and-leaf display above). What is the most appropriate measure of spread for the data?

(a) Range

(b) IQR

(c) Z-score

(d) Variance

(e) Standard Deviation

Solution: Since we have a bell-shaped, and symmetric distribution.

22. Data were collected on annual rain fall (y), and maximum daily temperature (x) in 11 sites in Mongolia (Central Asia). The scatterplot of y against x was approximately linear. The estimated correlation coefficient was r = -0.92. Which of the following statement(s) correctly describes the correlation between these two variables in the context of the study?

I As maximum daily temperature increases, the annual rain falls decrease.
II Higher maximum daily temperatures are associated with lower annual rain falls.
III As maximum daily temperature increases, the annual rain falls tend to decrease.

(a) I only
(b) I and II only
(c) II and III only
(d) I, II, and III
(e) None of the I, II, and III statements is correct.

23. Agricultural scientists are working on developing an improved variety of Roma tomatoes. Marking research indicated that customers are likely to bypass Romas that weigh less than 70 grams. The current variety of Roma plants produce fruit that average 74 grams, but 11% of the tomatoes are too small. It is reasonable to assume that a normal model applies. What is the standard deviation of weights of Romas now being grown?

(a) 3.25
(b) -4
(c) 1.23
(d) 11
(e) 1

Solution: 11% is equivalent to 0.11 proportion in the lower part of the normal distribution (below the mean). We need to find the associated Z-score (for the value of 70 weight) such that its area below that Z-score is about 0.11. However, based on our Z-table, the closest proportion to 0.11 is 0.1093 with Z-score of -1.23. We can, now, substitute numbers in the formula below to find the value of SD:

$$Z = \frac{Obervation - Mean}{SD} \qquad -1.23 = \frac{70 - 74}{SD} \rightarrow \text{-1.23 SD = -4} \rightarrow \text{SD = 3.25}$$

24. The contingency table below shows the conditional distribution of cannabis use during the past 12 months on the sex of respondents. This information was collected by the Canadians Alcohol and Drug Use Monitoring Survey (CADUMS) in 2012.

Contingency table results:

Rows: Sex
Columns: Cannabis Use

Cell format
Count
(Row percent)
(Column percent)
(Percent of total)

	No/Never	Yes/Used	Total
Male	3821 (87.6%) (37.8%) (34.6%)	541 (12.4%) (58.6%) (4.9%)	4362 (100%) (39.5%) (39.5%)
Female	6300 (94.3%) (62.2%) (57.0%)	382 (5.7%) (41.4%) (3.5%)	6682 (100%) (60.5%) (60.5%)
Total	10121 (91.6%) (100%) (91.6%)	923 (8.4%) (100%) (8.4%)	11044 (100%) (100%) (100%)

What can we tell about the apparent association, if at all, between Canadian respondents' cannabis usage during the past 12 months and their sex?

(a) There appears to be no association between cannabis usage during the past 12 months and sex of the respondents.

(b) The percentages of males and females' cannabis usage during the past 12 months appear to be about the same.

(c) Males are more likely (12.4%) than the females (5.7%) to use cannabis during the past 12 months.

(d) Males are more likely (58.6%) than the females (41.4%) to use cannabis during the past 12 months.

(e) Males are more likely (4.9%) than the females (3.5%) to use cannabis during the past 12 months.

25. Use this information for this question and the next three questions. Below is the relative frequency distribution table for responses to a Canadian Access and Support to Education and Training Survey (ASET, 2008) statement: "Learning gives you more self-confidence".

Frequency table results for Responses:
Count = 7921

Response	Relative Frequency
1: Strongly Disagree	0.002
2: Somewhat Disagree	0.002
3: Neither Agree nor Disagree	0.001
4: Somewhat Agree	0.092
5: Strongly Agree	0.903

What is the mean response to the ASET statement?

(a) 0

(b) 4.90

(c) 3

(d) 0.98

(e) 0.20

Solution: mean = (1 x 0.002) + (2 x 0.002) + (3 x 0.001) + (4 x 0.903) = 4.90

26. Refer to the information in question 25. What is the median response to the ASET statement?

(a) 1

(b) 3

(c) 0.50

(d) 50

(e) 5

Solution: Median is the 50^{th} percentile: 0.50 of data values are below the median and 0.50 are above it. We add proportions (relative frequencies) till we get 0.50:
0.002 + 0.002 + 0.001 + 0.092 = 0.097 < 0.50. We need 0.50 – 0.097 = 0.403 proportions to get to the median; We can take from the next bin since it has 0.903 proportion. Thus, median is 5.

27. Refer to the information in question 25. Below is the same relative frequency distribution table as the one presented in question 25.

Frequency table results for Responses:
Count = 7921

Response	Relative Frequency
1: Strongly Disagree	0.002
2: Somewhat Disagree	0.002
3: Neither Agree nor Disagree	0.001
4: Somewhat Agree	0.092
5: Strongly Agree	0.903

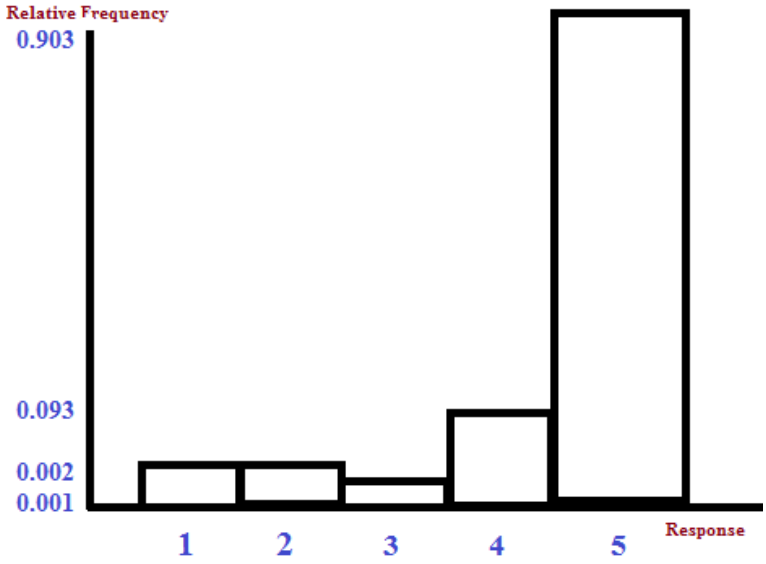
What proportion of respondents either "somewhat agreed" or "strongly agreed" with the ASET statement?

(a) 0.092
(b) 0.903
(c) 0.995
(d) 0
(e) 0.498

Solution: Proportion (Somewhat Agree) + Proportion (Strongly Agree) = 0.092 + 0.903 = 0.995

28. Refer to the information in question 25. What is the shape of the distribution of responses regarding the ASET statement?

(a) Symmetric
(b) Approximately symmetric, bell-shaped
(c) Skewed to the right
(d) Skewed to the left
(e) Bimodal



29. Use this information for this question and the next two questions. The Canadian Alcohol and Drug Use Monitoring Survey (CADUMS, 2012) asked a random sample of 11,047 adult Canadians whether they used cannabis in their lifetime. The result of this survey question revealed that 60.4% of the randomly selected never used Cannabis in their lifetime.

Based on the above information, identify the variable of interest and its scale of measurement.

(a) Lifetime Cannabis use: Categorical variable measured on a nominal scale
(b) Lifetime Cannabis use: Quantitative variable measured on an interval scale
(c) Lifetime Cannabis use: Categorical variable measured on an ordinal scale
(d) CADMUS, 2012
(e) Adult Canadians who only reported using cannabis in their lifetime: Quantitative variable

30. Refer to the information in in question 29. Identify the population of interest.

(a) Adult Canadians
(b) A randomly selected adult Canadians
(c) The 11,047 randomly selected adult Canadians
(d) All adult Canadians
(e) Adult Canadians who used cannabis in their lifetime

31. Refer to the information in in question 29. Identify the sample for this study.

(a) Adult Canadians
(b) A randomly selected adult Canadians
(c) The 11,047 randomly selected adult Canadians
(d) All adult Canadians
(e) Adult Canadians who used cannabis in their lifetime

32. Use this information for this question and the next three questions. The Canadian Alcohol and Drug Use Monitoring Survey (2012) asked adult Canadians to indicate their lifetime alcohol drinking pattern. Below is the contingency table of responses by the sex of the participants.

Sex	Lifetime Drinking Pattern						
	"Lifetime Abstainer"	"Former Drinker"	"Light/ Infrequent"	"Light/ Frequent"	"Heavy/ Infrequent"	"Heavy/ Frequent"	Total
Male	210	603	1340	1655	232	289	4329
Female	578	1234	2901	1671	173	87	6644
Total	788	1837	4241	3326	405	376	10973

Suppose the researchers were interested to investigate the association between respondents' lifetime alcohol drinking pattern and their sex. What is the response variable in the researchers' investigation?

(a) Sex of the respondents
(b) Respondents' lifetime alcohol pattern indication
(c) The 10,973 cases
(d) Both variables, the sex of the respondents and their lifetime alcohol pattern indication
(e) The Canadian Alcohol and Drug Use Monitoring Survey (2012)

33. Refer to the information in question 32. Among *female* Canadians, what is the conditional percentage of those who were a "light/frequent" drinker?

(a) 25.2 percent
(b) 30.3 percent
(c) 60.5 percent
(d) 51.8 percent
(e) 15.2 percent

Solution: $\frac{1671}{6644} \cong 0.252$ (in proportion): 0.252 x 100 = 25.2%

34. Refer to the information in question 32. Below is the same contingency table as the one presented in question 32.

Sex	Lifetime Drinking Pattern						
	"Lifetime Abstainer"	"Former Drinker"	"Light/ Infrequent"	"Light/ Frequent"	"Heavy/ Infrequent"	"Heavy/ Frequent"	Total
Male	210	603	1340	1655	232	289	4329
Female	578	1234	2901	1671	173	87	6644
Total	788	1837	4241	3326	405	376	10973

Among those who were either "heavy/infrequent" or "heavy/frequent" lifetime drinker, what is the conditional percentage of those who were *male*?

(a) 12.1 percent

(b) 78.1 percent

(c) 4.7 percent

(d) 66.7 percent

(e) 7.1 percent

Solution: $\frac{232+289}{405+376} = \frac{521}{781} \cong 0.667$ (in proportion): 0.667 x 100 = 66.7%

35. Refer to the information in question 32. The same contingency table is displayed above. What percentage of *female* respondents were neither a "heavy/infrequent" nor a "heavy/frequent" lifetime drinker?

(a) 33.3 percent

(b) 3.9 percent

(c) 96.1 percent

(d) 2.3 percent

(e) 66.7 percent

Solution: $\frac{578+1234+2901+1671}{6644} = \frac{6384}{6644} \cong 0.961$(in proportion): 0.961 x 100 = 96.1%

Or: $1 - (\frac{173+87}{6644}) \cong 1 - 0.0391 = 0.9609$ (in proportion): 0.9609 x 100 = 96.1%

~The End~