University of Toronto Scarborough
STAB22 Final Examination

December 2011

For this examination, you are allowed two handwritten letter-sized sheets of notes (both sides) prepared by you, a non-programmable, non-communicating calculator, and writing implements.

This question paper has 22 numbered pages, with statistical tables at the back. Before you start, check to see that you have all the pages. You should also have a Scantron sheet on which to enter your answers. If any of this is missing, speak to an invigilator.

This examination is multiple choice. Each question has equal weight, and there is no penalty for guessing. To ensure that you receive credit for your work on the exam, fill in the bubbles on the Scantron sheet for your correct student number (under "Identification"), your last name, and as much of your first name as fits.

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.)

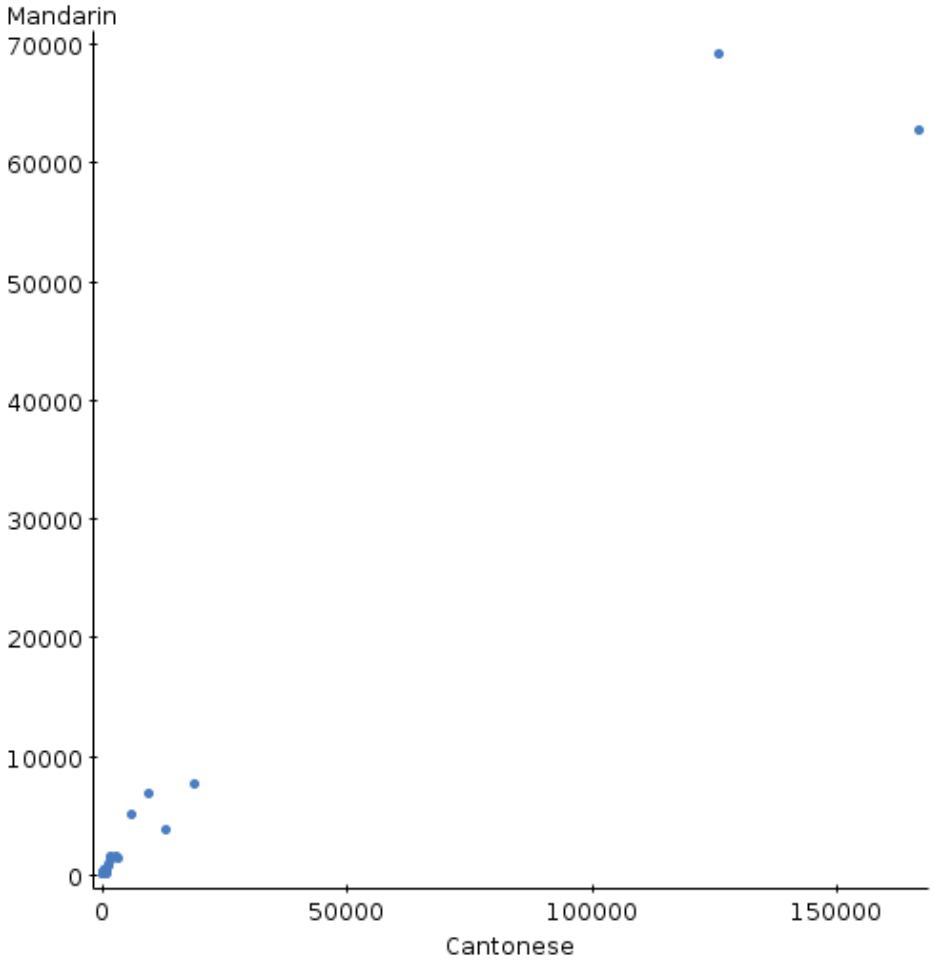
If you need paper for rough work, use the back of the sheets of this question paper.

Before you begin, two more things:

- Check that the colour printed on your Scantron sheet matches the colour of your question paper. If it does not, get a new Scantron from an invigilator.
- Complete the signature sheet, but *sign it only when the invigilator collects it.* The signature sheet shows that you were present at the exam.

At the end of the exam, you *must* hand in your Scantron sheet (or you will receive a mark of zero for the examination). You will be graded *only* on what appears on the Scantron sheet. You may take away the question paper after the exam, but whether you do or not, anything written on the question paper will *not* be considered in your grade.

1. For each of a number of cities across Canada, the number of people with Cantonese as a mother tongue and the number of people with Mandarin as a mother tongue were recorded. A scatterplot of the data is shown below.



The correlation between the number of Cantonese speakers in a city and the number of Mandarin speakers is 0.980. Do you think this correlation is a reasonable summary of the relationship? Why, or why not?

(a) If the two largest cities (Toronto and Vancouver) were removed, the correlation would be even higher.

(b) No, because there are outliers and/or influential points.

(c) Yes, because the relationship is more or less linear.

(d) No, because the relationship is clearly curved.

(e) Yes, because there are no influential points.

2. In a game you toss a fair coin once. You win \$20 if a head comes up and you lose \$10 if a tail comes up. Let the random variable X denote the amount of money you win. (Losing \$10 means winning $-$\$10.) What is the standard deviation of X, in dollars?

   (a) 15
   (b) 12.50
   (c) 17.50
   (d) 10
   (e) 5

3. A *98%* confidence interval was calculated for a population mean, based on a simple random sample of size 25. The population standard deviation was not known. The interval was from 54 to 66. What must the standard deviation of this sample have been?

   (a) 6
   (b) 12
   (c) 10
   (d) 14
   (e) 5

4. Researchers speculate that drivers who do not wear a seatbelt are more likely to speed than drivers who do wear one. A random sample of 20 drivers had their speed measured at a certain point, and were then observed to see whether they were wearing a seatbelt. Summary statistics for the drivers are shown below:

```
Summary statistics for speed:
Group by: seatbelt
seatbelt n  Mean      Std. Dev.
n        8  72.5      8.815571
y        12 65.333336 7.487363
```

A test was carried out to test the null hypothesis that the mean speed was equal for the drivers wearing and not wearing seatbelts, against the alternative that drivers not wearing seatbelts travel faster on average. The test statistic was -1.889 (seatbelt mean minus non-seatbelt mean). Using the methods learned in class, what can you say about the P-value for this test statistic?

(a) 0.0588

(b) larger than 0.10

(c) 0.0294

(d) between 0.05 and 0.10

(e) between 0.025 and 0.05

5. How does a systematic sample differ from a simple random sample?

(a) If parts of the population differ from each other, a systematic sample will give more accurate results.

(b) A systematic sample is less convenient to take than a simple random sample.

(c) Knowing about one item in the sample tells you about which other items will be in the sample.

(d) Each item in the population has the same chance to be in the sample.

6. A simple random sample of 50 measurements is taken from a population whose standard deviation is known to be 10. Assume that the sample is large enough for the Central Limit Theorem to apply. We are testing a null hypothesis that the population mean is 60 against the alternative that it is not equal to 60, using a significance level of $\alpha = 0.05$. The sample mean is 63. What do you conclude?

   (a) conclude that the population mean could be 60 because the P-value is less than 0.05

   (b) conclude that the population mean is not 60 because the P-value is between 0.025 and 0.05

   (c) conclude that the population mean is not 60 because the P-value is less than 0.01

   (d) conclude that the population mean is not 60 because the P-value is between 0.01 and 0.025

   (e) conclude that the population mean could be 60 because the P-value is greater than 0.05

7. The distribution of the grades in an exam has a Normal distribution. Approximately 2.5% of the students scored 50 or below. 16% of them scored 85 and above. What can you say about the mean grade of all the students?

   (a) between 70 and 75

   (b) greater than 75

   (c) between 65 and 70

   (d) less than 60

   (e) between 60 and 65

8. The probability that a randomly chosen calculus student passes a certain calculus course is 0.70. Use this information for this question and the next two questions.

   If 5 students are sampled at random, what is the probabilty that exactly 4 of them pass the course?

   (a) 0.028

   (b) 0.640

   (c) 0.335

   (d) impossible to determine from the information in this course

   (e) 0.360

9. In Question 8, the probability was 0.7 that a randomly chosen calculus student would pass a certain calculus course. Under these circumstances, suppose a simple random sample of 15 calculus students is taken. What is the probability that 13 or more of these students pass the course?

   (a) 0.079

   (b) 0.092

   (c) 0.000

   (d) impossible to determine from the information in this course

   (e) 0.127

10. In Question 8, the probability was 0.7 that a randomly chosen calculus student would pass a certain calculus course. Under these circumstances, suppose a simple random sample of 150 calculus students is taken. What is the probability that 115 or more of these students pass the course? (You may assume that the total number of students who take this course is over 2000.)

    (a) 0.50

    (b) impossible to determine from the information in this course

    (c) 0.21

    (d) 0.04

    (e) less than 0.01

11. A simple random sample of 64 men has a sample mean foot length of 27.5 cm. Assuming that the standard deviation of foot lengths for all men is 2 cm, obtain a 95% confidence interval for the mean foot length of all men. What is the **upper** limit of this interval, in centimetres?

    (a) 27.5

    (b) 27.0

    (c) 28.5

    (d) 28.0

    (e) 26.5

12. Sixty percent of all customers at an automotive service station pay for their purchases with a credit card. If a random sample of ten customers is selected, what is the probability that at least eight of them pay with a credit card?

    (a) 0.17

    (b) 0.13

    (c) 0.05

    (d) 0.83

    (e) 0.60

13. A simple random sample of students was taken from Penn State University. For each student, their sex was recorded and their pulse rate measured. The results were as follows:

Sex	n	sample mean	sample SD
Women	35	76.9	11.6
Men	57	70.42	9.95

    It is desired to see whether there is evidence of a difference in mean pulse rates between women and men. Calculate the test statistic for assessing this evidence. Assess the difference as women minus men. What do you get?

    (a) greater than 3

    (b) 2.74

    (c) 2.84

    (d) less than 2

    (e) 2.50

14. The null hypothesis $H_0 : \mu = 15$ is being tested against the alternative $H_a : \mu \neq 15$. The P-value for the test is 0.12. What can you say about the 90% and 95% confidence intervals for the population mean μ?

    (a) The 90% confidence interval contains 15 but the 95% confidence interval does not.

    (b) The P-value of the test tells us nothing about whether 15 is inside or outside a confidence interval.

    (c) The 95% confidence interval contains 15 but the 90% confidence interval does not.

    (d) Both confidence intervals contain 15.

    (e) Neither confidence interval contains 15.

15. Should cell phone use be banned by drivers? 188 people took part in a survey. The answers are classified by the gender of the respondent as below:

	Agree	Disagree
Female	68	37
Male	26	46

Use this information for this question and the two following.

What is the conditional proportion of females who agree?

(a) 0.60

(b) 0.55

(c) 0.45

(d) 0.65

(e) 0.72

16. Refer to the information given in Question 15. What is the marginal proportion of people who agree?

(a) 0.72

(b) 0.55

(c) 0.60

(d) 0.45

(e) 0.65

17. Refer to the information given in Question 15. Is there an association between gender and the response given on the survey?

(a) Yes, because the proportion of females who agree is higher than the proportion of males who agree.

(b) Association here has nothing to do with what proportions of males and females agree.

(c) No, because the proportions of females and males who agree are very similar.

(d) We need to look at a scatterplot to judge association.

(e) Yes, because the proportion of females who agree is lower than the proportion of males who agree.

18. The marks in an exam have a Normal distribution with mean 65 and standard deviation 15. A mark of 80 or above qualifies for an A grade. Adam, Bob and Cindy are three students writing this exam. Assume that their marks are independent. What is the probability that that at least one of them will get an A grade?

    (a) 0.60
    (b) 0.15
    (c) 0.50
    (d) 0.40
    (e) 0.05

19. Vehicle speeds at a certain highway location have a mean of 100 kmh and a standard deviation of 10 km/h, with a distribution that is approximately normal. A simple random sample of 25 vehicles is taken. There is a 95% chance that the mean speed of the sampled vehicles is between which of the values below, in km/h?

    (a) 98 and 102
    (b) 99.6 and 100.4
    (c) 80 and 120
    (d) 97 and 103
    (e) 96 and 104

20. A sample of 20 measurements is taken from an approximately normal population whose standard deviation is known. A 95% confidence interval for the population mean goes from 64.3 to 72.1. What is the test statistic for testing $H_0 : \mu = 75$ against $H_a : \mu \neq 75$?

    (a) 0
    (b) 3.5
    (c) -2.0
    (d) 2.0
    (e) -3.5

21. Let $\bar{x}$ be the mean of a random sample of size 4 from a Normally distributed population with mean 10 and standard deviation 20. What can we say about the sampling distribution of $\bar{x}$?

    (a) Not Normal because of the small sample size.

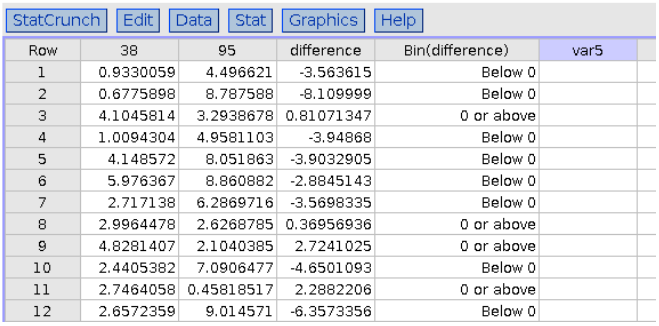
    (b) Normal with mean 10 and standard deviation 10.

    (c) Normal with mean 10 and standard deviation 20.

    (d) Only approximately Normal because of the small sample size.

    (e) Normal with mean 20 and standard deviation 20.

22. Suppose you are waiting for a bus at UTSC. The time until departure of the next number 38 bus is a (continuous) uniformly distributed random variable between 0 and 6 minutes, and the time until departure of the next number 95 bus is uniform between 0 and 10 minutes. You are travelling to Ellesmere and Markham Road, so you can catch either of these buses. You are interested in the chance that the number 38 bus leaves first. This is not possible to calculate using the methods of this course, so you do a simulation instead, using StatCrunch. The first two things you simulate are the times until departure of the next number 38 bus and the next number 95 bus. (These are in the columns labelled "38" and "95" below.) After some more calculation, the worksheet looks as shown below, in part. The "bin column" uses a single cutpoint of zero.



StatCrunch | Edit | Data | Stat | Graphics | Help

Row	38	95	difference	Bin(difference)	var5
1	0.9330059	4.496621	-3.563615	Below 0	
2	0.6775898	8.787588	-8.109999	Below 0	
3	4.1045814	3.2938678	0.81071347	0 or above	
4	1.0094304	4.9581103	-3.94868	Below 0	
5	4.148572	8.051863	-3.9032905	Below 0	
6	5.976367	8.860882	-2.8845143	Below 0	
7	2.717138	6.2869716	-3.5698335	Below 0	
8	2.9964478	2.6268785	0.36956936	0 or above	
9	4.8281407	2.1040385	2.7241025	0 or above	
10	2.4405382	7.0906477	-4.6501093	Below 0	
11	2.7464058	0.45818517	2.2882206	0 or above	
12	2.6572359	9.014571	-6.3573356	Below 0	

Some output was obtained, as below:

```
Summary statistics:
Column     Median     Min          Max      Q1          Q3
38         2.9611802  8.438758E-4 5.997734 1.4819387  4.458607
95         4.9292355  0.015312281 9.938501 2.6126897  7.5193777
difference -1.9829733 -9.351653   5.625704 -4.4648438 0.36462167

Frequency table results for Bin(difference):
Bin(difference) Frequency Relative Frequency
Below 0         701       0.701
0 or above      299       0.299
```

What is your best guess at the probability that the number 38 bus leaves first?

(a) 0.60

(b) 0.99

(c) 0.30

(d) 0.70

(e) 0.75

23. The random variable X has a Normal distribution with mean 60 and standard deviation 10. One of the following probabilities is the same as $P(40 < X \leq 48)$. Which one? (Hint: It is not necessary to use a normal table to answer this question.)

    (a) $P(72 < X \leq 80)$

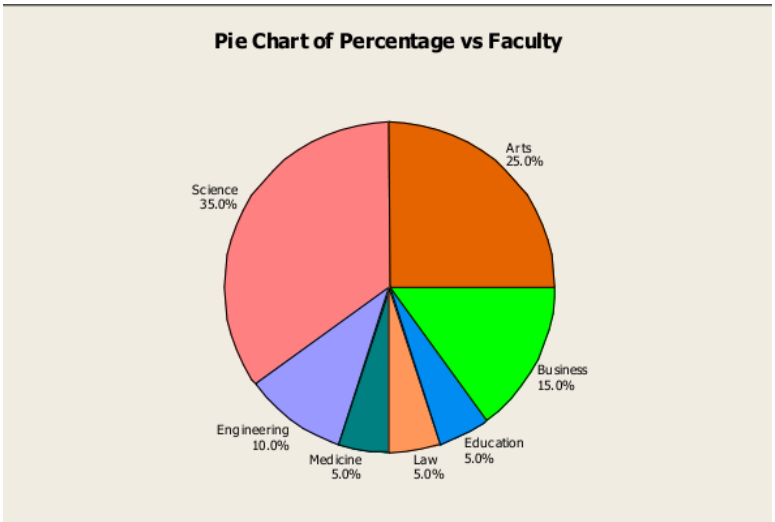
    (b) $P(56 < X \leq 64)$

    (c) $P(50 < X \leq 58)$

    (d) $P(80 < X \leq 88)$

    (e) $P(64 < X \leq 72)$

24. The pie chart below shows the percentage of students in each faculty at a university.



    If there are 3000 students in the faculty of Arts, then **how many** students are there in the faculty of Science?

    (a) 4200

    (b) 4000

    (c) 3800

    (d) 3600

    (e) 3400

25. The random variable X has the probability distribution shown below:

Value	1	2	3	7
Probability	0.65	0.20	0.10	0.05

What is the mean of X?

(a) 1.5

(b) 1

(c) 2.5

(d) 1.7

(e) more than 3

26. A study was carried out on some people who had developed a cold within the previous 24 hours. The people were randomly divided into two groups; the people in the first group had to take zinc lozenges, and the people in the second group had to take placebo lozenges. Everyone was instructed to take the lozenges every 2–3 hours until the cold was gone. The lozenges were designed so that they could not be distinguished by anyone involved in giving the lozenges to the subjects. For each person, the overall duration of cold symptoms was measured.

What kind of study is this?

(a) a double-blind experiment

(b) an observational study

(c) a voluntary-response sample

(d) a stratified sample

(e) an experiment, but not double-blind

27. A social psychologist wants to determine whether restaurant servers will get better tips if they introduce themselves by name to the people they serve. Do you think the social psychologist will use an experiment or an observational study to find out what she wants to know?

(a) Observational study, because an experiment would not be ethical.

(b) Observational study, because an experiment is difficult to do.

(c) Experiment, because data from an observational study would be worthless.

(d) Experiment, because an experiment can easily be done.

(e) Observational study, because that would allow the researcher to infer cause and effect.

28. Pulse rates of women have a normal distribution with mean 75 and standard deviation 8. Use this information for this question and the next three questions.

    What proportion of women have pulse rates less than 71?

    (a) 0.2
    (b) 0.3
    (c) 0.4
    (d) 0.1
    (e) 0.5

29. Using the information in Question 28, what proportion of women have pulse rates greater than 85?

    (a) 0.7
    (b) 0.9
    (c) 0.1
    (d) 0.3
    (e) 0.5

30. Using the information in Question 28, what proportion of women have pulse rates between 59 and 95?

    (a) less than 0.50
    (b) 0.85
    (c) 0.99
    (d) 0.95
    (e) 0.97

31. Using the information in Question 28, what pulse rate x is such that only 5% of women have a higher pulse rate than x?

    (a) 99
    (b) 103
    (c) 80
    (d) 88
    (e) 91

32. When computing a confidence interval for the population mean μ when the population SD σ is known, what value of z^* should be used for an 85% confidence interval? (The formula for the confidence interval is $\bar{x} \pm z^*\sigma/\sqrt{n}$.)

(a) 1.26

(b) 1.96

(c) 1.44

(d) 1.84

(e) 1.64

33. A fair coin is tossed 400 times. Find the standard deviation of the number of heads that will be obtained.

(a) 30

(b) 10

(c) 20

(d) 50

(e) 40

34. A newly-designed highway sign is being examined. 27 drivers, of various ages, were tested to see at what maximum distance (in feet) each driver could clearly read the sign. Each driver's age was also recorded. Some summary statistics are shown below:

Column	n	Mean	Std. Dev.	Median	Range	Min	Max	Q1	Q3
age	27	51.33	21.944	55	64	18	82	28	71
distance	27	423.33	82.927	420	310	280	590	360	460

The correlation between age and distance is -0.796. Calculate the **intercept** of the least-squares regression line for predicting distance from age.

(a) 575

(b) 425

(c) 500

(d) 350

(e) 275

35. In a certain game of chance, your probability of winning is 0.2. If you win, you win \$3. If you lose, you lose \$1 (which is the same as winning $-$\$1). You are going to play the game four times. Assume that the outcomes are independent. Let T be the total winnings from the four games. What is the standard deviation of T?

    (a) 4.8

    (b) 8.0

    (c) 1.6

    (d) 6.4

    (e) 3.2

36. A simple random sample of 25 observations is taken from a population with mean 80 and SD 20. Assume that this sample size is large enough for the Central Limit Theorem to apply. Use this information for this question and the next one.

    What is the probability that the sample mean is less than 77?

    (a) 0.51

    (b) 0.23

    (c) 0.00

    (d) 0.44

    (e) 0.72

37. Question 36 referred to taking simple random samples from a certain population. This time, two simple random samples are drawn, both of size 25. What is the probability that the two sample means will differ by more than 3?

    (a) 0.05

    (b) 0.30

    (c) 0.60

    (d) 0.90

    (e) 0.45

38. Bob and Carol go swimming every morning. They each swim a fixed distance. The time each person takes to complete the swim has a normal distribution, and the times are independent of each other. Bob has a mean time of 10 minutes, with a standard deviation of 1 minute, while Carol has a mean time of 11 minutes with a standard deviation of 0.5 minutes. What is the probability that, on a randomly chosen morning, Carol will complete her swim more quickly (in fewer minutes) than Bob does?

    (a) 0.7

    (b) 0.3

    (c) 0.1

    (d) 0.2

    (e) 0.5

39. Part of the probability distribution of a random variable X is shown below. Unfortunately, somebody spilled coffee on the original piece of paper, and some of the probabilities got lost.

Value	0	1	6
Probability	0.2	lost	lost

    It is known, however, that the mean of X is 3.3. What must $P(X = 1)$ be?

    (a) 0.2

    (b) 0.7

    (c) 0.6

    (d) 0.3

    (e) 0.5

40. The random variable Y has the distribution shown below:

Value	1	4
Probability	0.2	0.8

    The mean of Y is 3.4. What is the standard deviation of Y?

    (a) 2.5

    (b) 1.4

    (c) 5.0

    (d) 1.2

    (e) 1.9

41. The salaries paid to the 13 employees of a small market research company are as follows: the five telephone interviewers are each paid $32,000; three administrative assistants are paid $48,000; three data analysts are paid $55,000, one supervisor is paid $65,000 and one senior manager is paid $160,000.

    What is the median salary of these 13 employees?

    (a) $60,000
    (b) $55,000
    (c) $40,000
    (d) $52,500
    (e) $48,000

42. You toss 2 fair coins and count the number of heads. Independently, your friend tosses 3 fair coins and counts the number of heads. The winner is the player who gets more heads when they toss their coins. What is the probability that you win?

    (a) 0.50
    (b) 0.04
    (c) 0.39
    (d) 0.19
    (e) 0.01

43. A sample of size 20 is taken from a population with unknown mean and unknown standard deviation. What table value should be used for a 90% confidence interval for the population mean?

    (a) 1.729
    (b) 2.093
    (c) 1.645
    (d) 1.960
    (e) 2.861

44. A test of significance is significant at the 5% level. Which one of the following statements must also be true about this test result? (At most one of the statements below must be true.)

    (a) The test is not significant at the 10% level.

    (b) The test is significant at the 1% level.

    (c) The test is not significant at the 1% level.

    (d) The test is significant at the 10% level.

    (e) None of the other statements must be true.

45. A Statistics course has two tutorials (TUT01 and TUT02). TUT01 has 8 men and 12 women. TUT02 has 24 men and an unknown number of women. A student is selected at random from each tutorial. The probability that both these students are of the same gender is 0.44. How many women are in TUT02?

    (a) 24

    (b) 4

    (c) 12

    (d) 6

    (e) 30

46. A car rental company records the number of kilometres driven per day by each of its customers, and finds that the number of kilometres driven has a mean of 110 km and a standard deviation of 80 km. Based only on this information, what do you think is the shape of the distribution of the number of kilometres driven?

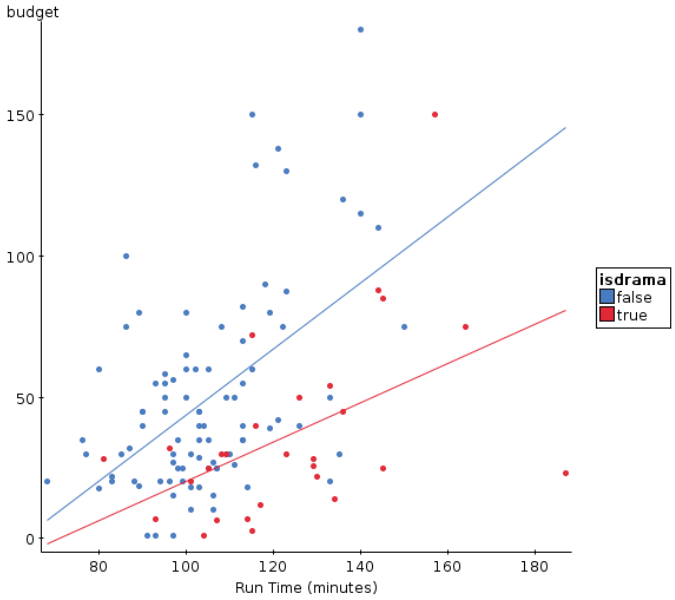
    (a) Like a normal distribution

    (b) Skewed to the right

    (c) Symmetric but not normal

    (d) Skewed to the left

47. A study was made of movies in 2005. For each movie, the following were recorded: its budget (the amount it cost to make, in millions of dollars), its running time (from start to finish, in minutes), and the genre (drama, comedy, action, etc.). The scatterplot below shows the budget vs. running length for all movies. Two regression lines are shown on the scatterplot, for predicting budget from running length. The lower line is for movies of the "drama" genre, while the upper line is for movies of all other genres. Which of the statements below do you most strongly agree with?



(a) The difference in budget between dramas and other movies of the same length remains constant as the movies get longer.

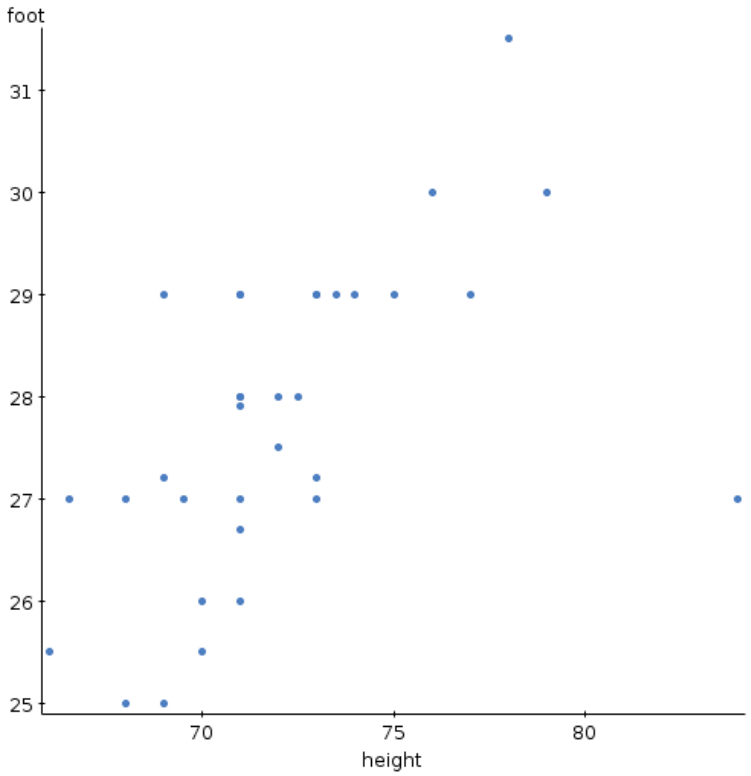
(b) Dramas cost less to make than other movies of the same length, and that difference increases as the movies get longer.

(c) Dramas cost more to make than other movies of the same length, and that difference increases as the movies get longer.

(d) Dramas cost more to make than other movies of the same length, and that difference decreases as the movies get longer.

(e) Dramas cost less to make than other movies of the same length, and that difference decreases as the movies get longer.

48. For each of the 33 male students in a college class, their height (in inches) and foot length (in centimetres) were measured. None of the men appeared to be unusually tall or short. A scatterplot of the data collected is shown below.



How would you describe what you see on the scatterplot?

(a) There is an outlier, and that outlier appears to be an error.

(b) There is an outlier, and that outlier appears to be a legitimate data value.

(c) There is a curved association.

(d) There is no association.

(e) There is a straight-line association.

49. Some people have been complaining that the children's playground at a certain city park is in need of repair. Any repairs will need to be paid for from city taxes. The city will commission a survey on this issue. The survey question is planned to be "the city should allot more funds for the maintenance and repair of children's playgrounds in city parks". What would be the best way to conduct the survey?

    (a) hand out surveys to parents of all children at some randomly chosen city parks

    (b) hand out surveys to parents of all children at the playground in this city park

    (c) draw a probability sample from all city taxpayers and contact the sampled taxpayers by phone. If the phone is not answered, ignore this taxpayer and move on to the next.

    (d) draw a probability sample from all city taxpayers and contact the sampled taxpayers by phone, following up if necessary

    (e) use a web site like `surveymonkey.com` to host the survey and advertise it to all city taxpayers.

50. Suppose you arrive at the UTSC bus stop. You have exactly 4 minutes to wait for your bus. While you are waiting, you observe other buses that are leaving. The time until the next number 38 bus leaves is a random variable with a (continuous) uniform distribution between 0 and 6 minutes; independently of that, the time until the next number 95 bus leaves is a random variable with a uniform distribution between 0 and 10 minutes. What is the probability that, while you are waiting, you observe a number 38 bus leaving, but you do not observe a number 95 bus leaving?

    (a) 0.53

    (b) 0.40

    (c) 0.27

    (d) 0.13

    (e) 0.20