University of Toronto Scarborough
STAB22 Final Examination

December 2009

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 18 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.

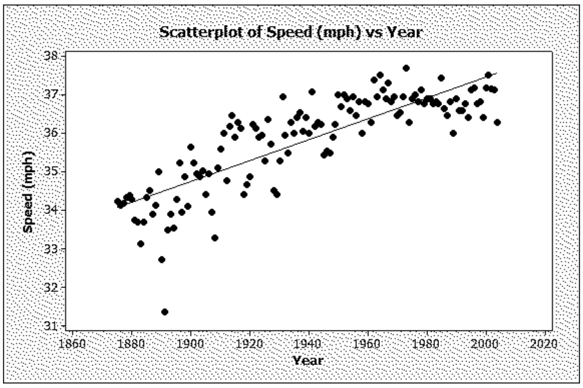
At the end of the exam, you *must* hand in your Scantron sheet (or you will receive a mark of zero for the examination). You will be graded *only* on what appears on the Scantron sheet. You may take away the question paper after the exam, but whether you do or not, anything written on the question paper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

1. Heart problems can be examined via a small tube (called a catheter) threaded into the heart from a vein in the patient's leg. It is important that the company that manufactures the catheter maintains a diameter of 2 mm. Suppose μ denotes the population mean diameter. Each day, quality control personnel make measurements to test a null hypothesis that $\mu = 2.00$ against an alternative that $\mu \neq 2.00$, using $\alpha = 0.05$. If a problem is discovered, the manufacturing process is stopped until the problem is corrected.

   What, in this context, is a type II error?

   (a) Concluding that the mean catheter diameter is not 2 mm when it is actually less than 2 mm.
   (b) Concluding that the mean catheter diameter is satisfactory when in fact it is either bigger or smaller than 2 mm.
   (c) Using a sample size that is too small.
   (d) Concluding that the mean catheter diameter is 2 mm when it actually is 2 mm.
   (e) Concluding that the mean catheter diameter is unsatisfactory when in fact it is equal to 2 mm.

2. The Kentucky Derby is a famous horse race that has been run every year since the late 19th century. The scatterplot below shows the speed (in miles per hour) of the winning horse, plotted against the year. The regression line is shown on the plot.

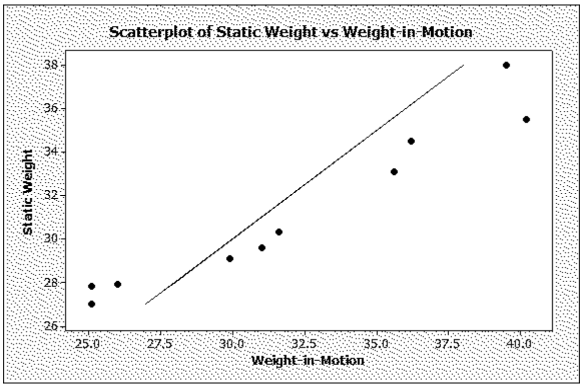


   What kind of association do you see between speed and year?

   (a) no association
   (b) negative and non-linear
   (c) negative and linear
   (d) positive and linear
   (e) positive and non-linear

3. Weighing large trucks is a slow business because the truck has to stop exactly on a scale. This is called the "static weight". The Minnesota Department of Transportation developed a new method, called "weight-in-motion", to weigh a truck as it drove over the scale without stopping. To test the effectiveness of the "weight-in-motion" method, 10 trucks were each weighed using both methods. All weights are measured in thousands of pounds.

   A scatterplot of the results is shown below. Superimposed on the scatterplot is the line that the data would follow if the weight-in-motion was always equal to the static weight. Use the scatterplot to answer this question and the one following.



How would you describe the positive association?

(a) not useful since the data are not close to the line

(b) approximately linear

(c) definitely curved

(d) no association of note

4. Question 3 described some data on two methods of weighing trucks. The Minnesota Department of Transportation wants to predict the static weight of trucks from the weight-in-motion. Which of the following statements best describes what they can do?

(a) taking the weights-in-motion and modifying them in some linear way would accurately predict the static weight.

(b) The weights-in-motion can be used to predict the static weights, but a non-linear transformation would have to be applied to do it.

(c) The static weight is accurately predicted by the weight-in-motion itself.

(d) There is no way to use the weights-in-motion to predict the static weight.

5. A factory hiring people to work on an assembly line gives job applicants a test of manual agility. This test involves fitting strangely-shaped pegs into matching holes on a board. In the test, each job applicant has 60 seconds to fit as many pegs into their holes as possible. For one job application cycle, the results were as follows:

	Male applicants	Female applicants
Subjects	41	51
Mean pegs placed	17.9	19.4
SD of pegs placed	2.5	3.4

The factory wishes to see if there is evidence for a difference between males and females. Which is more appropriate, a matched-pairs t-test or a two-sample t-test? Using the more appropriate test, what P-value do you obtain?

(a) bigger than 0.05

(b) between 0.01 and 0.02

(c) between 0.005 and 0.01

(d) less than 0.005

(e) between 0.02 and 0.05

6. A discrete random variable X has the distribution shown below:

Value	0	1	2
Probability	0.2	0.4	0.4

The random variable Y is defined as $Y = 2X + 10$. What is the mean (expected value) of Y?

(a) 12.4
(b) 10
(c) 1.2
(d) 2.4
(e) there is not enough information

7. A consumer magazine tested 14 (randomly chosen) brands of vanilla yogurt and measured the number of calories per serving of each one. Some Minitab output from the analysis is shown below.

```
One-Sample T: Calories

Variable   N         90% CI
Calories  14   (136.676, 179.038)


One-Sample T: Calories

Variable   N         95% CI
Calories  14   (132.018, 183.696)
```

Suppose that a diet guide claims that there is an average of 130 calories per serving in vanilla yogurt. You wish to assess the evidence against this claim. From the information above, what can you say about the P-value of your test of significance?

(a) between 0.01 and 0.05
(b) less than 0.01
(c) less than 0.05
(d) between 0.05 and 0.10
(e) greater than 0.10

8. A simple random sample of 10 observations is taken from a normally-distributed population with mean 9 and standard deviation 2. Let $\bar{X}$ denote the mean of this sample. Independently of this, another simple random sample of 10 observations is taken from another normally-distributed population with mean 10 and standard deviation 1. Let $\bar{Y}$ denote the mean of this second sample.

What is the probability that $\bar{X}$ is greater than $\bar{Y}$?

(a) 0.00
(b) 0.92
(c) 0.08
(d) 0.52
(e) 0.48

9. The probability that a US resident has visited Canada is 0.18, the probability that a US resident has visited Mexico is 0.09, and the probability that a US resident has visited both countries is 0.04. Consider the events "has visited Canada" and "has visited Mexico", as applied to a randomly-chosen US resident. Are these two events independent? Are they disjoint? What can you say about these events?

   (a) They are independent but not disjoint
   (b) They are disjoint but not independent
   (c) They are both independent and disjoint
   (d) They are neither independent nor disjoint

10. A researcher expects a relationship between two variables, but finds that the correlation between them is close to zero. The researcher has plenty of data. What is a possible explanation?

   (a) The relationship is a curve
   (b) The relationship is strongly linear but the correlation happened to come out close to zero
   (c) The relationship is a curved upward trend.
   (d) There cannot actually be a relationship between the variables if the correlation is close to zero.

11. A shooter fires shots at a target. Each shot is independent, and each shot hits the bull's eye with probability 0.7. Use this information for this question and the next one.

   Suppose the shooter fires 5 shots. What is the probability that the shooter hits the bull's eye exactly 4 times?

   (a) 0.07
   (b) 0.36
   (c) 0.80
   (d) 0.24
   (e) 0.69

12. Question 11 gave some information about a shooter. Suppose now the shooter fires 50 shots. What is the approximate probability that the shooter hits the bull's eye at least 40 times, using a suitable approximation? (Do not use a continuity correction.)

   (a) 0.24
   (b) 0.06
   (c) 0.61
   (d) 0.47
   (e) 0.32

13. A researcher studies children in school and finds a strong positive linear association between height and reading ability. What would the researcher's best conclusion be?

   (a) The observed association was an accident.
   (b) In any grade, taller children are better readers.
   (c) There is a lurking variable that explains the correlation.
   (d) Height and reading ability are confounded.

14. Whitefish have lengths that vary according to a normal distribution. We want to see if there is any evidence that the population mean length μ is not 9 cm. A simple random sample of 10 whitefish has sample mean length 8.5 cm and sample standard deviation 1.2 cm. Let μ denote the mean length of all whitefish. Test $H_0 : \mu = 9$ against a suitable alternative. What do you conclude, using $\alpha = 0.05$?

    (a) accept the null hypothesis
    (b) reject the null hypothesis because the sample mean is not equal to 9
    (c) fail to reject the null hypothesis
    (d) reject the null hypothesis

15. A waiter believes that the distribution of his tips is slightly skewed to the right, with a mean of \$9.60 and a standard deviation of \$5.40. The waiter is interested in the sample mean tip from his next 100 customers (which you can treat as a random sample of all possible customers). Calculate the probability that this sample mean is greater than \$10.00.

    (a) 0.53
    (b) 0.10
    (c) 0.77
    (d) 0.47
    (e) 0.23

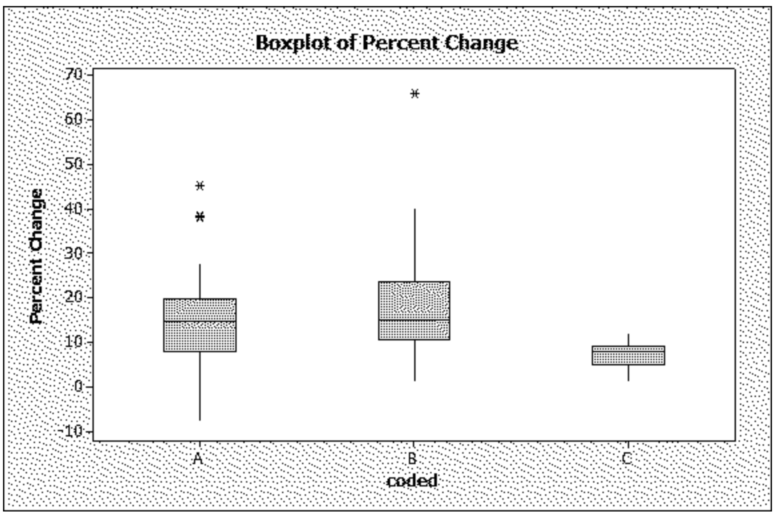
16. Data were collected on the percent change in populations for each US state (comparing the 1990 and 2000 census). The states were classified either as "south/west" (S/W) or "northeast/midwest" (NE/MW).

    Descriptive statistics for the population percent changes for each region are shown below:

Variable	region	N	Mean	SE Mean	StDev	Minimum	Q1	Median
Percent change	S/W	29	18.86	2.32	12.50	1.00	10.50	15.00
	NE/MW	19	7.158	0.677	2.949	1.000	5.000	8.000

Variable	region	Q3	Maximum
Percent change	S/W	23.50	66.00
	NE/MW	9.000	12.000

    Three side-by-side boxplots are shown below, labelled A, B, and C.



    Which boxplot is which?

(a) B is S/W and C is NE/MW
(b) A is NE/MW and B is S/W
(c) A is S/W and C is NE/MW
(d) A is NE/MW and C is S/W
(e) None of the other alternatives is correct.

17. A large city school system has 20 elementary schools. The school board is considering the adoption of a new policy that would require each student to pass a test in order to be promoted to the next grade. Before adopting this policy, the school board commissions a survey to determine whether parents agree with this plan. Use this information for this question and the 3 questions following.

    Which of the following would be a multistage sample?

    (a) Randomly select 100 parents of students from within the school board's schools.
    (b) Randomly select 4 of the elementary schools, and randomly select 25 parents of students within each of the schools chosen.
    (c) Put a large advertisement in the local newspaper asking parents to visit the school board's website to fill out the survey.
    (d) Randomly select 10 parents of students at each elementary school. Mail them a survey, and follow up with a phone call if the survey is not returned within a week.

18. In the situation of Question 17, two proposed questions to ask in seeking parents' opinions are these:

    I. Should elementary-school children have to pass high-stakes tests in order to remain with their classmates?

    II. Should schools and students be held accountable for meeting yearly learning goals by testing students before they advance to the next grade?

    Which question do you think will have a higher percentage of parents agreeing with it?

    (a) I
    (b) II
    (c) the percentages will be about the same

19. In the situation of Question 17, suppose the elementary schools are numbered 01, 02, ..., 20, and it is desired to select a simple random sample of 2 of them. Use the following excerpt from Table B to make the selection.

    `3450860202005316190748327392`

    Which two schools did you choose?

    (a) 16 and 19
    (b) some other selection of schools not listed here
    (c) 02 and 16
    (d) the excerpt from Table B is not long enough
    (e) 02 twice

20. Suppose a simple random sample of 10 observations $X_1, X_2, \ldots, X_{10}$ is taken from a normally distributed population with mean 20 and standard deviation 2. Let $\bar{X}$ denote the mean of this sample. Independently of this, another simple random sample of 10 observations $Y_1, Y_2, \ldots, Y_{10}$ is taken from another normally distributed population with mean 10 and standard deviation 1. Let $\bar{Y}$ denote the mean of this sample. Use this information for this question and the question following.

    Consider the random variable $X_{10} - Y_1$ (the 10th value in the first sample minus the first value in the second sample). This random variable has a normal distribution. What are its mean and standard deviation (respectively)?

    (a) 0 and 1
    (b) 10 and 3
    (c) 0 and $\sqrt{5}$
    (d) 10 and $\sqrt{5}$
    (e) 10 and 1

21. Using the information in Question 20, consider the random variable $\bar{X} - \bar{Y}$. What is the distribution of this random variable?

    (a) normal with mean 10 and standard deviation 0.7
    (b) standard normal
    (c) a t distribution with 9 degrees of freedom
    (d) normal with mean 10 and standard deviation 0.3
    (e) some distribution not mentioned in the other alternatives

22. In a study of the health risks of smoking, the cholesterol levels were measured for 43 smokers. A stemplot of the results is shown below.

```
Stem-and-leaf of Smokers  N  = 43
Leaf Unit = 10

  1   1  5
  1   1
  3   1  89
 18   2  000000011111111
(7)   2  2223333
 18   2  4444555
 11   2  67
  9   2  888888
  3   3  00
  1   3
  1   3  5
```

    What is the interquartile range of the cholesterol levels?

    (a) 260
    (b) 50
    (c) 60
    (d) 210
    (e) 5

23. A sample of 40 observations is taken from a population whose standard deviation is known to be 1. We are interested in testing the null hypothesis $H_0 : \mu = 12$ against the alternative $H_a : \mu < 12$. Suppose we decide to reject the null hypothesis if we observe a sample mean $\bar{x}$ that is less than 11.7. What is the probability of making a type I error with this test?

(a) less than 0.02
(b) between 0.03 and 0.04
(c) between 0.04 and 0.06
(d) greater than 0.06
(e) between 0.02 and 0.03

24. When you are conducting a test about a population mean, in which one of the following situations would you use a t-test (based on the t distribution) instead of a z procedure (based on the normal distribution)?

(a) When the population standard deviation is not known
(b) When the population mean is not known
(c) When there is some doubt about whether the sampling distribution is normal
(d) When the sample size is large

25. Are more people generally admitted to emergency rooms for vehicular accidents on Friday 13th than on other Fridays? A study compared emergency room admissions for vehicular accidents on six different Friday 13th dates, and compared with Friday 6th *in the same months*. The results are as shown:

Year	Month	Friday 6th	Friday 13th
1989	October	9	13
1990	July	6	12
1991	September	11	14
1991	December	11	10
1992	March	3	4
1992	November	5	12

Some Minitab output for these data is as shown. The first part ("two-sample") is appropriate for two independent samples, and the second part ("paired") is appropriate for pairs of measurements which are in some way dependent. Column C1 contains the Friday 6th counts, and column C2 contains the Friday 13th counts.

```
Two-sample T for C1 vs C2

    N   Mean  StDev  SE Mean
C1  6   7.50   3.33      1.4
C2  6  10.83   3.60      1.5

Difference = mu (C1) - mu (C2)
Estimate for difference:  -3.33
95% CI for difference:  (-7.86, 1.20)
T-Test of difference = 0 (vs not =): T-Value = -1.66  P-Value = 0.130  DF = 9


Paired T-Test and CI: C1, C2

Paired T for C1 - C2

            N   Mean  StDev  SE Mean
C1          6   7.50   3.33     1.36
C2          6  10.83   3.60     1.47
Difference  6  -3.33   3.01     1.23


95% CI for mean difference: (-6.49, -0.17)
T-Test of mean difference = 0 (vs not = 0): T-Value = -2.71  P-Value = 0.042
```

Using the more appropriate one of these analyses, what would be an appropriate P-value for this study?

(a) 0.021
(b) the P-value is larger than 0.5 because the difference in sample means should be positive.
(c) 0.130
(d) 0.065
(e) 0.042

26. In a certain population, 10% of the people are beautiful, 10% are intelligent but only 1% are beautiful and intelligent. Use this information for this question and the next one.

For a person picked at random from the population what is the probability that the person is not intelligent?

(a) 0.19
(b) 0.90
(c) 0.11
(d) 0.01

27. In the situation of Question 26, suppose a person is picked at random from the population. What is the probability that the person is either beautiful or intelligent? (Two of the alternatives below are reasonable answers to the question; if you mark either of them, you will get this question correct.)

(a) Cannot be done using the methods of this course
(b) 0.99
(c) 0.19
(d) 0.20
(e) 0.50

28. A simple random sample was taken of 288 teachers in the state of Utah. A 95% confidence interval for the population mean μ is from \$37,500 to \$41,400. Which statement below correctly describes what the confidence interval tells us?

(a) If we took many random samples of Utah teachers, about 95% of them would produce a confidence interval that contained the mean salary of all Utah teachers.

(b) If we took many random samples of Utah teachers, about 95% of them would produce this confidence interval.

(c) About 95% of Utah teachers earn between \$37,500 and \$41,400.

(d) We are 95% confident that the mean salary of all teachers in the United States is between \$37,500 and \$41,400.

(e) About 95% of the teachers in the sample earn between \$37,500 and \$41,400.

29. A researcher studied the times taken by mice to learn to run a simple maze. The times were measured in minutes. The researcher took samples of 6 white mice and 6 brown mice. Some output is given below, but unfortunately the important part of the output has been lost. What is the lower limit of a 95% confidence interval for the difference in maze learning times between white mice and brown mice, in that order?

```
Two-sample T for white vs brown

       N   Mean  StDev  SE Mean
white  6  17.00   4.56      1.9
brown  6  16.67   5.05      2.1
```

(a) 5.9

(b) -5.9

(c) 6.6

(d) -6.6

(e) too much of the output was lost: impossible to calculate.

30. “Shingles” is a painful, but not life-threatening, skin rash. A doctor has discovered a new ointment that he believes will be more effective in the treatment of shingles than the current medication. Eight patients are available to participate in the initial trials of this new ointment. Use this information for this question and the three questions following.

What would be the best way to assess the doctor’s belief? (You can assume that the available patients are a mixture of males and females and that there is no difference between males and females in the effectiveness of the current medication and the new ointment).

(a) Allow the doctor to decide which four patients should receive the new ointment and which four should receive the current medication.

(b) Divide the eight patients at random into a treatment group and a control group, with the patients in the control group getting the current medication.

(c) Give the new ointment to the four oldest patients, because they need it most, and give the current medication to the other patients.

(d) There are only eight patients, so test the new ointment on them all.

31. In the situation of Question 30, how could this experiment be made most nearly double-blind?

(a) Employing a nurse to apply the new ointment and current medication, rather than the doctor.
(b) Giving the current medication in the form of an ointment.
(c) Having a control group.
(d) Using randomization.

32. In the situation of Question 30, which of the following could be a response variable?

(a) The gender of the patient.
(b) The age of the patient.
(c) The time it takes for the skin rash to disappear.
(d) Whether or not a patient receives the new ointment.
(e) Whether or not the patient survives for a year.

33. In the experiment of Question 30, it turns out that the new ointment does appear to be effective, but that the new ointment is more effective for younger patients. A second study is planned, and many more patients are available. Based on the knowledge gained from the first study, what would you do in the second study?

(a) Use age as the response variable.
(b) Use treatment and control groups and also ensure that the patients in the two groups are similar in terms of age.
(c) Decide which patients receive the new ointment using a multistage sample.
(d) Design the second study in the same way as the first.

34. You roll a (fair, 6-sided) die. If you get a 6 on the first roll, you win \$100. If not, you roll a second time, and if you get a 6 on the second roll, you win \$50. Otherwise, you win nothing. What are your mean winnings from this game?

(a) \$10.50
(b) \$30.00
(c) \$75.00
(d) \$25.00
(e) \$23.50

35. A smelt is a type of small fish. Suppose that smelt lengths have a standard deviation of 2 cm. A simple random sample of 30 smelts has a sample mean length of 7.5 cm. Use this information for this question and the next one.

What is the upper limit of a 95% confidence interval for the mean length of all smelts?

(a) 8.22 cm
(b) 8.50 cm
(c) 7.50 cm
(d) 8.36 cm
(e) 6.78 cm

36. In the situation of Question 35, and using the same data, suppose now that you are testing the null hypothesis $H_0 : \mu = 8.5$ cm against $H_a : \mu < 8.5$ cm, where μ is the mean length of all smelts. What do you conclude from your test? Use $\alpha = 0.01$.

(a) reject the null hypothesis
(b) fail to reject the null hypothesis
(c) cannot do test without knowing what μ is
(d) need to use a larger value of α
(e) accept the null hypothesis

37. Out of all American workers, 56% have a workplace retirement plan, 68% have health insurance, and 49% have both benefits. For a randomly sampled worker, are "has a workplace retirement plan" and "has health insurance" independent events?

(a) No, more workers have both benefits than you would expect if the two events were independent.
(b) No, fewer workers have both benefits than you would expect if the two events were independent.
(c) There is not enough information to decide.
(d) Yes.

38. A dogfish is a kind of shark. A simple random sample of 10 of one type of dogfish produced a sample mean length of 1.2m and a sample standard deviation of 0.1m. Dogfish lengths are believed to have a shape close to a normal distribution. A 95% confidence interval for the mean length of all dogfish of this type has what lower limit?

(a) 1.099
(b) need to know the population mean
(c) 1.138
(d) 1.128
(e) need to know the population standard deviation

39. A and B are disjoint events with $P(A) = 0.2$ and $P(B) = 0.7$. Use this information for this question and the next one.

What is the probability that A and B both happen?

(a) there is not enough information
(b) 0.00
(c) 0.90
(d) 0.14

40. In the situation of Question 39, what is the probability that exactly one of the two events occurs?

(a) 0.62
(b) there is not enough information
(c) 0.76
(d) 0.90

41. A continuous random variable X has a probability density function $f(x)$ that is equal to 1 for $0 < x < 1$. What is $P(3/4 < X < 1)$?

(a) 3/4
(b) 1
(c) 1/4
(d) 0

42. A discrete random variable X has this distribution:

Value	2	3	4
Probability	0.1	0.5	0.4

What is the mean (expected value) of X?

(a) 3.3
(b) less than 2
(c) 3.6
(d) 3.0
(e) 2.5

43. The residents in a certain street are concerned that vehicles are driving too fast along their street, where the speed limit is 40 km/h. The residents collect a large random sample of vehicles driving along their street, and obtained a sample mean of 42 km/h (2 km/h over the speed limit). Let μ denote the mean speed of all vehicles driving along the street. For a test of $H_0 : \mu = 40$ against $H_a : \mu > 40$, the P-value is less than 0.05.

What would be the most appropriate reaction to these results?

(a) Because the P-value is small, there is evidence that cars are travelling dangerously fast on the street.
(b) An average of 2 km/h over the speed limit is not of practical importancein this case.
(c) The P-value is small, so there is no evidence that vehicles are travelling too fast on average.
(d) The test says that the average of 2 km/h over the speed limit must be of practical importance.

44. A random survey of cars parked in a university's parking lot revealed the following information about the country of origin of the car and whether its driver was a student or a staff member:

	Driver	
Origin	Student	Staff
America	107	105
Europe	33	12
Asia	55	47

Use this information for this question and the two following.

In the joint distribution, what proportion of cars are from Europe and driven by a student?

(a) 0.17
(b) 0.25
(c) 0.03
(d) 0.73
(e) 0.09

45. Using the information from Question 44, in the marginal distribution of car origins, what proportion of cars are from America?

(a) 0.55
(b) 0.64
(c) 0.75
(d) 0.72
(e) 0.59

46. Look again at the information in Question 44. Given that a driver is a member of staff, what proportion of these drivers drive a car from Asia?

(a) 0.35
(b) 0.28
(c) 0.29
(d) 0.54
(e) 0.46

47. Suppose two fair 6-sided dice are rolled, and the numbers of spots on the uppermost faces are recorded. Use this information for this question and the next one.

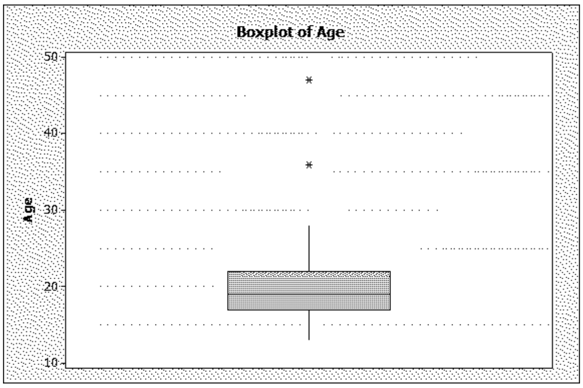
What is the probability that the total number of spots is exactly 5?

(a) 4/36
(b) 2/36
(c) 1/36
(d) 4/12

48. In Question 47, two fair dice were rolled. What is the probability of getting at least one 5? (That is, what is the probability that at least one of the dice shows 5 spots?)

(a) 10/36
(b) 11/36
(c) 2/6
(d) 5/6

49. A company monitors accidents at rock concerts. When someone is injured due to “crowd crush”, the company collects information about the victim. A boxplot of the victims’ ages is shown below, along with some other information about the ages. Use these to answer this question and the next one.



Variable	N	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Age	66	0	20.136	0.628	5.099	13.000	17.000	19.000	22.000

Variable	Maximum
Age	47.000

How would you describe the shape of the distribution?

(a) Like a normal distribution.
(b) Skewed to the right.
(c) Skewed to the left.
(d) Approximately symmetric.

50. In Question 49, some information was given about the ages of victims of “crowd crush”. According to the usual rule, how high an age would be considered an outlier?

(a) 28.2
(b) 29.5
(c) 25.5
(d) 22.0
(e) 19.0

51. A simple random sample of 17 cows were given a special feed supplement to see if it will promote weight gain. The sample mean weight gain was 55.5 pounds, and the population standard deviation was known to be 10 pounds. The population distribution of weight gain is very close to normal. Calculate a 95% confidence interval for the population mean weight gain for all cows. What is the lower limit of this confidence interval, to the nearest pound?

(a) 56
(b) 50
(c) 60
(d) 61
(e) 51

52. Question 51 discussed an investigation of the effect of a special feed for cows on weight gain. Use the information in that question, except suppose now that the figure of 10 pounds is the *sample* standard deviation. What is the lower limit of a 95% confidence interval for the population mean, to the nearest pound?

   (a) 60
   (b) 50
   (c) 56
   (d) 61
   (e) 51

53. A psychology researcher is testing how long it takes for rats to run through a certain maze (to reach the food at the other end). The researcher tests a maze with 10 rats, obtaining a sample mean of 48 seconds. Other researchers using similar mazes have found a mean time of 55 seconds, so the first researcher carries out a test of $H_0 : \mu = 55$ against $H_a : \mu \neq 55$ for his data, and obtains a P-value of 0.08, which is not smaller than the $\alpha = 0.05$ chosen.

   What would be the best conclusion from these data?

   (a) Because the P-value is greater than α, we have proved that the population mean is equal to 55 seconds.
   (b) It might still be true that the mean time is not equal to 55 seconds, but the test does not have enough power to prove this.
   (c) The researcher should use a smaller sample size.
   (d) The P-value is small enough to conclude that the mean is not 55 seconds.

54. Suppose a simple random sample of size 10 is drawn from a normal population with mean $\mu = 20$ and standard deviation $\sigma = 2$. Denote the 10 sampled values $X_1, X_2, \ldots, X_{10}$, let $\bar{X}$ denote the sample mean, and let S denote the sample standard deviation. Use this information for this question and the two following questions.

   The random variable $(X_1 - 20)/2$ has a normal distribution. What are its mean and standard deviation (respectively)?

   (a) 0 and 2
   (b) 0 and 1
   (c) 0 and $1/\sqrt{10}$
   (d) 20 and 2
   (e) 20 and $2/\sqrt{10}$

55. Using the information in Question 54, the random variable $\bar{X} - 10$ has a normal distribution. What are its mean and standard deviation (respectively)?

   (a) 10 and $2/\sqrt{10}$
   (b) 0 and 1
   (c) 0 and $2/\sqrt{10}$
   (d) 10 and 2
   (e) 0 and 2

56. Using the information in Question 54, consider the random variable $(\bar{X} - 20)/(S/\sqrt{10})$. What distribution does this have?

(a) a standard normal distribution
(b) some distribution not given in the other alternatives
(c) a t-distribution with 19 degrees of freedom
(d) a normal distribution, but not a standard normal distribution
(e) a t-distribution with 9 degrees of freedom

57. The company that makes a well-known brand of chocolate-chip cookies advertises that the 500g bags contain an average of "at least 1100 chocolate chips". To test this claim, a dedicated group of students purchased (a random sample of) 16 500g bags of these cookies and counted the number of chocolate chips in each bag. The students found a sample mean of 1188 chocolate chips per bag. The population standard deviation is known to be 100, from knowledge of the manufacturing process.

Calculate the P-value for the test of $H_0 : \mu = 1100$ against $H_a : \mu \neq 1100$, where μ is the mean number of chocolate chips in all 500g bags of these cookies. What do you get?

(a) between 0.001 and 0.0025
(b) less than 0.0006
(c) between 0.002 and 0.005
(d) greater than 0.005
(e) less than 0.0003

58. According to real estate data, 21% of homes for sale have swimming pools. A simple random sample of 5 homes for sale is taken. What is the probability that at least one of the homes in the sample has a swimming pool?

(a) 0.31
(b) 0.21
(c) 0.69
(d) 0.79
(e) 0.50

59. What does the Central Limit Theorem say?

(a) The sampling distribution of the sample mean is approximately normal for large samples.
(b) If the sample is large, the population from which the sample comes is approximately normal.
(c) The sampling distribution of the sample standard deviation is approximately normal for large samples.
(d) The sample mean is likely to be very close to the population mean if the sample is large.

60. A 2004 survey of the world's countries found a strong positive correlation between the percentage of the country's population regularly using cellphones and life expectancy at birth (in years). What can you conclude from this?

(a) Cellphone use and life expectancy are confounded
(b) In countries where cellphone use is low, cellphone use should be encouraged in order to increase life expectancy
(c) Using cellphones is good for your health
(d) Some other variable is the cause of the high correlation