THE UNIVERSITY OF AUCKLAND

STATS 101
STATS 101G
STATS 108

FIRST SEMESTER, 2014
Campus: CITY

VERSION 1

STATISTICS

Introduction to Statistics
Statistics for Commerce

(Time allowed: THREE hours)

ANSWERS ON PAGE 42

NOTE:

* **This examination consists of 50 multiple-choice questions.**
* **All questions have a single correct answer.**
* **If you give more than one answer to any question, you will receive zero marks for that question.**
* **No mark is deducted for an incorrect answer.**
* **All questions carry the same mark value.**
* **Answers must be written on the special answer sheet provided.**
* **Calculators are permitted.**

ATTACHMENT:

* **Appendix A: Wolves Data** pages 27 and 28
* **Appendix B: Ice Slurry Data** pages 29 and 30
* **Appendix C: Anchoring Data** pages 31 and 32
* **Appendix D: Alcohol-related Offences Data** pages 33 and 34
* **Appendix E: Bullying Data** page 35
* **Appendix F: Olympics Data** pages 36 to 38
* **Appendix G: Battery Data** pages 39 and 40
* **Appendix H: Blood Pressure Data** pages 41 and 42

CONTINUED

BLANK PAGE

CONTINUED

Questions 2 to **4** refer to the information in **Appendix A**, pages 27 and 28.

1. **No longer examined**

2. In the study, as well as the 10 male wolves there were 6 female wolves from the Arctic region. To explore the relationship between **Braincase** and **Sex** for the wolves from the Arctic region, the **most** appropriate display tool to use is:

(1) back-to-back histograms of **Braincase** for both levels of **Sex**, using the same scale for each plot.

(2) a two-way table of counts with values of **Braincase** for each level of **Sex** forming each of the columns.

(3) side-by-side box plots of **Braincase** for both levels of **Sex**, using the same scale for each plot.

(4) a scatter plot of **Braincase** against **Sex**.

(5) side-by-side dot plots of **Braincase** for both levels of **Sex**, using the same scale for each plot.

3. Which **one** of the following statements for the plots shown in Figures 2 and 3, page 28, is **false**?

(1) Of the 1000 re-samples taken with replacement from the original sample data, in more than 25 of them the re-sample mean for the Arctic male wolves was greater than that for the Rocky Mountains male wolves.

(2) The bootstrap distribution plot shows the extent of the variation in the differences in re-sample means of 1000 re-samples taken with replacement from the original sample data.

(3) The bootstrap distribution plot displays the distribution of the differences in crown lengths between male wolves from the Arctic and Rocky Mountains populations.

(4) The mean crown length of these Rocky Mountains male wolves is 0.20 mm greater than the mean crown length of these Arctic male wolves.

(5) The bootstrap confidence interval is obtained by taking the lower and upper limits of the central 95% of the bootstrap distribution.

CONTINUED

4. Which **one** of the following statements about crown lengths is **false**?

(1) It is plausible that the mean crown length for Rocky Mountains male wolves is 0.75 mm greater than that for Arctic male wolves.

(2) It is not plausible that the mean crown length for Rocky Mountains male wolves is 0.4 mm less than that for Arctic male wolves.

(3) It is plausible that the mean crown length for Rocky Mountains male wolves is the same as that for Arctic male wolves.

(4) It is a fairly safe bet that the mean crown length for Rocky Mountains male wolves is somewhere between 0.19 mm less than and 0.59 mm greater than that for Arctic male wolves.

(5) It is plausible that the mean crown length for Rocky Mountains male wolves is 0.5 mm greater than that for Arctic male wolves.

CONTINUED

Questions 5 to **8** refer to the information in **Appendix B**, pages 29 and 30.

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

(1) The volunteers were not blinded because they would have known the type of liquid they had drunk before each trial.

(2) Blocking was used in this study because, in the first trial, five volunteers drank ice slurry and five drank cold water.

(3) Randomising the order of the type of liquid to be drunk was used to account for any carryover effects such as fatigue or learning effects.

(4) This study is a well-designed experiment because the order in which each type of liquid was to be drunk was randomly allocated to each volunteer.

(5) If the person measuring the time taken to run to exhaustion was not told which type of liquid the volunteer had drunk, then a form of blinding was used.

6. Which **one** of the following statements about this paired-data t-procedure is **false**?

(1) It is critical that the two times taken to run to exhaustion for any one of the ten volunteers are independent of the two times taken to run to exhaustion for the other nine volunteers.

(2) The dot plot of the differences displayed in Figure 5, page 29, should be examined when considering the validity of this t-procedure.

(3) It is critical that the ten differences displayed in Figure 5, page 29, are independent of each other.

(4) For each of the ten volunteers, it is critical that the time taken to run to exhaustion after drinking ice slurry is independent of the time taken to run to exhaustion after drinking cold water.

(5) The dot plots displayed in Figure 4, page 29, are not relevant when considering the validity of this t-procedure.

Question 8 assumes that the use of the paired-data t-procedure is appropriate. (Note that it may not be.)

This questions also assumes that these ten volunteers form a random sample from some male population.

7. **No longer examined**

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

With 95% confidence, we estimate:

(1) that, on average, when a member of this population drinks ice slurry before running to exhaustion the time taken to run to exhaustion is somewhere between 6.9 minutes and 12.1 minutes longer than when he drinks cold water.

(2) that in this population, the time taken to run to exhaustion after drinking ice slurry is somewhere between 6.9 minutes and 12.1 minutes longer than that after drinking cold water.

(3) that in this population, the time taken to run to exhaustion after drinking ice slurry is, on average, 9.5 minutes longer than that after drinking cold water, with a margin of error of 2.6 minutes.

(4) that in this population, the mean difference in the time taken to run to exhaustion after drinking ice slurry compared to drinking cold water is somewhere between 6.9 minutes and 12.1 minutes, with the time taken after drinking ice slurry being longer.

(5) that in this population, the time taken to run to exhaustion after drinking ice slurry is, on average, somewhere between 6.9 minutes and 12.1 minutes longer than that after drinking cold water.

Questions 9 to **12** refer to the information in **Appendix C**, pages 31 and 32.

9. Which **one** of the following statements gives the **best** explanation for why the assignment of the 21 students to one of the two groups was done randomly?

The assignment of the 21 students to one of the two groups was done randomly mainly to try to:

(1) reduce the extent of sampling error.

(2) make Group 1 and Group 2 as similar as possible in all aspects apart from the question (anchor) they had been asked.

(3) make Group 1 representative of some larger group and to make Group 2 representative of some larger group.

(4) increase the accuracy of the results.

(5) minimise the extent of nonsampling error.

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

(1) Chance was acting alone means that the observed difference between the two group medians was purely and simply the result of luck as to which students happened to be assigned to which group.

(2) Under chance alone, it would not have mattered to which group a student had been randomly assigned, the student would still have given the same answer.

(3) Chance was acting alone means that the students' answers had nothing to do with the size of the numbers given with the two questions.

(4) Chance was acting alone means that the observed difference between the medians of Group 1 and Group 2 was due solely to the random allocation of the students to the groups.

(5) If chance was acting alone, then if the student in Group 2 who gave an answer of 200 million had been in Group 1 they would have given an answer less than 200 million.

CONTINUED

11. Which **one** of the following statements about the re-randomisation distribution is **true**?

Each dot in the re-randomisation distribution represents the difference between the group medians from:

(1) one re-sample of size 21, without replacement, from the original 21 students.

(2) a repeat of the experiment with the same 21 students.

(3) a repeat of the experiment with 21 different students.

(4) one re-randomisation under chance alone.

(5) one re-sample of size 21, with replacement, from the original 21 students.

12. Which **one** of the following statements is a valid interpretation of the randomisation test result?

(1) The *P-value* is about 0.012 and there is strong evidence of an anchoring effect for all Statistics students at university in the United States.

(2) The *P-value* is about 0.012 and there is strong evidence of an anchoring effect for these 21 Statistics students.

(3) The *P-value* is about 0.006 and there is strong evidence of an anchoring effect for these 21 Statistics students.

(4) The *P-value* is about 0.006 and there is strong evidence of an anchoring effect for all Statistics students at university in the United States.

(5) It would be unwise to make a conclusion from this test because the group sizes are too small.

CONTINUED

Questions 13 to **23** refer to the information in **Appendix D**, pages 33 and 34.

13. Which **one** of the following statements about the variables is **true**?

(1) The variables **Day** and **TimePeriod** are both ordinal.

(2) The variable **Day** is categorical and the variable **TimePeriod** is numeric.

(3) The variable **Day** is discrete and the variable **TimePeriod** is ordinal.

(4) The variable **Day** is ordinal and the variable **TimePeriod** is nominal.

(5) The variables **Day** and **TimePeriod** are both discrete.

Use Table 2, page 33, to answer **Questions 14** and **15**.

14. What proportion of the offences did **not** occur in the Morning period?

(1) 0.843

(2) 0.039

(3) 0.804

(4) 0.196

(5) 0.157

15. What percentage of the offences occurred in the Night period on Fridays or Saturdays?

(1) 80.9%

(2) 50.2%

(3) 66.5%

(4) 53.8%

(5) 33.4%

Questions 16 to **23** refer to the following information.

A Chi-square test for independence was conducted on the data. The results of this test are shown in Table 3, page 34.

16. Which **one** of the following is an **incorrect null** hypothesis for this Chi-square test?

(1) The variables **Day** and **TimePeriod** are not related.

(2) The distribution of **TimePeriod** is the same for each level of **Day**.

(3) There is no association between the variables **Day** and **TimePeriod**.

(4) The distribution of **Day** is the same for each level of **TimePeriod**.

(5) At least two levels of **Day** have distributions of **TimePeriod** which are different.

17. Which **one** of the following statements gives the **best** justification concerning the validity of this Chi-square test?

(1) Because one of the cell contributions to the test statistic is less than 1 there is cause for concern with the validity.

(2) Because all of the observed counts are greater than or equal to 5 there is **no** cause for concern with the validity.

(3) Because at least three of the cell contributions to the test statistic are less than 5 there is cause for concern with the validity.

(4) Because none of the expected counts is less than 1 there is **no** cause for concern with the validity.

(5) Because all of the expected counts are greater than or equal to 5 there is **no** cause for concern with the validity.

Questions 18 to **23** assume that the use of the Chi-square test is appropriate.
(Note that it may not be.)

18. Consider the cell in Table 3, page 34, for offences in the Evening period on Thursdays. Under the null hypothesis, the expected count for this cell is approximately:

(1) 1291.3
(2) 914.4
(3) 1242.5
(4) 950.3
(5) 1277.0

19. Consider the cell in Table 3, page 34, for offences in the Midday period on Wednesdays. This cell's contribution to the Chi-square test statistic is approximately:

(1) 0.04
(2) 5.15
(3) 6.45
(4) 0.25
(5) 0.06

20. The *P-value* for this Chi-square test is calculated by:

(1) $\text{pr}(\chi^2 \geq 2130.548)$ where $\chi^2 \sim \text{Chi-square}(df = 27)$
(2) $\text{pr}(\chi^2 \geq 2130.548)$ where $\chi^2 \sim \text{Chi-square}(df = 18)$
(3) $2\times\text{pr}(\chi^2 \geq 2130.548)$ where $\chi^2 \sim \text{Chi-square}(df = 27)$
(4) $2\times\text{pr}(\chi^2 \geq 2130.548)$ where $\chi^2 \sim \text{Chi-square}(df = 18)$
(5) $2\times\text{pr}(\chi^2 \geq 2130.548)$ where $\chi^2 \sim \text{Chi-square}(df = 28)$

CONTINUED

21. Which **one** of the following statements is **true**?

Under the null hypothesis, we would expect to see the proportion of these offences that occur in the Evening period to be approximately:

(1) $\frac{314}{6401}, \frac{509}{6401}, \frac{786}{6401}, \frac{1277}{6401}, \frac{1399}{6401}, \frac{1388}{6401}$ and $\frac{728}{6401}$ for the seven respective days.
(2) $\frac{1076}{34791}, \frac{1476}{34791}, \frac{2547}{34791}, \frac{5165}{34791}, \frac{7484}{34791}, \frac{9965}{34791}$ and $\frac{7078}{34791}$ for the seven respective days.
(3) $\frac{314}{1076}, \frac{509}{1476}, \frac{786}{2547}, \frac{1277}{5165}, \frac{1399}{7484}, \frac{1388}{9965}$ and $\frac{728}{7078}$ for the seven respective days.
(4) $\frac{6401}{34791}$ for each of the days.
(5) $\frac{1}{7}$ for each of the days.

22. Which **one** of the following statements is **not** a valid conclusion from the results of this Chi-square test?

(1) The difference between the observed count and the expected count in at least one of the cells cannot be attributed to sampling variation alone.
(2) There is very strong evidence that the variables **Day** and **TimePeriod** are not independent.
(3) At the 5% level of significance, it may be claimed that the variables **Day** and **TimePeriod** are not associated.
(4) At the 5% level of significance, it may be claimed that there is a link between the variables **Day** and **TimePeriod**.
(5) At the 5% level of significance, the analysis has resulted in a statistically significant result.

23. Which **one** of the following statements gives the **best** explanation for the small *P-value* of 0.000?

(1) There were far more offences in the Midday and Evening periods on Mondays than would have been expected when the null hypothesis is true.
(2) The number of offences in the Night period on Thursdays and the number of offences expected to be in this category when the null hypothesis is true, are quite similar.
(3) There were far fewer offences in the Morning period on Tuesdays than would have been expected when the null hypothesis is true.
(4) There were far more offences in the Morning period on Sundays than would have been expected when the null hypothesis is true.
(5) There is very little difference between the number of offences in the Evening period on Fridays and the number of offences expected to be in this category when the null hypothesis is true.

CONTINUED

Questions 24 to **27** refer to the information in **Appendix E**, page 35.

24. For the **boys** in this study, the relative risk of being bullied for the short group compared to the not-short group is:

(1) 0.45
(2) 2.24
(3) 3.32
(4) 1.78
(5) 1.92

Questions 25 to **27** refer to the following additional information.

Let:

p_S be the underlying proportion of **short pupils** who have been bullied at some time at secondary school

and

p_{NS} be the underlying proportion of **not-short pupils** who have been bullied at some time at secondary school.

25. For the purposes of calculating the standard error of the estimate, $se(\widehat{p}_S - \widehat{p}_{NS})$, the sampling situation is **best** described as:

(1) one sample of size 72, several response categories.
(2) two independent samples of sizes 92 and 117.
(3) one sample of size 209, many yes/no items.
(4) one sample of size 209, several response categories.
(5) two independent samples of sizes 72 and 137.

26. From this survey, a 95% confidence interval for $p_S - p_{NS}$ is $(0.071, 0.329)$.

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

With 95% confidence, we estimate that the underlying proportion of short pupils who have been bullied at some time at secondary school is:

(1) somewhere between 0.071 and 0.329 less than the corresponding underlying proportion of not-short pupils.
(2) 0.071 and the corresponding underlying proportion of not-short pupils is 0.329.
(3) up to 0.258 more than the corresponding underlying proportion of not-short pupils.
(4) somewhere between 0.071 less than and 0.329 more than the corresponding underlying proportion of not-short pupils.
(5) somewhere between 0.071 and 0.329 more than the corresponding underlying proportion of not-short pupils.

27. If a two-tailed t-test for no difference between p_S and p_{NS} was conducted on this data, which **one** of the following statements is **true**?

(1) The *P-value* would be less than 0.05.
(2) There is not enough information to give an indication of the size of the *P-value*.
(3) The *P-value* would be between 0.05 and 0.1.
(4) The *P-value* would be greater than 0.2.
(5) The *P-value* would be between 0.1 and 0.2.

Questions 28 to **34** refer to the information in **Appendix F**, pages 36 to 38.

28. Which **one** of the following statements about the data in Figure 8, page 37, is **false**?

(1) In 2000, more than half of the runners had a reaction time slower than 0.200 seconds.

(2) In 2008, the interquartile range of the reaction times is less than 0.05 seconds.

(3) In at least three of the four years, at least one runner had an unusually slow reaction time.

(4) The reaction times in 2004 are positively (right) skewed.

(5) In 2000, fewer than a quarter of the runners had a reaction time faster than 0.150 seconds.

29. Which **one** of the following is **not** a correct pair of hypotheses for this F-test?

(1) H_0: The underlying mean reaction times are the same for all four years.
H_1: The underlying mean reaction times are not the same for all four years.

(2) H_0: There is no year effect on the mean reaction times.
H_1: There is a year effect on the mean reaction times.

(3) H_0: The sample mean reaction times are the same for all four years.
H_1: The sample mean reaction times are different for at least one pair of years.

(4) H_0: **Reaction Time** is independent of **Year**.
H_1: **Reaction Time** is not independent of **Year**.

(5) H_0: The underlying mean reaction times are the same for all four years.
H_1: The underlying mean reaction times are different for at least one pair of years.

30. Which **one** of the following statements is **not** an assumption of this F-test?

(1) The reaction times within each year are independent of each other.

(2) The underlying mean reaction times are the same for all four years.

(3) The underlying distribution of reaction times for each year has a Normal distribution.

(4) The reaction time of a runner in one year is independent of the reaction time of any runner in one of the other three years.

(5) The standard deviations of the underlying distributions of reaction times for each year are all equal.

31. Which **one** of the following statements gives the **best** description of the appropriateness of using a one-way analysis of variance (ANOVA) F-test on these data?

(1) We have **no** concerns about using an F-test because the F-test is sufficiently robust to withstand the amount of skewness evident in the data.

(2) We have **no** concerns about using an F-test because the F-test can withstand both the departures from Normality evident in the data and the variability in the population standard deviations as estimated by the sample standard deviations.

(3) We have **no** concerns about using an F-test because the ratio of the largest sample standard error to the smallest sample standard error is less than 2.

(4) We should be wary about using an F-test because the ratio of the largest sample standard deviation to the smallest sample standard deviation is greater than 2.

(5) We should be wary about using an F-test because the difference between the largest sample size and the smallest sample size is too big.

Questions 32 and **33** assume that the use of the F-test is appropriate.
(Note that it may not be.)

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

(1) There is very strong evidence against the four underlying mean reaction times being equal.

(2) At the 5% level of significance, the difference between at least one pair of sample means cannot be explained as being due to sampling variation alone.

(3) There is very strong evidence that the underlying mean reaction times are not all the same.

(4) The variability between the four sample means is large relative to the variability within the four samples.

(5) There is no evidence that the underlying mean reaction times for the four years are equal.

33. Which **one** of the following statements is **not** a correct conclusion, at the 5% level of significance, from the Tukey multiple comparisons output in Table 5?

(1) The underlying mean reaction time for runners in 2008 was faster than that for runners in 2000.

(2) The slowest underlying mean reaction time was for runners in 2000.

(3) There is a statistically significant difference between the mean reaction times for the runners in 2000 and those in 2008.

(4) The difference between the mean reaction times for the runners in 2008 and those in 2012 is **not** statistically significant.

(5) The fastest underlying mean reaction time was for runners in 2004.

34. Which **one** of the following forms of analysis could be used to investigate the relationship between **Reaction Time** and **Finishing Time**?

(1) Simple linear regression

(2) A Chi-square test for independence

(3) A two independent samples t-test for no difference between two means

(4) A Chi-square test for goodness of fit

(5) A t-test for no difference between two proportions

CONTINUED

Questions 35 to **40** refer to the information in **Appendix G**, pages 39 and 40.

35. Which **one** of the following statements is a correct pair of hypotheses for this t-test?

(1) H_0: $\mu_E - \mu_U \neq 0$
H_1: $\mu_E - \mu_U = 0$

(2) H_0: $\mu_E - \mu_U = 0$
H_1: $\mu_E - \mu_U > 0$

(3) H_0: $\mu_E - \mu_U = 0$
H_1: $\mu_E - \mu_U \neq 0$

(4) H_0: $\overline{x}_E - \overline{x}_U \neq 0$
H_1: $\overline{x}_E - \overline{x}_U \neq 0$

(5) H_0: $\overline{x}_E - \overline{x}_U = 0$
H_1: $\overline{x}_E - \overline{x}_U \neq 0$

36. Which **one** of the following statements gives the **best** justification about the validity of using a t-test on these data?

(1) We should be wary about the validity of this test because the sample of Energizer batteries is not independent of the sample of Ultracell batteries.

(2) There are **no** concerns with the validity of this test because both samples have two observations with unusually small playing times.

(3) There are **no** concerns with the validity of this test because, with sample sizes of 9 each, the t-test is sufficiently robust to withstand the departures from Normality suggested by the data.

(4) We should be wary about the validity of this test because of the one unusually large playing time of just over 8.8 hours in the Ultracell data.

(5) There are **no** concerns with the validity of this test because any concerns caused by the negative (left) skew in the Energizer data is offset by the positive (right) skew in the Ultracell data.

CONTINUED

Questions 37 to **40** assume that the use of the t-test is appropriate.
(Note that it may not be.)

37. The value of the test statistic, to 2 decimal places, for this t-test is:

(1) 0.39
(2) 2.55
(3) 0.04
(4) 0.70
(5) 2.13

38. Which **one** of the following statements about the *P-value* of 0.700 is **true**?

(1) When there is a difference between $\mu_E - \mu_U$, then the probability that sampling variation alone would produce sample means that differ by at least 0.03556 is about 0.700.
(2) When there is no difference between $\mu_E - \mu_U$, then the probability that sampling variation alone would produce sample means that differ by at least 0.03556 is about 0.700.
(3) When the sample means differ by 0.03556, then the probability that there is no difference between $\mu_E - \mu_U$ is about 0.700.
(4) When there is no difference between $\mu_E - \mu_U$, then the probability that sampling variation alone would produce sample means that differ by 0.03556 is about 0.700.
(5) When the sample means differ by 0.03556, then the probability that there is a difference between $\mu_E - \mu_U$ is about 0.700.

39. Which **one** of the following interpretations of the t-test is **true**?

The observed difference between the mean playing time for the Energizer batteries and that for the Ultracell batteries is:

(1) sufficiently large to reject a claim that the underlying mean playing times for the two brands are equal.
(2) sufficiently large to claim that Energizer batteries last longer, on average, than Ultracell batteries.
(3) sufficiently large to be statistically significant at the 1% level of significance.
(4) not large enough to be statistically significant at the 5% level of significance.
(5) small enough to claim that there is no difference between the underlying mean playing times for the two brands.

40. A confidence interval for $\mu_E - \mu_U$ is given in Table 7, page 40. Suppose we had to calculate this 95% confidence interval by hand (rather than use the statistical software available to us).

Which **one** of the following statements about the resulting confidence interval would be **true**?

(1) $df = 16$ and be slightly narrower
(2) $df = 8$ and be slightly narrower
(3) $df = 17$ and be slightly wider
(4) $df = 17$ and be slightly narrower
(5) $df = 8$ and be slightly wider

41. Which **one** of the following scenarios would **most likely** produce **selection bias** in a survey?

(1) Using the results from a survey of 1000 randomly selected voters in the Epsom electorate to estimate the proportion of voters in the North Shore electorate who would give their party vote to the National Party.

(2) Using the results from a survey that was in a suburban newspaper delivered to households in the Mangere electorate to estimate the proportion of voters in the Mangere electorate who would give their party vote to the Maori Party.

(3) Using the results from a survey, conducted by the Green Party, of 1000 randomly selected voters in the Auckland Central electorate to estimate the proportion of voters in the Auckland Central electorate who would give their party vote to the Green Party.

(4) Using the results from a survey of 1000 voters in the Botany electorate obtained from a random selection of Botany addresses in the Auckland phone directory to estimate the proportion of voters in the Botany electorate who would give their party vote to the New Zealand First Party.

(5) Using the results from a survey of 1000 randomly selected voters in the Te Atatu electorate, from which 230 responded, to estimate the proportion of voters in the Te Atatu electorate who would give their party vote to the Labour Party.

42. Which **one** of the following is **not** an assumption of the simple linear regression model?

(1) The random errors follow a linear trend.

(2) The random errors are Normally distributed.

(3) The random errors have the same standard deviation, regardless of the value of x.

(4) The random errors have a mean of zero.

(5) The random errors are all independent.

Questions 43 to **48** refer to the information in **Appendix H**, pages 41 and 42.

Questions 43 to **48** assume that the use of a simple linear regression analysis is appropriate. (Note that it may not be.)

43. The equation of the least squares regression line for this analysis is:

(1) Predicted **SBP** = 3.704 + 0.101 × **Age**

(2) Predicted **SBP** = 96.415 + 0.592 × **Age**

(3) Predicted **SBP** = 96.415 + 3.704 × **Age**

(4) Predicted **SBP** = 0.101 + 0.592 × **Age**

(5) Predicted **SBP** = 0.592 + 96.415 × **Age**

44. For two female Maori workers like those in the study whose ages differ by 15 years this regression analysis predicts that the difference in their systolic blood pressures to be approximately:

(1) 103.8 mm Hg

(2) 0.6 mm Hg

(3) 8.9 mm Hg

(4) 6.4 mm Hg

(5) 1.5 mm Hg

45. The residual for the 45 year old female Maori worker who has a systolic blood pressure of 107.5 mm Hg is approximately:

(1) 15.6 mm Hg

(2) −7.8 mm Hg

(3) −15.6 mm Hg

(4) −11.1 mm Hg

(5) 11.1 mm Hg

46. In a test for no linear relationship between **SBP** and **Age**, the hypotheses are:

(1) H_0: $\widehat{\beta}_1 = 0$ H_1: $\widehat{\beta}_1 \neq 0$

(2) H_0: $\beta_1 = 0$ H_1: $\beta_1 \neq 0$

(3) H_0: $\beta_1 \neq 0$ H_1: $\beta_1 = 0$

(4) H_0: $\beta_0 = 0$ H_1: $\beta_0 \neq 0$

(5) H_0: $\widehat{\beta}_0 \neq 0$ H_1: $\widehat{\beta}_0 = 0$

47. Which **one** of the following statements about the test for no linear relationship between **SBP** and **Age** is **true**?

(1) There is very strong evidence against a positive linear association between **SBP** and **Age** and with 95% confidence we estimate that, on average, a 10 year increase in age is associated with an increase in systolic blood pressure of somewhere between 0.4 mm Hg and 0.8 mm Hg.

(2) There is very strong evidence of a positive linear association between **SBP** and **Age** and with 95% confidence we estimate that, on average, a 10 year increase in age is associated with an increase in systolic blood pressure of somewhere between 0.4 mm Hg and 0.8 mm Hg.

(3) There is very strong evidence of a positive linear association between **SBP** and **Age** and with 95% confidence we estimate that, on average, a 10 year increase in age is associated with an increase in systolic blood pressure of somewhere between 3.9 mm Hg and 7.9 mm Hg.

(4) There is evidence of a very strong positive linear association between **SBP** and **Age** and with 95% confidence we estimate that, on average, a 10 year increase in age is associated with an increase in systolic blood pressure of somewhere between 3.9 mm Hg and 7.9 mm Hg.

(5) There is evidence of a very strong positive linear association between **SBP** and **Age** and with 95% confidence we estimate that, on average, a 10 year increase in age is associated with an increase in systolic blood pressure of somewhere between 0.4 mm Hg and 0.8 mm Hg.

CONTINUED

48. Which **one** of the following statements is **true**?

With 95% confidence, it is estimated that for:

(1) female Maori workers like those in the study who are 53 years old their mean systolic blood pressure is somewhere between 123.7 mm Hg and 131.9 mm Hg.

(2) female Maori workers like those in the study who are 20 years old their mean systolic blood pressure is somewhere between 81.9 mm Hg and 134.6 mm Hg.

(3) a female Maori worker like those in the study who is 36 years old her systolic blood pressure is somewhere between 115.7 mm Hg and 119.8 mm Hg.

(4) female Maori workers like those in the study who are 29 years old their mean systolic blood pressure is somewhere between 87.4 mm Hg and 139.8 mm Hg.

(5) a female Maori worker like those in the study who is 43 years old her systolic blood pressure is somewhere between 119.3 mm Hg and 124.4 mm Hg.

49. Which **one** of the following statements about the sample correlation coefficient, r, is **false**?

(1) The sample correlation coefficient, r, changes sign if the two axes are swapped around—a positive sample correlation becomes negative and vice versa.

(2) The sample correlation coefficient, r, tries to answer the question "how close do these points in the scatter plot come to falling on a straight line?"

(3) Outliers can inflate the value of a sample correlation coefficient, r.

(4) It only makes sense to interpret a sample correlation coefficient, r, for two numeric variables if the overall pattern seen in a scatter plot of the two variables can be described as a straight line.

(5) Outliers can deflate the value of a sample correlation coefficient, r.

CONTINUED

50. Figure 1 shows data from three hypothetical studies. In each study, each of three independent groups of subjects was given a different treatment. The treatment groups correspond to Group 1, Group 2 and Group 3.

Suppose that a one-way analysis of variance (ANOVA) F-test was conducted for each of the studies, to investigate whether there were any differences between the three groups.

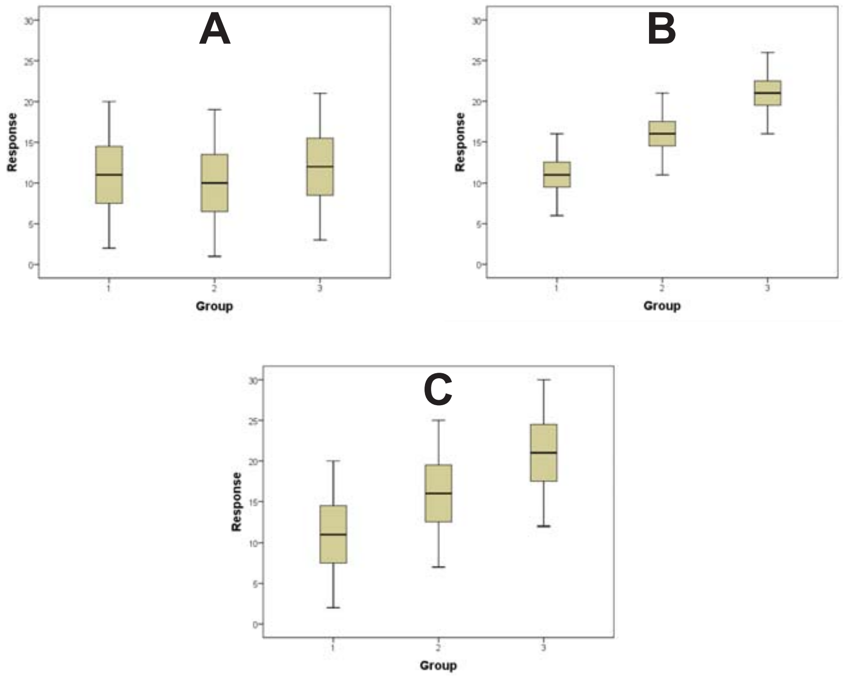

Figure 1: Box plots for three studies

Consider the F-test statistic for each study. Which **one** of the following statements is **true**?

(1) B has the largest test statistic and C has the smallest.

(2) B has the largest test statistic and A has the smallest.

(3) C has the largest test statistic and A has the smallest.

(4) A has the largest test statistic and C has the smallest.

(5) A has the largest test statistic and B has the smallest.

ATTACHMENT FOLLOWS

INCLUSIONS:

* **Appendix A: Wolves Data** for use in Questions **2** to **4**
* **Appendix B: Ice Slurry Data** for use in Questions **5** to **8**
* **Appendix C: Anchoring Data** for use in Questions **9** to **12**
* **Appendix D: Alcohol-related Offences Data** for use in Questions **13** to **23**
* **Appendix E: Bullying Data** for use in Questions **24** to **27**
* **Appendix F: Olympics Data** for use in Questions **28** to **34**
* **Appendix G: Battery Data** for use in Questions **35** to **40**
* **Appendix H: Blood Pressure Data** for use in Questions **43** to **48**

References

Dunn, P. K. (2013). Comparing the lifetimes of two brands of batteries. *Journal of Statistics Education*, **21**(1), 1–19.

Jolicoeur, P. (1959). Multivariate geographical variation in the wolf canis lupus l. *Evolution*, **13**, 283–299.

MacMahon, S., Norton, R., Jackson, R., Mackie, M. J., Cheng, A., Vander Hoorn, S., Milne, A., and McCulloch, A. (1995). Fletcher Challenge-University of Auckland Heart & Health Study: Design and baseline findings. *New Zealand Medical Journal*, **108**, 499–502.

MOT (2010). *Motor vehicle crashes in New Zealand 2009, Statistical Statement Calendar Year 2009*. Reseach and Statistics, Ministry of Transport.

Siegel, R., Mate, J., Brearley, M. B., Watson, G., Nosaka, K., and Laursen, P. B. (2010). Ice slurry ingestion increases core temperature capacity and running time in the heat. *Medicine and Science in Sports and Exercise*, **42**(4), 717–725.

Utts, J. M. and Heckard, R. F. (2014). *Mind on Statistics*. 5th edition.

Voss, L. D. and Mulligan, J. (2000). Bullying in school: are short pupils at risk? Questionnaire study in a cohort. *British Medical Journal*, **320**, 612–613.

CONTINUED

Appendix A: Wolves Data

Questions 2 to **4** refer to the information in this appendix.

Jolicoeur (1959) conducted a study on wolves from two regions in North America: the Arctic and the Rocky Mountains.

Some of the variables used in the study were:

Location	The region where a wolf was found – Arctic, Rocky Mountains
Sex	The sex of a wolf – Female, Male
Braincase	The least width of a wolf's braincase, in millimetres
Crown	The crown length of a first upper molar (a tooth), in millimetres

Assume the study involves two random samples: a random sample of wolves from the population of all Arctic wolves and a random sample of wolves from the population of all Rocky Mountains wolves.

Questions 3 and **4** refer to the following additional information.

A key purpose of the study was to explore differences in the crown length of the first upper molar for male wolves from the Arctic and Rocky Mountains regions. A bootstrap confidence interval for the difference between the two population means was constructed.

Two plots from the associated VIT output are shown in Figures 2 and 3. Figure 2 shows the data from the study and Figure 3 shows the bootstrap distribution and the bootstrap confidence interval.

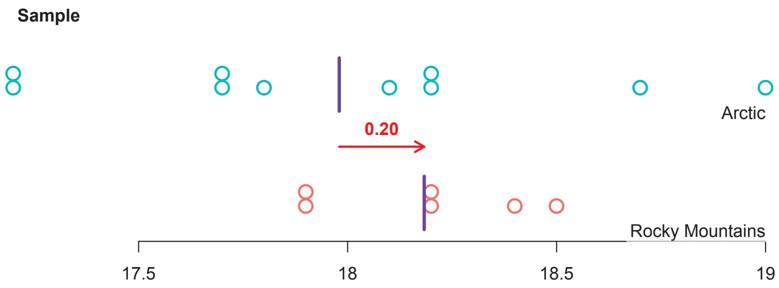

Figure 2: Male wolves' crown lengths (millimetres)

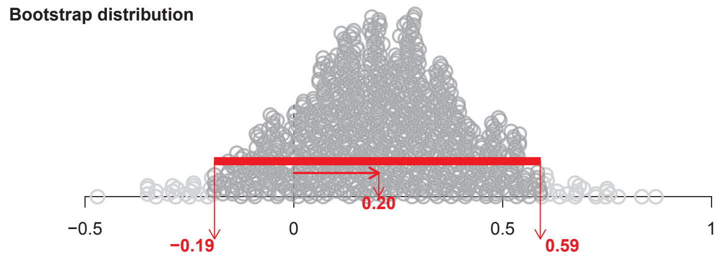

Figure 3: Bootstrap distribution and confidence interval

Appendix B: Ice Slurry Data

Questions 5 to **8** refer to the information in this appendix.

The effects, on the time taken to run to exhaustion, of drinking ice slurry (at −1°C) shortly before exercise were compared with drinking cold water (at 4°C) shortly before exercise (Siegel *et al.*, 2010). The subjects were ten healthy male volunteers who were participating in recreational sport and considered moderately active.

The study consisted of two trials. Five subjects were randomly assigned to drink ice slurry in the first trial and cold water in the second. For the remaining five subjects the order in which each type of liquid was to be drunk was reversed: cold water in the first trial and ice slurry in the second. The second trial was conducted about ten days after the first. In both trials, the time taken to run to exhaustion was recorded in minutes to one decimal place.

A paired-data t-procedure was conducted. Dot plots are shown in Figures 4 and 5 below and output is shown in Table 1 on page 30.

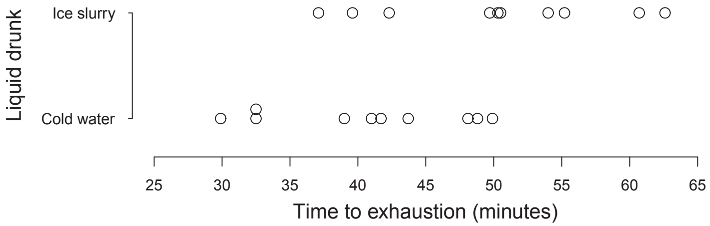

Figure 4: Time taken to run to exhaustion for both liquids

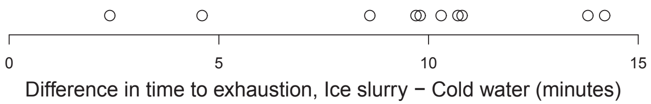

Figure 5: Difference in time taken to run to exhaustion

CONTINUED

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Ice slurry	50.200	10	8.4995	2.6878
	Cold water	40.710	10	7.2196	2.2831

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Ice slurry - Cold water	9.4900	3.6471	++	6.8811	12.0989	8.229	9	.000

Note: One value has been repleced by ++.

Table 1: Paired-data t-procedure output

CONTINUED

Appendix C: Anchoring Data

Questions 9 to **12** refer to the information in this appendix.

An experiment, described in Utts and Heckard (2014), was conducted to investigate whether the answers to a question may be influenced by information given with, or in, the question. Such an influence is called an 'anchoring effect'. It has been suggested that presenting a larger number with, or in, a question would generally produce larger values in the responses to the question. In the experiment 21 graduate Statistics students at a university in the United States were randomly allocated to one of two groups; Group 1 (11 students), Group 2 (10 students).

Group 1 was asked:

> **Anchor 1:** *The population of Australia is about 18 million. To the nearest million, what do you think is the population of Canada?*

Group 2 was asked:

> **Anchor 2:** *The population of the U.S. is about 270 million. To the nearest million, what do you think is the population of Canada?*

Figure 6 shows the distribution of the responses for both groups and the difference between the median responses for the two groups.

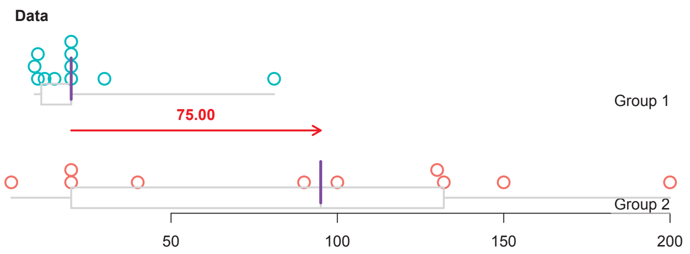

Figure 6: Population of Canada responses (millions)

Let:

> $\widetilde{\mu}_1$ be the median response if all 21 students in the study had responded to **Anchor 1**

and

> $\widetilde{\mu}_2$ be the median response if all 21 students in the study had responded to **Anchor 2**.

A one-tailed randomisation test was conducted with:

$$H_0\colon\ \widetilde{\mu}_2 - \widetilde{\mu}_1 = 0 \text{ and } H_1\colon\ \widetilde{\mu}_2 - \widetilde{\mu}_1 > 0$$

The re-randomisation distribution, including the tail proportion, is shown in Figure 7.

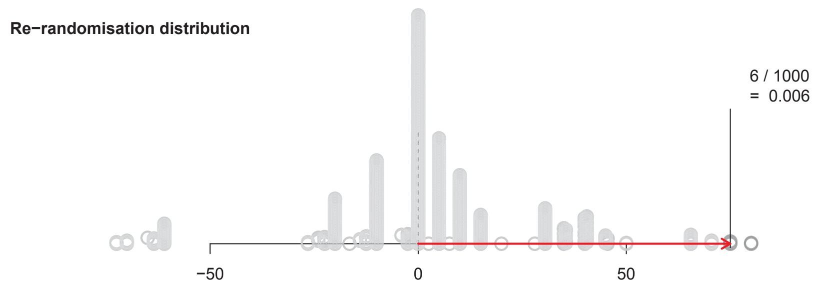

Figure 7: Randomisation test output

Appendix D: Alcohol-related Offences Data

Questions 13 to **23** refer to the information in this appendix.

In New Zealand the Ministry of Transport collects a wide range of traffic-related data, including alcohol-related offences (MOT, 2010).

Two of the variables recorded for alcohol-related offences are:

Day The day of an offence
– Mon, Tue, Wed, Thu, Fri, Sat, Sun

TimePeriod The time period within the day of an offence
– Morning (from 3:00am to 8:59am)
– Midday (from 9:00am to 2:59pm)
– Evening (from 3:00pm to 8:59pm)
– Night (from 9:00pm to 2:59am)

Note that offences classified as occurring in the 'Night' period of a Monday, say, actually occurred somewhere between 9:00pm on Monday and 2:59am on the next day (Tuesday).

Table 2 shows the number of alcohol-related driving offences in New Zealand in 2009 identified as a result of breath screening drivers. The 34,791 offences were cross-classified according to **Day** and **TimePeriod**. Assume that these offences form a random sample of alcohol-related driving offences in New Zealand.

	TimePeriod				
Day	Morning	Midday	Evening	Night	**Total**
Mon	137	63	314	562	1076
Tue	131	111	509	725	1476
Wed	173	125	786	1463	2547
Thu	565	218	1277	3105	5165
Fri	913	209	1399	4963	7484
Sat	1621	312	1388	6644	9965
Sun	1924	323	728	4103	7078
Total	5464	1361	6401	21565	34791

Table 2: Alcohol-related offences data

A Chi-square test for independence was conducted on the data. The results of this test are shown in Table 3.

			TimePeriod				
			Morning	Midday	Evening	Night	Total
Day	Mon	Count	137	63	314	562	1076
		Expected count	168.99	42.09	197.97	666.95	1076.00
		Cell contribution	6.06	10.38	68.01	16.52	
	Tue	Count	131	111	509	725	1476
		Expected count	231.81	57.74	271.56	914.89	1476.00
		Cell contribution	43.84	49.13	207.60	39.41	
	Wed	Count	173	125	786	1463	2547
		Expected count	400.01	99.64	468.61	1578.74	2547.00
		Cell contribution	128.83	++	214.97	8.49	
	Thu	Count	565	218	1277	3105	5165
		Expected count	++	202.05	++	3201.50	5165.00
		Cell contribution	74.71	1.26	112.33	2.91	
	Fri	Count	913	209	1399	4963	7484
		Expected count	1175.38	292.77	1376.94	4638.91	7484.00
		Cell contribution	58.57	++	0.35	22.64	
	Sat	Count	1621	312	1388	6644	9965
		Expected count	++	389.82	++	6176.75	9965.00
		Cell contribution	2.00	15.54	108.21	35.35	
	Sun	Count	1924	323	728	4103	7078
		Expected count	1111.61	276.89	1302.24	4387.26	7078.00
		Cell contribution	593.70	7.68	253.22	18.42	
Total		Count	5464	1361	6401	21565	34791
		Expected count	5464.00	1361.00	6401.00	21565.00	34791.00

Chi-Square Tests

	Value	df	Sig.
Pearson Chi-Square	2130.548	++	.000
Likelihood Ratio	2032.517	++	.000
N of Valid Cases	34791		

Note: Some values have been replaced by ++.

Table 3: Chi-square test output

Appendix E: Bullying Data

Questions 24 to **27** refer to the information in this appendix.

In a United Kingdom study conducted on school pupils by Voss and Mulligan (2000), 92 short (in height) pupils and 117 not-short pupils (the controls) completed a questionnaire on bullying. The pupils were 13 to 15 years old and the controls were selected so that the distributions of age and sex were similar to the distributions of the short pupils. The short pupils had been below the 30th percentile for height when they started school.

Table 4 summarises the responses from the pupils as to whether or not they had been bullied at some time at secondary school.

	Short			**Not-short (controls)**			**Total**		
	Boy	Girl	Total	Boy	Girl	Total	Boy	Girl	Total
Bullied	25	17	42	13	17	30	38	34	72
Not bullied	29	21	50	50	37	87	79	58	137
Total	54	38	92	63	54	117	117	92	209

Table 4: Bullying of short and not-short pupils

Assume that the use of *t*-procedures is appropriate.

Appendix F: Olympics Data

Questions 28 to **34** refer to the information in this appendix.

One of the feature events in the Summer Olympic Games is the men's 100 metres in which the runners place their feet on starting blocks to help give them a quick start when the starting gun is fired.

A sport researcher was interested in comparing reaction times in the men's 100 metres in the four most recent Olympic Games (London, 2012; Beijing, 2008; Athens, 2004; Sydney, 2000). Some runners are eliminated after the first round so data was collected on the runners in the first round only.

Some of the variables measured were:

Reaction Time The time between the firing of the starting gun and a runner leaving the starting blocks, in seconds (to three decimal places)

Finishing Time The time for a runner to complete the 100 m race, in seconds (to two decimal places)

Year The year in which an Olympic Games was held – 2000, 2004, 2008, 2012

A one-way analysis of variance (ANOVA) F-test was conducted to see whether mean reaction times depend on the Olympic Games year. Box plots are shown in Figure 8 below and test output is shown in Table 5, page 38.

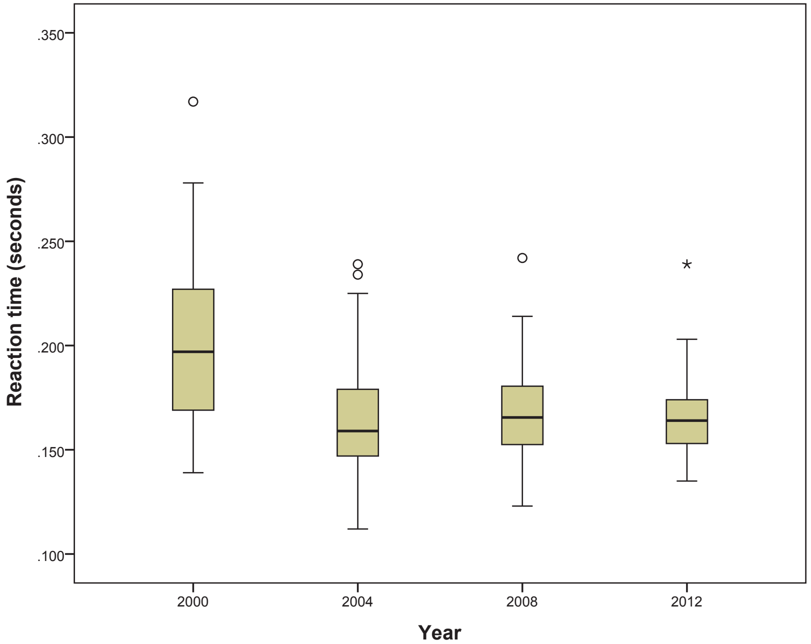

Figure 8: Reaction times data

Descriptives

Reaction time (seconds)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
2000	97	.20022	.036845	.003741	.19279	.20764	.139	.317
2004	81	.16514	.025327	.002814	.15954	.17074	.112	.239
2008	80	.16760	.021111	.002360	.16290	.17230	.123	.242
2012	73	.16532	.016784	.001964	.16140	.16923	.135	.239
Total	331	.17605	.031021	.001705	.17270	.17941	.112	.317

ANOVA

Reaction time (seconds)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.080423	3	.026808	36.968	.000
Within Groups	.237127	327	.000725		
Total	.317550	330			

Multiple Comparisons

Dependent Variable: Reaction time (seconds)

Tukey HSD

(I) Year	(J) Year	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
2000	2004	.035081	.004053	.000	.02461	.04555
	2008	.032616	.004067	.000	.02211	.04312
	2012	.034901	.004172	.000	.02413	.04568
2004	2000	-.035081	.004053	.000	-.04555	-.02461
	2008	-.002464	.004245	.938	-.01343	.00850
	2012	-.000179	.004346	1.000	-.01140	.01104
2008	2000	-.032616	.004067	.000	-.04312	-.02211
	2004	.002464	.004245	.938	-.00850	.01343
	2012	.002285	.004359	.953	-.00897	.01354
2012	2000	-.034901	.004172	.000	-.04568	-.02413
	2004	.000179	.004346	1.000	-.01104	.01140
	2008	-.002285	.004359	.953	-.01354	.00897

Table 5: F-test for one-way analysis of variance (ANOVA) output

Appendix G: Battery Data

Questions 35 to **40** refer to the information in this appendix.

A study described by Dunn (2013) looked at the performance of two brands of 1.5 volt batteries, Energizer Max AA Alkaline (Energizer) and Ultracell AA Alkaline (Ultracell). Nine batteries of each brand were used to play a 250 mA electronic game. The playing time taken, in hours, for the battery voltage to fall to 0.9 volts was recorded.

Figure 9 shows the distribution of the playing times to fall to 0.9 volts for each brand of battery.

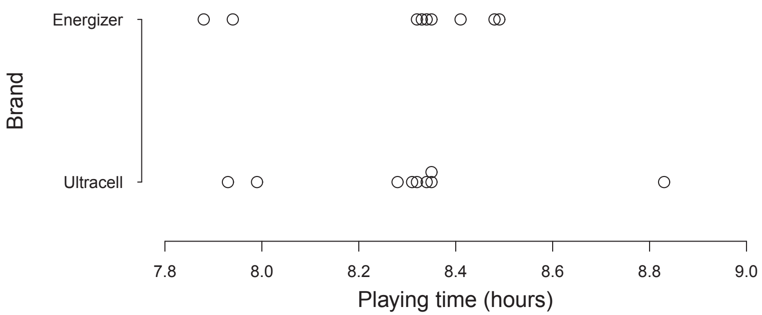

Figure 9: Time to fall to 0.9 volts for both brands

Assume that the samples of each brand are random samples.

Let:

μ_E be the underlying mean playing time to fall to 0.9 volts for **Energizer** Max AA Alkaline 1.5 volt batteries

and

μ_U be the underlying mean playing time to fall to 0.9 volts for **Ultracell** AA Alkaline 1.5 volt batteries.

A two-tailed two independent sample *t*-test was conducted on the data. Output is shown in Tables 6 and 7.

Group Statistics

	Brand	N	Mean	Std. Deviation	Std. Error Mean
Time	Energizer	9	8.2789	.21740	.07247
	Ultracell	9	8.2433	.16279	.05426

Table 6: Summary statistics

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Time	Equal variances assumed	.507	.487	++	16	.700	.03556	.09053	-.15636	.22747
	Equal variances not assumed			++	14.825	.700	.03556	.09053	-.15760	.22871

Note: Two values have been replaced with ++.

Table 7: *t*-test output

Appendix H: Blood Pressure Data

Questions 43 to **48** refer to the information in this appendix.

As part of a study, MacMahon *et al.* (1995) collected data on 163 Maori female workers who worked for a large New Zealand company.

Two of the variables recorded for each woman were:

SBP Systolic blood pressure (in mm Hg)

Age Age (in years)

A scatter plot of **SBP** against **Age** is shown in Figure 10.

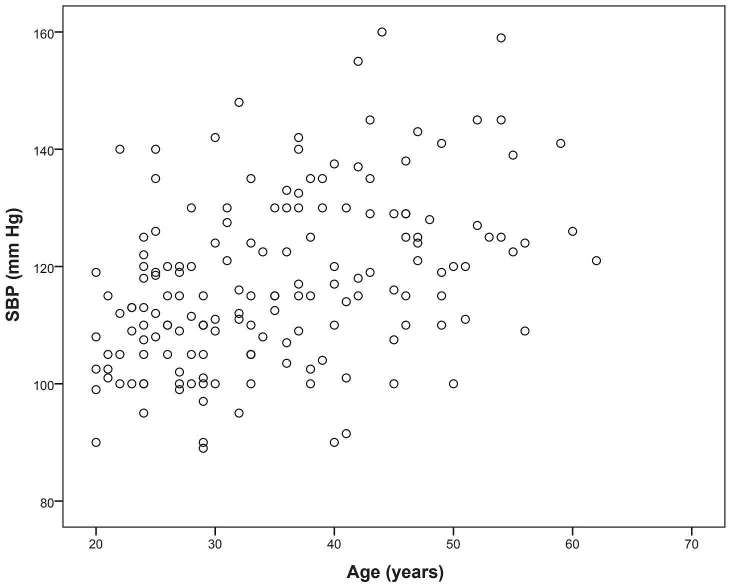

Figure 10: **SBP** against **Age**

CONTINUED

A simple linear regression analysis was carried out using the 163 observations. Some output from this analysis is shown in Tables 8 and 9.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	96.415	3.704		26.033	.000	89.101	103.729
	Age	.592	.101	.419	5.855	.000	.392	.792

a. Dependent Variable: SBP

Table 8: Regression output

	Age	SBP	LMCI_1	UMCI_1	LICI_1	UICI_1
1	20	119.0	104.60337	111.90981	81.93663	134.57655
2	36	122.5	115.68135	119.77789	91.58408	143.87517
3	53	125.0	123.69090	131.89854	101.40845	154.18100
4	43	145.0	119.30279	124.44536	95.68237	148.06579
5	29	110.0	111.20046	115.96988	87.41112	139.75923

Table 9: Confidence and prediction intervals

ANSWERS:

2. (5)	3. (3)	4. (1)	5. (2)	6. (4)
8. (2)	9. (2)	10. (5)	11. (4)	12. (3)
13. (1)	14. (1)	15. (5)	16. (5)	17. (5)
18. (4)	19. (3)	20. (2)	21. (4)	22. (3)
23. (4)	24. (2)	25. (2)	26. (5)	27. (1)
28. (1)	29. (3)	30. (2)	31. (4)	32. (5)
33. (5)	34. (1)	35. (3)	36. (4)	37. (1)
38. (2)	39. (4)	40. (5)	41. (4)	42. (1)
43. (2)	44. (3)	45. (3)	46. (2)	47. (3)
48. (1)	49. (1)	50. (2)		