THE UNIVERSITY OF AUCKLAND

STATS 101
STATS 101G
STATS 108

SUMMER SEMESTER, 2014
Campus: CITY

VERSION 1

STATISTICS

Introduction to Statistics
Statistics for Commerce

(Time allowed: THREE hours)

ANSWERS ON PAGE 38

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: E-EPA Data** pages 24 to 26
* **Appendix B: Driving Behaviour Data** pages 27 to 29
* **Appendix C: Pain Data** pages 30 to 35
* **Appendix D: Testosterone Data** pages 36 and 37

Questions 1 and **2** refer to the following information.

Based on information from the Real Estate Institute of New Zealand, Table 1 shows the median price and the number of sales for houses in the Auckland region for September 2012 and September 2013. It also includes the percentage changes over this time period.

District	**Median Price**			**Number**		
	Sep-13	Sep-12	% Change	Sep-13	Sep-12	% Change
North Shore	$715,000	$628,000	13.9	505	469	7.7
Waitakere	$531,000	$429,500	23.6	374	324	15.4
Central Auckland	$640,000	$595,000	7.6	874	765	14.2
Manukau	$490,000	$485,000	1.0	557	376	48.1
Rodney	$531,000	$488,500	8.7	173	144	20.1
Outer Auckland	$468,000	$428,000	9.3	339	257	31.9

Table 1: Auckland region house sales

1. Suppose the main purpose of Table 1 is to compare the percentage change in median price from September 2012 to September 2013 across the six districts.

Which **one** of the following would **most** improve the presentation of the table for this purpose?

(1) Round the values of % Change within **Median Price** to the nearest whole number.

(2) Keep the order of the table as it stands but put the information about **Median Price** to the right of the information about **Number**.

(3) List the information in the table in ascending (or descending) order of % Change within **Median Price**.

(4) List the information in the table in alphabetical order of **District** and swap the % Change column within **Median Price** so that it comes in the first column.

(5) Keep the order of the table as it stands but swap the % Change column within **Median Price** so that it comes in the first column.

2. Based on the information in Table 1, page 2, four of the following statements are **definitely true**.

Which **one** of the following statements **may not be true**?

(1) Both the median house price and the number of house sales have increased for all districts in Auckland from September 2012 to September 2013.

(2) North Shore has the highest median house price in both September 2012 and September 2013 but the lowest percentage change in number of house sales over the same time period.

(3) The median of the six median house prices for each of the districts in September 2013 would give us the median house price for all of the Auckland region.

(4) Although Manukau has the lowest percentage change in median house price, it has the highest percentage change in number of house sales.

(5) The district with the highest percentage change in median house price is Waitakere.

3. Which **one** of the following statements about confidence intervals for a single population proportion, produced using the Student's t-distribution, is **false**?

(1) Calculated from the same sample data, the 95% confidence interval will be narrower than the 99% confidence interval.

(2) The size of the t-multiplier used to calculate the confidence interval depends on the desired level of confidence.

(3) The mid-point of the 95% confidence interval is the sample proportion.

(4) The margin of error is the distance between the upper limit and the lower limit of the confidence interval.

(5) A 95% confidence interval does not always include the true value of the population proportion.

Questions 4 to **9** refer to the information in **Appendix A**, pages 24 to 26.

4. Which **one** of the following statements gives the **best** explanation for why the assignment of the 20 patients to one of the two groups was done randomly?

The assignment of the 20 patients to one of the two groups was done randomly:

(1) to try to increase the accuracy of the results.

(2) to try to ensure that, overall, the E-EPA and placebo groups would each be representative of some larger group.

(3) so that any sampling error would be minimised.

(4) so that any non-sampling error would be minimised.

(5) to try to ensure that, overall, patients in the E-EPA group would be similar to patients in the placebo group.

Questions 5 and **6** refer to the information on page 25.

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

(1) If the experiment was repeated 1000 times, then about 18 out of these 1000 re-runs would produce an observed difference between the group means of somewhere between 1.5 and 7.4.

(2) 18 out of 1000 re-samples with replacement produced a difference between the group means of 5.9 or more.

(3) 18 out of 1000 re-randomisations under chance alone produced a difference between the group means of 5.9 or more.

(4) 18 out of 1000 re-randomisations under chance alone produced a difference between the group means of somewhere between 1.5 and 7.4.

(5) If the experiment was repeated 1000 times, then about 18 out of these 1000 re-runs would produce an observed difference between the group means of 5.9 or more.

6. Which **one** of the following statements is **not** a valid interpretation of the randomisation test result?

(1) It may be concluded that that the observed difference in decreases in Hamilton depression ratings between the two groups is the result of the effect of E-EPA together with some chance effect.

(2) Chance is acting alone is not a plausible explanation for the observed difference between the two group means.

(3) For patients in this study, it may be claimed that, on average, E-EPA had the effect of reducing their Hamilton depression ratings.

(4) An observed difference between the two group means of 5.9 or more is unlikely when chance is acting alone.

(5) The observed difference between the group means is consistent with there being no difference in decreases in Hamilton depression ratings between the two groups.

Questions 7 to **9** refer to the information on page 26.

7. **No longer examined**

8. **No longer examined**

9. **No longer examined**

Questions 10 to **22** refer to the information in **Appendix B**, pages 27 to 29.

Use Table 4, page 27, to answer **Questions 10** and **11**.

10. The percentage of the drivers who gave way to the pedestrian, to the nearest whole percentage, is:

(1) 69%
(2) 76%
(3) 35%
(4) 55%
(5) 65%

11. Of the drivers, what proportion were driving a medium status vehicle and gave way to the pedestrian?

(1) 0.276
(2) 0.401
(3) 0.776
(4) 0.424
(5) 0.689

Questions 12 to **15** assume that the 152 subjects in the first study are a random sample of drivers and vehicles in the San Francisco area.

Questions 12 and **13** refer to the following additional information.

For drivers of **medium** status vehicles in the San Francisco area, let:

p_{MY} be the underlying proportion who give way to pedestrians on uncontrolled pedestrian crossings,

and for drivers of **high** status vehicles in the San Francisco area, let:

p_{HY} be the underlying proportion who give way to pedestrians on uncontrolled pedestrian crossings.

12. The sampling situation associated with $\text{se}(\widehat{p}_{\text{MY}} - \widehat{p}_{\text{HY}})$ can be described as:

(1) one single sample of size 152, several response categories.

(2) two independent samples, one of size 61 and one of size 58.

(3) one single sample of size 152, many yes/no items.

(4) two independent samples, one of size 42 and one of size 32.

(5) one single sample of size 99, several response categories.

13. A 95% confidence interval for $p_{\text{MY}} - p_{\text{HY}}$ is $(-0.0361, 0.3097)$.

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

(1) It would be very surprising to see a different sample of the same size produce a result with $\widehat{p}_{\text{MY}}$ smaller than $\widehat{p}_{\text{HY}}$.

(2) At the 5% level of significance, it may not be claimed that p_{MY} is larger than p_{HY}.

(3) At the 5% level of significance, it is plausible that the observed difference, 0.1368, between $\widehat{p}_{\text{MY}}$ and $\widehat{p}_{\text{HY}}$ could be due to sampling error alone.

(4) The observed difference, 0.1368, between $\widehat{p}_{\text{MY}}$ and $\widehat{p}_{\text{HY}}$ is not significant at the 1% level of significance.

(5) The observed difference, 0.1368, between $\widehat{p}_{\text{MY}}$ and $\widehat{p}_{\text{HY}}$ is not significant at the 5% level of significance.

Questions 14 and **15** refer to the information on page 28.

14. **No longer examinable**

15. **No longer examinable**

Questions 16 to **22** refer to the information on page 29.

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

(1) There is no cause for concern because only one cell has an observed count less than 5 and all observed counts are greater than 1.

(2) There is no cause for concern because no cells have an expected count less than 5.

(3) There is concern because at least one cell has a contribution to the Chi-square test statistic that is less than 1.

(4) There is concern because, although all cells have an observed count greater than 1, one of them is less than 5.

(5) There is concern because one cell has an expected count less than 10.

Questions 17 to **22** assume that the Chi-square test is valid.
(Note: This may not be true.)

17. Consider the cell in Table 7, page 29, for drivers who cut in front of another vehicle and were driving a medium status vehicle. Under the null hypothesis, the expected count for this cell is approximately:

(1) 14.146

(2) 57.000

(3) 11.333

(4) 45.667

(5) 15.000

18. Consider the cell in Table 7, page 29, for drivers who cut in front of another vehicle and were driving a low status vehicle. This cell's contribution to the Chi-square test statistic, to 2 decimal places, is:

(1) 2.76

(2) 0.78

(3) 1.51

(4) 0.21

(5) 0.45

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

Under the null hypothesis, the proportion of these drivers who did not cut in front of another vehicle at the intersection is expected to be about:

(1) $\frac{240}{274}$ for all three levels of **Status**.

(2) $\frac{1}{3}$ for all three levels of **Status**.

(3) $\frac{55}{59}$, $\frac{99}{114}$ and $\frac{86}{101}$ for the respective levels of **Status**.

(4) $\frac{55}{240}$, $\frac{99}{240}$ and $\frac{86}{240}$ for the respective levels of **Status**.

(5) $\frac{59}{274}$, $\frac{114}{274}$ and $\frac{101}{274}$ for the respective levels of **Status**.

20. Based on the results of this Chi-square test, which **one** of the following statements is **true**?

(1) If the alternative hypothesis is true, then the probability of getting a test statistic at least as large as 2.333 is approximately 0.311.

(2) If the null hypothesis is true, then the probability of getting a test statistic of 2.333 or smaller is approximately 0.311.

(3) The probability that the alternative hypothesis is false is approximately 0.311.

(4) The probability that the null hypothesis is true is approximately 0.311.

(5) If the null hypothesis is true, then the probability of getting a test statistic at least as large as 2.333 is approximately 0.311.

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

(1) There is no evidence that the variables **Cut in front** and **Status** are related.

(2) There is no evidence of a link between the variables **Cut in front** and **Status**.

(3) There is no evidence that the underlying distribution of **Cut in front** is the same for each level of **Status**.

(4) There is no evidence that the variables **Cut in front** and **Status** are not independent.

(5) There is no evidence that the underlying distribution of **Status** is different for each level of **Cut in front**.

22. Which **one** of the following conclusions may **not** be made, based on the results of this Chi-square test?

(1) Considering drivers of low status vehicles who did not cut in front of another vehicle, the difference between the observed count of 55 and the expected count of 51.679 is small enough that it could be due to sampling variation alone.

(2) The difference between the proportion of the drivers who did not cut in front of another vehicle that were driving a low status vehicle, $\frac{55}{240}$, and the proportion of the drivers who did not cut in front of another vehicle that were driving a medium status vehicle, $\frac{99}{240}$, could be due to sampling variation alone.

(3) The difference between the proportion of the drivers of low status vehicles who cut in front of another vehicle, $\frac{4}{59}$, and the proportion of the drivers of medium status vehicles who cut in front of another vehicle, $\frac{15}{114}$, could be due to sampling variation alone.

(4) The difference between the proportion of the drivers who cut in front of another vehicle that were driving a medium status vehicle, $\frac{15}{34}$, and the proportion of the drivers who did not cut in front of another vehicle that were driving a medium status vehicle, $\frac{99}{240}$, could be due to sampling variation alone.

(5) Considering drivers of high status vehicles who did not cut in front of another vehicle, the difference between the observed count of 86 and the expected count of 88.467 is small enough that it could be due to sampling variation alone.

23. Garcia-Retamero and Hoffrage (2013) used the following information in their research on physicians' understanding of probability.

> For a woman at age 40 who participates in routine screening, the probability of breast cancer is 1%. If a woman has breast cancer, the probability is 80% that she will have a positive mammogram. If a woman does not have breast cancer, the probability is 10% that she will have a positive mammogram.

Based on this, for 40-year-old women who have a positive mammogram, approximately what percentage will have breast cancer?

(1) 7.5%
(2) 0.8%
(3) 80.0%
(4) 47.1%
(5) 23.7%

Questions 24 to **44** refer to the information in **Appendix C**, pages 30 to 35.

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

(1) Blocking has not been used in this study.

(2) Randomising the order of the type of word to be repeated was used to control for any order effect on the two treatments.

(3) The subjects were not blinded because they would have known which type of word they were repeating.

(4) The use of a placebo would have improved this study because the study units are people.

(5) This study is an example of an experiment even though there are not two distinct treatment groups.

Questions 25 to **27** refer to the information on page 31.

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

(1) It is critical that the measurements for **SwearT** and **NonSwearT** for any one female participant are independent of the measurements for **SwearT** and **NonSwearT** for all other female participants.

(2) For each female participant, it is critical that the measurement for **SwearT** is independent of the measurement for **NonSwearT**.

(3) The underlying distribution of the differences, **SwearT** – **NonSwearT**, is assumed to have a Normal distribution.

(4) It is critical that the differences, obtained from **SwearT** – **NonSwearT** for each female participant, are independent of each other.

(5) The test is equivalent to a one sample t-procedure on the differences, obtained from **SwearT** – **NonSwearT**.

Questions 26 and **27** assume that the paired-data t-procedure is valid.
(Note: This may not be true.)

26. **Lm jml ecp cv_kglcb**

27. Which **one** of the following statements is a **correct** interpretation of the confidence interval in Table 8, page 31?

With 95% confidence, it is estimated that:

(1) a female's hand-immersion time when repeating a swear word is, on average, somewhere between 25.3 and 48.7 seconds longer than her time when repeating a non-swear word.

(2) the hand-immersion time for the 29 female participants when repeating a swear word is, on average, somewhere between 25.3 and 48.7 seconds longer than their time when repeating a non-swear word.

(3) a female's hand-immersion time when repeating a swear word is somewhere between 25.3 and 48.7 seconds longer than her time when repeating a non-swear word.

(4) a female's hand-immersion time when repeating a swear word is, on average, somewhere between 25.3 and 48.7 seconds shorter than her time when repeating a non-swear word.

(5) the hand-immersion time for the 29 female participants when repeating a swear word is, on average, somewhere between 25.3 and 48.7 seconds shorter than their time when repeating a non-swear word.

Questions 28 and **29** refer to the information on page 32.

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

(1) Each dot in the bootstrap distribution plot represents the median of a re-sample obtained by randomly re-sampling with replacement from the 29 differences in hand-immersion times shown in the Sample plot.

(2) The bootstrap distribution estimates the extent of the variation in the sample median difference in hand-immersion times, which allows a bootstrap confidence interval to be determined.

(3) We would not expect the bootstrap distribution to be similar to the distribution of the 29 differences in hand-immersion times shown in the Sample plot.

(4) The bootstrap distribution shows the extent of the variation of the medians of 1000 random re-samples taken with replacement from the 29 differences in hand-immersion times shown in the Sample plot.

(5) The bootstrap distribution plot displays the distribution of the differences in hand-immersion times for the underlying population of females.

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

(1) It's a fairly safe bet that repeating a swear word, compared to a non-swear word, has the effect of increasing hand-immersion times for females by somewhere between 25 and 37 seconds.

(2) It is not certain that the true value of the median difference in hand-immersion times (**SwearT** – **NonSwearT**) for females will be somewhere from 25 seconds to 37 seconds but there is a very good chance that it is.

(3) It's a fairly safe bet that the time a female's hand is immersed when repeating a swear word is, on average, somewhere between 25 and 37 seconds longer than her time when repeating a non-swear word.

(4) If another person used iNZightVIT software to construct a bootstrap confidence interval for the median using the same sample of 29 differences then it would not be surprising if its limits were slightly different from 25 and 37 seconds.

(5) Any time in the interval from 25 seconds to 37 seconds may be interpreted as a plausible value for the median difference in hand-immersion times (**SwearT** – **NonSwearT**) for females.

Questions 30 to **35** refer to the information on page 33.

30. Based on the scatter plot in Figure 6, page 33, which **one** of the following statements is **false**?

(1) For any two female participants, the one with the longer hand-immersion time when repeating a non-swear word did not always have the longer hand-immersion time when repeating a swear word.

(2) The female who had the longest hand-immersion time when repeating a non-swear word had the longest hand-immersion time when repeating a swear word.

(3) The relationship between **SwearT** and **NonSwearT** is moderate to strong.

(4) An increase in hand-immersion time when repeating a swear word tends to be associated with an increase in hand-immersion time when repeating a non-swear word.

(5) Of the eight females who had hand-immersion times when repeating a non-swear word between 60 and 120 seconds, four of them had hand-immersion time when repeating a swear word considerably longer than the other four.

Questions 31 to **35** assume that the simple linear regression analysis on **SwearT** and **NonSwearT** is valid.
(Note: This may not be true.)

31. For females like those in the study whose hand-immersion times when repeating a non-swear word is 10 seconds this regression analysis predicts their mean hand-immersion time when repeating a swear word to be approximately:

(1) 12 seconds.

(2) 40 seconds.

(3) 182 seconds.

(4) 193 seconds.

(5) 30 seconds.

32. The residual for the female whose hand-immersion time when repeating a non-swear word was 172 seconds and 250 seconds when repeating a swear word is approximately:

(1) 229.2 seconds.

(2) −20.8 seconds.

(3) 78.0 seconds.

(4) −78.0 seconds.

(5) 20.8 seconds.

33. For two females like those in the study whose hand-immersion times when repeating a non-swear word differ by 30 seconds, this regression analysis predicts the difference between their mean hand-immersion times when repeating a swear word to be approximately:

(1) 36.8 seconds.

(2) 32.5 seconds.

(3) 54.9 seconds.

(4) 11.4 seconds.

(5) 2.6 seconds.

34. Table 2 gives t-multipliers for constructing 95% confidence intervals for three different values of degrees of freedom.

df	t-multiplier
27	2.0518
28	2.0484
29	2.0452

Table 2: t-multipliers for 95% confidence intervals

Based on this regression analysis, a 95% confidence interval for the slope of the true regression line, β_1, is calculated by:

(1) $18.033 \pm 2.0518 \times 8.828$

(2) $1.228 \pm 2.0452 \times 0.086$

(3) $1.228 \pm 2.0518 \times 0.086$

(4) $1.228 \pm 2.0484 \times 0.086$

(5) $18.033 \pm 2.0484 \times 8.828$

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

With 95% confidence, it is estimated that for:

(1) females like those in the study whose hand-immersion time when repeating a swear word is 192 seconds, their mean hand-immersion time when repeating a swear word is somewhere between 192 and 225 seconds.

(2) a female like those in the study whose hand-immersion time when repeating a swear word is 131 seconds, her hand-immersion time when repeating a non-swear word is somewhere between 77 and 194 seconds.

(3) a female like those in the study whose hand-immersion time when repeating a non-swear word is 96 seconds, her hand-immersion time when repeating a swear word is somewhere between 125 and 147 seconds.

(4) females like those in the study whose hand-immersion time when repeating a non-swear word is 115 seconds, their mean hand-immersion time when repeating a swear word is somewhere between 147 and 171 seconds.

(5) females like those in the study whose hand-immersion time when repeating a non-swear word is 155 seconds, their mean hand-immersion time when repeating a swear word is somewhere between 149 and 268 seconds.

Questions 36 to **38** refer to the information on page 34.

36. In a two-tailed t-test for no linear relationship between **FearScore** and **NonSwearT** the correct hypotheses are:

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

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

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

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

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

37. Assume that the linear regression analysis described on page 34 is valid. (Note: This may not be true.)

Based on Table 11, page 34, which **one** of the following statements is **true**?

(1) There is strong evidence of a non-linear relationship between **FearScore** and **NonSwearT**.

(2) There is strong evidence of no linear relationship between **FearScore** and **NonSwearT**.

(3) There is no evidence against a linear relationship between **FearScore** and **NonSwearT**.

(4) There is strong evidence against a non-linear relationship between **FearScore** and **NonSwearT**.

(5) There is no evidence of a linear relationship between **FearScore** and **NonSwearT**.

38. For the data in Figure 7, page 34, which **one** of the following values would be the **closest** to the sample correlation coefficient, r?

(1) -0.8

(2) 0.9

(3) 0.7

(4) 0.1

(5) -0.4

Questions 39 to **44** refer to the information on page 35.

39. Based on the dot plots in Figure 8, page 35, which **one** of the following statements is **false**?

(1) For both plots, there are no outliers.

(2) The fear of pain scores for the females are more variable than those for the males.

(3) Both plots are multi-modal in nature.

(4) For both plots, the amount of skewness is, at most, slight.

(5) The fear of pain scores for the females are centred higher than those for the males.

40. In considering the appropriateness of using a two independent sample t-test, which **one** of the following statements is **true**?

(1) There are concerns because the spreads of the two samples are quite different.

(2) There are no concerns because, when considering the sample sizes, the t-test is sufficiently robust to withstand any non-Normal features evident in the dot plots.

(3) There are concerns because, when considering the sample sizes, the t-test is not sufficiently robust to withstand the amount of skewness evident in the dot plots.

(4) There are concerns because the male fear of pain scores are not unimodal.

(5) There are no concerns because the sample sizes are similar.

41. Which **one** of the following numbers from Table 12, page 35, is the **best** estimate of the average distance of the male fear of pain scores (in Figure 8, page 35) from the sample mean of these fear of pain scores?

(1) 2.3149

(2) 14.2700

(3) 4.1242

(4) 3.9870

(5) 15.3344

Questions 42 to **44** assume that the two independent sample t-test is valid.
(Note: This may not be true.)

42. In this two independent sample t-test, approximately how far is the estimated difference above the hypothesised value?

(1) 3.72 standard errors

(2) 1.64 standard errors

(3) 4.12 standard deviations

(4) 4.12 standard errors

(5) 3.72 standard deviations

43. Based on the *P-value* of 0.000 for this t-test, which **one** of the following statements is **false**?

(1) There is very strong evidence that the underlying mean fear of pain score for males is different from that for females.

(2) At the 5% level of significance, the observed difference between the mean fear of pain score for the females and that for the males is sufficiently large to be statistically significant.

(3) At the 5% level of significance, there is enough evidence to reject the hypothesis that the underlying mean fear of pain scores are the same for males and females.

(4) There is very strong evidence that the underlying mean fear of pain score for females is greater than that for males.

(5) At the 5% level of significance, it may be claimed that there is a very large gender effect on fear of pain scores.

44. The 95% confidence interval for $\mu_F - \mu_M$ given in the output in Table 12, page 35, is $(7.0564, 23.6124)$. Suppose we had to calculate this 95% confidence interval by hand rather than use the statistical software that produced Table 12.

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

The confidence interval calculated by hand would use:

(1) $df = 28$ and would be slightly wider than the one in Table 12.

(2) $df = 28$ and would be slightly narrower than the one in Table 12.

(3) $df = 66$ and would be slightly narrower than the one in Table 12.

(4) $df = 66$ and would be slightly wider than the one in Table 12.

(5) $df = 37$ and would be slightly wider than the one in Table 12.

Questions 45 to **50** refer to the information in **Appendix D**, pages 36 and 37.

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

(1) The underlying distribution of testosterone level for each group has a Normal distribution.

(2) The testosterone level of a female in one group is independent of the testosterone level of any female in one of the other two groups.

(3) The underlying mean testosterone level is the same for all three groups.

(4) The testosterone levels within each group are independent of each other.

(5) The standard deviations of the underlying distributions of testosterone level for each group are all equal.

Questions 46 to **50** assume that the F-test is valid.
(Note: This may not be true.)

46. The test statistic, f_0, for this F-test approximately:

(1) 2.56

(2) 6.57

(3) 43.14

(4) 0.15

(5) 0.31

47. The *P-value* for this F-test is calculated by:

(1) $2 \times \text{pr}(F \geq f_0)$ where $F \sim \text{F}(df_1 = 3, df_2 = 43)$

(2) $2 \times \text{pr}(F \geq f_0)$ where $F \sim \text{F}(df_1 = 2, df_2 = 43)$

(3) $\text{pr}(F \geq f_0)$ where $F \sim \text{F}(df_1 = 2, df_2 = 43)$

(4) $\text{pr}(F \geq f_0)$ where $F \sim \text{F}(df_1 = 2, df_2 = 45)$

(5) $\text{pr}(F \geq f_0)$ where $F \sim \text{F}(df_1 = 3, df_2 = 43)$

48. Which **one** of the following statements gives the **best** reason for the small *P-value* of 0.003 in the ANOVA output in Table 13, page 37?

(1) The variability between the three sample means is small and so is the variability within the three samples.

(2) The variability between the three sample means is large relative to the variability within the three samples.

(3) The variability between the three sample means is large.

(4) The variability between the three sample means is small.

(5) The variability between the three sample means is small relative to the variability within the three samples.

49. Based on the *P-value* of 0.003 for this F-test, which **one** of the following statements is **false**?

(1) There is strong evidence that the underlying mean testosterone levels of the three groups are not all the same.

(2) At the 1% level of significance, it may be claimed that the difference between at least one pair of observed mean testosterone levels is large enough to be statistically significant.

(3) There is strong evidence that the underlying mean testosterone level of at least one group is different from at least one of the other two underlying means.

(4) At the 5% level of significance, it may be claimed that the underlying mean testosterone levels of the three groups are all different.

(5) At the 5% level of significance, it may be claimed that the underlying mean testosterone levels of the three groups are not all the same.

50. At the 5% level of significance, which **one** of the following statements is **true**?

(1) It may be claimed that the underlying mean testosterone level for the professional group is the highest and that the underlying mean testosterone level for the housewife group is the lowest.

(2) It may be claimed that the underlying mean testosterone level for the housewife group is the lowest but the group with the highest underlying mean testosterone level cannot be determined.

(3) It may be claimed that the underlying mean testosterone level for the professional group is the highest but the group with the lowest underlying mean testosterone level cannot be determined.

(4) The group with the highest underlying mean testosterone level cannot be determined and the group with the lowest underlying mean testosterone level cannot be determined

(5) It may be claimed that the underlying mean testosterone level for the professional group is the highest and that the underlying mean testosterone level for the office worker group is the lowest.

INCLUSIONS:

* **Appendix A: E-EPA Data** for use in Questions **4** to **9**
* **Appendix B: Driving Behaviour Data** for use in Questions **10** to **22**
* **Appendix C: Pain Data** for use in Questions **24** to **44**
* **Appendix D: Testosterone Data** for use in Questions **45** to **50**

References

Garcia-Retamero, R. and Hoffrage, U. (2013). Visual representation of statistical information improves diagnostic inferences in doctors and their patients. *Social Science and Medicine*, **83**, 27–33.

Nemets, B., Stahl, Z., and Belmaker, R. H. (2002). Addition of omega-3 fatty acid to maintenance treatment for recurrent unipolar depressive disorder. *The American Journal of Psychiatry*, **159**(3), 477–479.

Piff, P. K., Stancato, D. M., Cote, S., Mendoza-Denton, R., and Keltner, D. (2012). Higher social class predicts increased unethical behavior. *PNAS*, **109**(11), 4086–4091.

Purifoy, F. E. and Koopmans, L. H. (1979). Androstenedione, testosterone, and free testosterone concentration in women of various occupations. *Biodemography and Social Biology*, **26**(3), 179–188.

Stephens, R., Atkins, J., and Kingston, A. (2009). Swearing as a response to pain. *NeuroReport*, **20**(12), 1056–1060.

CONTINUED

Appendix A: E-EPA Data

Questions 4 to **9** refer to the information in this appendix.

Several studies have reported that countries with high rates of fish oil consumption tend to have low rates of depression disorders. Fish oil contains omega-3 fatty acids. Nemets *et al.* (2002) investigated the effect of one omega-3 fatty acid, E-EPA, on 20 patients with recurrent unipolar depressive disorder all of whom were receiving maintenance antidepressant therapy.

The patients were randomly assigned to two groups. One group of 10 patients received E-EPA in 1 g doses, twice a day, over the four-week study period. The other group received a matching placebo, twice a day, over the same four-week period. All patients continued their current antidepressant treatment. Patients did not know the group to which they were assigned.

The 24-item Hamilton Depression Rating Scale was used to measure severity of depression, with higher ratings indicating greater levels of depression. A Hamilton depression rating was conducted by an experienced psychiatrist just before the first treatment (baseline rating) and at weekly intervals for four weeks. The psychiatrist did not know which group a patient was in.

CONTINUED

Questions 5 and **6** refer to the information on this page.

A randomisation test was conducted on the decrease in Hamilton depression ratings from baseline to Week 2. A plot of the data showing the difference between the two group means ($7.4 - 1.5 = 5.9$) is shown in Figure 1 and the resulting randomisation test output is shown in Figure 2.

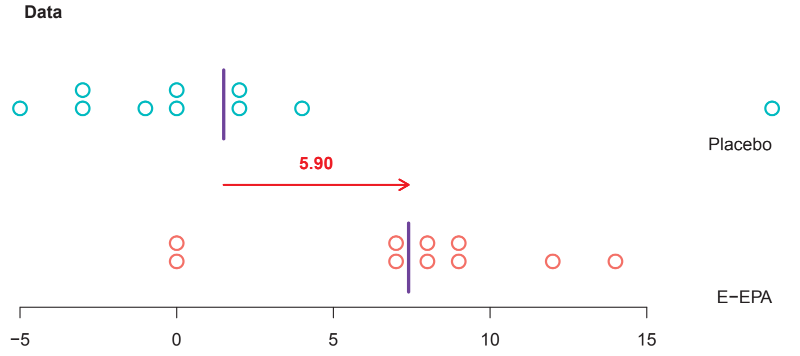

Figure 1: Decrease in Hamilton depression ratings

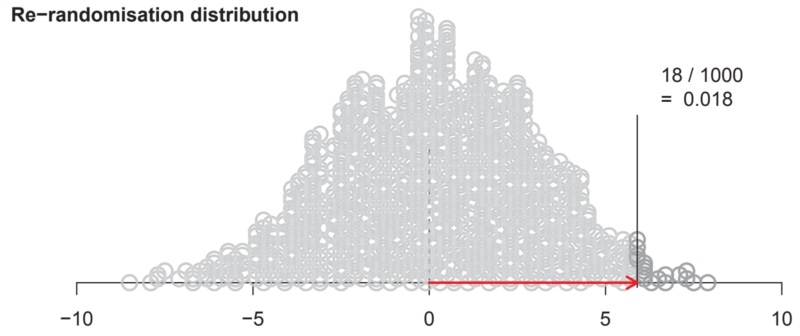

Figure 2: Randomisation test output

Questions 7 to **9** refer to the information on this page.

No longer examined

Appendix B: Driving Behaviour Data

Questions 10 to **22** refer to the information in this appendix.

Piff *et al.* (2012) conducted two studies in San Francisco which investigated possible relationships between driving behaviour and the status of the vehicle. The first study was carried out at a pedestrian crossing, not controlled by traffic lights, on a busy suburban road and the second study was carried out at a busy four-way intersection with compulsory stop signs on each entry to the intersection.

In both studies the data collection team selected an approaching vehicle in a random manner and rated its status (low, medium or high) based on the vehicle's make, age and physical appearance. Team members kept out of sight of the drivers and were not informed of the purpose of the study.

In the first study the team observed whether or not the driver gave way to a pedestrian intending to use the crossing. The area around the crossing was controlled so that the pedestrian was the only pedestrian visible to the driver, the pedestrian approached the crossing when the vehicle was 15 metres from the crossing and looked towards the vehicle to indicate to the driver their intention to cross.

Variables are defined as:

Gave way Whether or not the driver gave way to the pedestrian
- Yes
- No

Status The rated status of the vehicle
- Low
- Medium
- High

The results are shown in Table 4.

	Status			
Gave way	Low	Medium	High	**Total**
Yes	25	42	32	99
No	8	19	26	53
Total	33	61	58	152

Table 4: Driving behaviour at pedestrian crossing

Questions 14 and **15** refer to the information on this page.

No longer examined

Questions 16 to **22** refer to the information on this page.

In the second study, once a vehicle came to a stop the team observed whether or not the driver cut in front of another vehicle at the intersection. The intersection was busy enough that there were always other vehicles waiting to cross the intersection.

A further variable is defined as:

Cut in front Whether or not the driver cut in front of another vehicle
– Yes
– No

The results are shown in Table 6.

	Status			
Cut in front	Low	Medium	High	**Total**
Yes	4	15	15	34
No	55	99	86	240
Total	59	114	101	274

Table 6: Driving behaviour at busy intersection

Assume that these 274 drivers are a random sample of drivers and vehicles in the San Francisco area.

A Chi-square test was conducted to see whether this type of behaviour by drivers in the San Francisco area depends on the status of the vehicle. The results of this Chi-square test are shown in Table 7.

			Status			
			Low	Medium	High	Total
Cut in front	Yes	Count	4	15	15	34
		Expected count	7.321	++	++	34.0
		Cell contribution	++	++	0.486	
	No	Count	55	99	86	240
		Expected count	51.679	++	88.467	240.0
		Cell contribution	0.213	0.007	0.069	
Total		Count	59	114	101	274
		Expected count	59.0	114.0	101.0	274.0

Chi-Square Tests

	Value	df	Sig.
Pearson Chi-Square	2.333	2	.311
Likelihood Ratio	2.599	2	.273
Linear-by-Linear Association	2.000	1	.157
N of Valid Cases	274		

Note: Some values have been replaced by ++.

Table 7: Chi-square test output

Appendix C: Pain Data

Questions 24 to **44** refer to the information in this appendix.

Stephens *et al.* (2009) investigated swearing as a response to tolerating pain. Initially the participants completed a Fear of Pain Questionnaire, from which a fear of pain score was calculated. Higher scores indicate greater fear of pain.

In the study participants immersed their unclenched non-dominant hand in cold water (5 °C) for as long as they could bear the cold temperature. The time participants kept their hand immersed gives a measure of the participant's tolerance of pain. Each participant did this twice; once while repeatedly saying a swear word and once while repeatedly saying a non-swear word. The time between the two trials was long enough so that the first trial did not interfere with the second. The order of the type of word used (swear or non-swear) was randomised across the participants. For each participant, the time (in seconds) the hand was immersed was recorded.

Some of the variables recorded are defined as:

Gender The gender of the participant
– Female
– Male

FearScore A measure of the participant's fear of pain (from 30 to 150)

WordType The type of word repeated
– Swear
– Non-swear

SwearT The time the participant's hand was immersed when repeating a swear word (in seconds)

NonSwearT The time the participant's hand was immersed when repeating a non-swear word (in seconds)

Questions 25 to **44** assume that the 29 female participants are a random sample from some underlying population of females and **Questions 39** to **44** assume that the 38 male participants are a random sample from some underlying population of males.

Questions 25 to **27** refer to the information on this page.

A paired-data t-procedure was conducted to estimate the average difference in hand-immersion times when a female repeats a swear word or a non-swear word. A dot plot is shown in Figure 4 and computer output for the t-procedure is shown in Table 8.

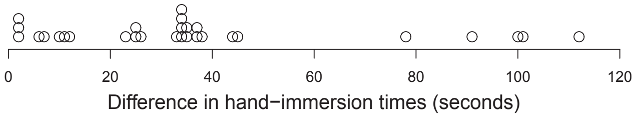

Figure 4: Differences in hand-immersion times, SwearT – NonSwearT

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SwearT	120.28	29	80.509	14.950
	NonSwearT	83.28	29	61.641	11.446

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 SwearT - NonSwearT	37.000	30.841	++	25.269	48.731	6.461	28	.000

Note: One value has been replaced by ++.

Table 8: Paired-data t-procedure output

CONTINUED

Questions 28 and **29** refer to the information on this page.

The differences in hand-immersion times shown in Figure 4, page 31, were used to form a bootstrap confidence interval for a median. Plots for the iNZightVIT bootstrap confidence interval output are shown in Figure 5.

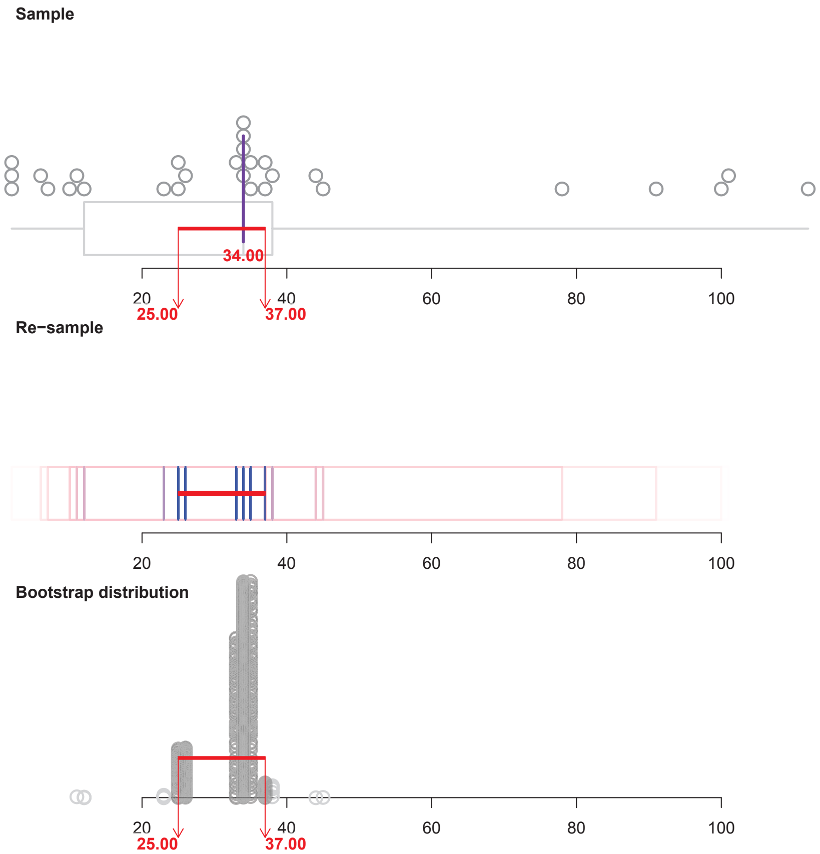

Figure 5: Bootstrap confidence interval construction output

CONTINUED

Questions 30 to **35** refer to the information on this page.

A simple linear regression analysis was carried out to investigate whether the time a female's hand is immersed in cold water when repeating a non-swear word could be used to predict the time her hand is immersed in cold water when repeating a swear word. A scatter plot of the data from the 29 females is shown in Figure 6 and output from the simple linear regression analysis is shown in Tables 9 and 10.

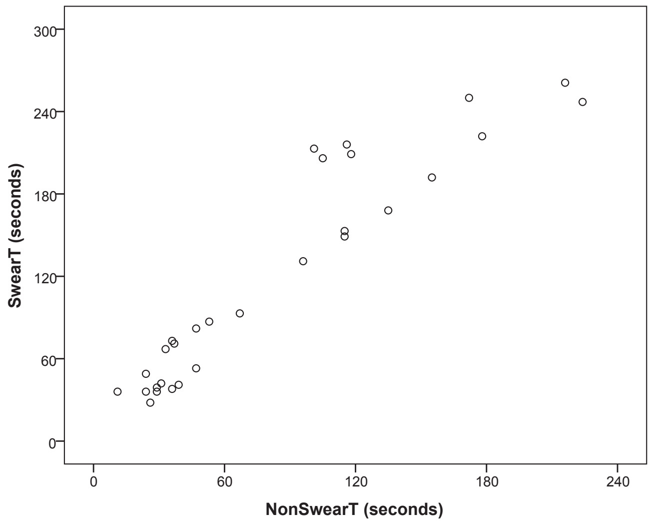

Figure 6: Hand-immersion times, swear word against non-swear word

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	18.033	8.828		2.043	.051	-.081	36.146
	NonSwearT	1.228	.086	.940	14.321	.000	1.052	1.404

a. Dependent Variable: SwearT

Table 9: Simple linear regression output

NonSwearT	SwearT	LMCI_1	UMCI_1	LICI_1	UICI_1
96	131	125.01106	146.78520	77.49810	194.29815
115	149	147.19821	171.25317	100.60236	217.84902
155	192	191.82268	224.84999	148.63093	268.04174

Table 10: Simple linear regression prediction output

CONTINUED

Questions 36 to **38** refer to the information on this page.

A simple linear regression analysis was also carried out to investigate the nature of the relationship between the fear of pain score for a female and the time her hand is immersed in cold water when repeating a non-swear word. A scatter plot is shown in Figure 7 and output from the simple linear regression analysis is shown in Table 11.

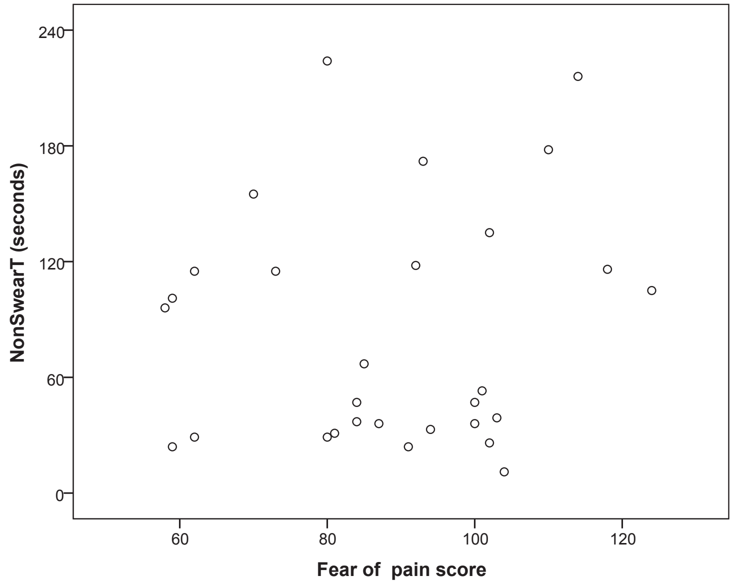

Figure 7: Hand-immersion times (non-swear word) against fear of pain scores

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	47.177	57.982		.814	.423	-71.791	166.146
	Fear of pain score	.407	.641	++	.635	.531	-.907	1.721

Note: One value has been replaced by ++.

a. Dependent Variable: NonSwearT (seconds)

Table 11: Simple linear regression output

CONTINUED

Questions 39 to **44** refer to the information on this page.

Let:

μ_F be the underlying mean fear of pain score for females

and

μ_M be the underlying mean fear of pain score for males.

A two-tailed t-test for no difference between the mean fear of pain score for males and females was conducted. Dot plots are shown in Figure 8 and computer output is shown in Table 12.

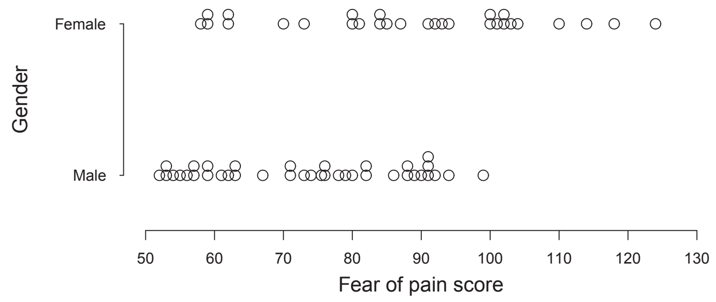

Figure 8: Fear of pain scores

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Fear of pain score	Female	29	88.690	18.3811	3.4133
	Male	38	73.355	14.2700	2.3149

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
Fear of pain score	Equal variances assumed	1.635	.206	3.846	65	.000	15.3344	3.9870	7.3718	23.2970
	Equal variances not assumed			3.718	51.445	.000	15.3344	4.1242	7.0564	23.6124

Table 12: t-test output

CONTINUED

Appendix D: Testosterone Data

Questions 45 to **50** refer to the information in this appendix.

Purifoy and Koopmans (1979) collected data on the testosterone levels (in milligrams per 100 millilitres of blood) of 46 healthy females from New Mexico. The women were classified into three groups: housewife, office worker and professional.

An F-test for one-way analysis of variance was carried out to investigate any differences in testosterone levels among the three groups. Dot plots are shown in Figure 9 and test output is shown in Table 13 on the next page.

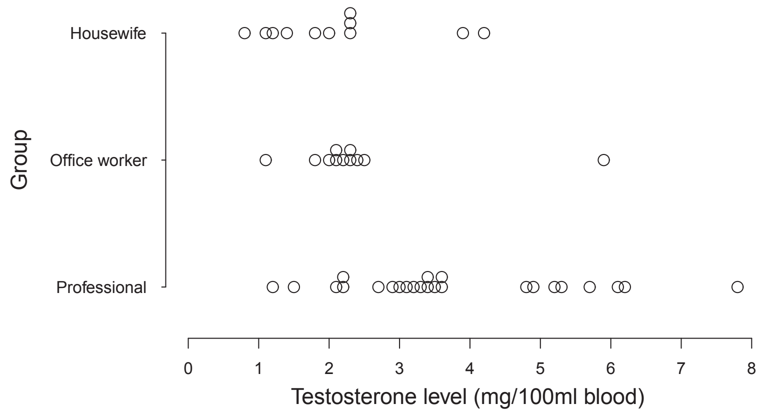

Figure 9: Testosterone levels

CONTINUED

Descriptives

Testosterone level

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Housewife	11	2.118	1.0889	.3283	1.387	2.850	.8	4.2
Office worker	11	2.427	1.2125	.3656	1.613	3.242	1.1	5.9
Professional	24	3.788	1.6337	.3335	3.098	4.477	1.2	7.8
Total	46	3.063	1.5973	.2355	2.589	3.537	.8	7.8

ANOVA

Testosterone level

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	26.863	++	13.431	++	.003
Within Groups	87.944	++	2.045		
Total	114.807	++			

Multiple Comparisons

Dependent Variable: Testosterone level

Tukey HSD

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Housewife	Office worker	-.3091	.6098	.868	-1.789	1.171
	Professional	-1.6693	.5207	.007	-2.933	-.405
Office worker	Housewife	.3091	.6098	.868	-1.171	1.789
	Professional	-1.3602	.5207	.032	-2.624	-.096
Professional	Housewife	1.6693	.5207	.007	.405	2.933
	Office worker	1.3602	.5207	.032	.096	2.624

Note: Some values have been replaced by ++.

Table 13: F-test for one-way analysis of variance output

CONTINUED

ANSWERS:

1. (3)	2. (3)	3. (4)	4. (5)	5. (3)
6. (5)	7. (2)	8. (2)	9. (4)	10. (5)
11. (1)	12. (2)	13. (1)	14. (5)	15. (4)
16. (2)	17. (1)	18. (3)	19. (1)	20. (5)
21. (3)	22. (2)	23. (1)	24. (4)	25. (2)
26. (4)	27. (1)	28. (5)	29. (1)	30. (2)
31. (5)	32. (5)	33. (1)	34. (3)	35. (4)
36. (4)	37. (5)	38. (4)	39. (3)	40. (2)
41. (2)	42. (1)	43. (5)	44. (1)	45. (3)
46. (2)	47. (3)	48. (2)	49. (4)	50. (3)