# The University of Sydney
School of Mathematics and Statistics

## STAT5002

This is a sample exam paper used for showing the format of the exam. The actual exam may cover different topics and have different number of questions in both sections.

Time Allowed: **Reading time — 10 minutes; Writing time — 2 hours**

Exam Conditions: This is a closed-book examination — no material permitted. Writing is not permitted at all during reading time.

Family Name: .............................................. SID: ...........................

Other Names: .............................................. Seat Number: .................

Please check that your examination paper is complete (27 pages) and indicate by signing below.
I have checked the examination paper and affirm it is complete.

Signature: .............................................. Date: .........................



**This examination has two sections: Multiple Choice and Extended Answer.**

The Multiple Choice Section is worth 40% of the total examination.
There are 16 questions. The questions are of equal value.
All questions may be attempted.

Answers to the Multiple Choice questions must be entered on the Multiple Choice Answer Sheet before the end of the examination.

The Extended Answer Section is worth 60% of the total examination. All questions may be attempted. Working must be shown.

**Non-programmable calculators may be used, as long as they have a University of Sydney approval sticker on them.**

**THE QUESTION PAPER MUST NOT BE REMOVED FROM THE EXAMINATION ROOM.**

| Marker's use only | |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

## Multiple Choice Section

*In each question, choose at most one option.*

**Your answers must be entered on the Multiple Choice Answer Sheet.**

**1.** In a dataset of size 8, the mean is 7 and standard deviation is 4. We add 4 to each observation in the dataset. The new mean and SD are respectively

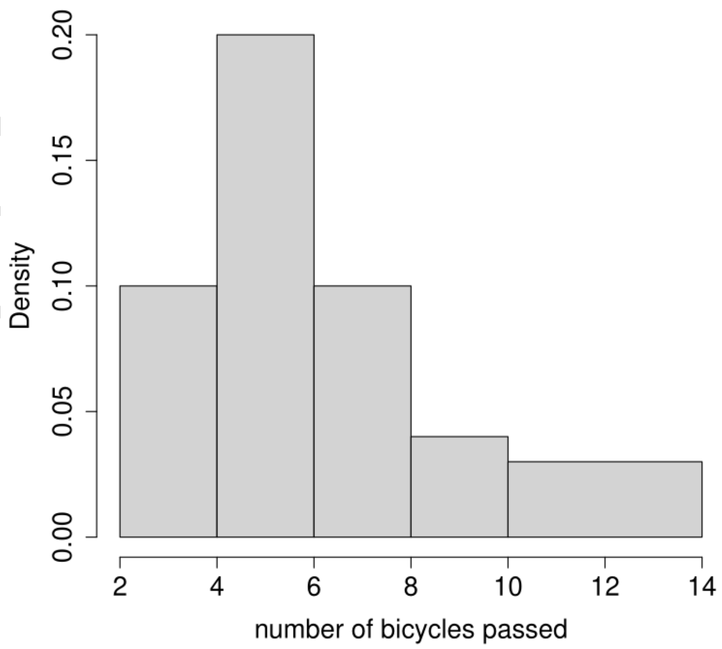
(a) 7 and 4

(b) 11 and 8

(c) 7 and 8

(d) 11 and 4

**2.** Each day for a 25-day period, Jimmy counted the number of bicycles he passed on his drive to work. A histogram of his counts is plotted below. How many days did he pass less than 6 bicycles? (Assume each bin is of the left-closed and right-open form $[a, b)$.)




(a) 8

(b) 10

(c) 12

(d) 15

This page is left blank for your working.
Working for the Multiple Choice section will **not** be marked.

**3.** A study conducted at the University of Sydney shows that the average height of the female staff members is 165.52cm with a SD of 5.59cm. Consider this the full population of the female staff members of the University of Sydney and that the height can be described by a normal model. Which one is the correct R code for calculating the percentage of female staff members with height between 152.40cm and 167.64cm?

(a) qnorm$(167.64, 165.52, 5.59)$-qnorm$(152.4, 165.52, 5.59)$

(b) pnorm$(167.64, 165.52, 5.59^2)$-pnorm$(152.4, 165.52, 5.59^2)$

(c) qnorm$(167.64, 165.52, 5.59^2)$-qnorm$(152.4, 165.52, 5.59^2)$

(d) pnorm$(167.64, 165.52, 5.59)$-pnorm$(152.4, 165.52, 5.59)$

**4.** The weight of the box is normally distributed with a mean of 5 kilograms and a SD of 2 kilograms. The box is going to be shipped to a country where weights are commonly reported in pounds. To convert from kilograms to pounds, one must use the formula lb $= 2.205 \times$ kg, where kg is the weight in kilograms and lb is the weight in pounds. What is the probability that the weight of the box in pounds is greater than 8 pounds, which is the weight limit for packages being shipped to that country? (Round to 3 d.pl.) You may find the following R output useful:

```
> pnorm(0.343)
[1] 0.6342008
> pnorm(0.566)
[1] 0.7143031
> pnorm(0.686)
[1] 0.7536434
```

(a) 0.754

(b) 0.714

(c) 0.634

(d) 0.246

**5.** Consider the following table with pairs of values for $x$ and $y$,

| $x$ | $y$ |
|---|---|
| 2 | 2 |
| 2 | 1 |
| 1 | 1 |

Now we add a 4th pair of values to make the correlation coefficient 0. By inspection, the added values for $x$ and $y$ are

(a) 0 and 0

(b) 2 and 1

(c) 1 and 2

(d) There is not enough information to tell.

This page is left blank for your working.
Working for the Multiple Choice section will **not** be marked.

**6.** Select the correct statement from below.

(a) If $A$ and $B$ are mutually exclusive events, then $A$ and $B$ are independent.

(b) If we take a sample of points of size 10 from a scatter plot with correlation coefficient 1, then the sampled points lie on a straight line.

(c) The slope of a regression line relating height and age among boys is 1.3 times higher than the slope of a regression line relating the height and age among girls. Therefore, the correlation coefficient between height and age among boys is 1.3 times higher than correlation coefficient between height and age among girls.

(d) All of the other listed options are incorrect.

**7.** Select the correct statement from below.

(a) If the correlation coefficient between two variables $X$ and $Y$ is $r = 0.95$, it means 95% of the points in the scatterplot lie on a straight line.

(b) If $r = 0$ for two variables, it means there is no relationship between them.

(c) In a University class, the correlation between the quantitative two variables, students' age and students' birth year, is -1.

(d) All of the other listed options are incorrect.

**8.** A random sample of size 400 is taken from a large population with mean $\mu = 12$ and standard deviation $\sigma = 2.5$. The probability that the sample average is between 11.875 and 12.25 is closest to

(a) 81%

(b) 67%

(c) 95%

(d) 47.5%

This page is left blank for your working.
Working for the Multiple Choice section will **not** be marked.

**9.** In general, which of the following descriptive summary measures *cannot* be easily approximated from a boxplot?

(a) The variance.

(b) The 25% percentile.

(c) The interquartile range.

(d) The median.

**10.** A hypothesis test was performed, and the calculated P-value was 0.02. Given a significance level $\alpha = 0.05$, which of the following statements correctly interprets the P-value?

(a) Since the P-value is very small, we should accept the null hypothesis.

(b) There is a 2% chance that the alternative hypothesis is true.

(c) There is a 2% chance that the null hypothesis is true.

(d) The data provides evidence against the null hypothesis.

**11.** In a "taste test" challenge, 100 randomly sampled students try two different drinks, Brand A and Brand B (in a double-blind arrangement) to see which drink they prefer. If $p$ denotes the proportion of the student population who prefer Brand A, which alternative hypothesis should be used to examine the question: "Is one of the brands more popular than the other?"

(a) $p \neq 0.5$

(b) $p = 0.5$

(c) $p > 0.5$

(d) $p < 0.5$

**12.** A six-sided die, with faces numbered 1, 2, 3, 4, 5, 6, is rolled 30 times. The numbers rolled each time are stored in the R vector `rolls`. Using the R output below, determine the value of the chi-squared test statistic for testing that each face is equally likely.

```
> rolls
[1] 6 4 1 2 1 3 2 4 2 5 1 3 4 4 5 6 1 5 4 3 5 4 1 2 6 5 6 6 4 5
> table(rolls)
rolls
1 2 3 4 5 6
5 4 3 7 6 5
> sum((table(rolls)-5)^2)
[1] 10
```

(a) 2

(b) 10

(c) 2.4

(d) 12

This page is left blank for your working.
Working for the Multiple Choice section will **not** be marked.

**13.** A random sample of size 100 is taken from a large population. The sample mean is 26.4 while the sample standard deviation is 7.0. The value of a t-statistic for testing the null hypothesis that the population mean equals 25 is given by

(a) 3.77

(b) 2.8

(c) 5.28

(d) 2

**14.** What is the purpose of the adjusted R-squared in a multiple linear regression model?

(a) To measure the correlation between two specific independent variables.

(b) To account for the number of independent variables and provide a more accurate model fit.

(c) To represent the slope of the regression line.

(d) To show the variability of the independent variables.

**15.** In the following list, select the **wrong** statement.

(a) In multiple linear regression, we need to assume that the dependent variable and residuals follow some normally shaped box models.

(b) In a logistic regression model, the range of predicted probability should be between 0 and 1.

(c) In multiple linear regression, we must assume that the independent variables are perfectly correlated with each other.

(d) In backward selection, we need to begin with a model with all available independent variables.

This page is left blank for your working.
Working for the Multiple Choice section will **not** be marked.

**16.** A random sample of size $n = 20$ is taken from a normally shaped box model with mean $\mu$ and SD $\sigma$ both unknown. Consider the following R output, where the sample values are stored in $x$:

```
> mean(x)
[1] 14.83756
> sd(x)
[1] 2.975263
> qt(c(0.90, 0.95, 0.975, 0.98, 0.99, 0.995), df = 18)
[1] 1.330391 1.734064 2.100922 2.213703 2.552380 2.878440
> qt(c(0.90, 0.95, 0.975, 0.98, 0.99, 0.995), df = 19)
[1] 1.327728 1.729133 2.093024 2.204701 2.539483 2.860935
> qt(c(0.90, 0.95, 0.975, 0.98, 0.99, 0.995), df = 20)
[1] 1.325341 1.724718 2.085963 2.196658 2.527977 2.845340
> qt(c(0.90, 0.95, 0.975, 0.98, 0.99, 0.995), df = 21)
[1] 1.323188 1.720743 2.079614 2.189427 2.517648 2.831360
> qt(c(0.90, 0.95, 0.975, 0.98, 0.99, 0.995), df = 22)
[1] 1.321237 1.717144 2.073873 2.182893 2.508325 2.818756
```

A 95% confidence interval for $\mu$ is given by $14.8 \pm c \times \text{SE}$ where $(c, \text{SE})$ are given by

(a) (2.09, 0.67)

(b) (1.73, 0.67)

(c) (2.09, 2.98)

(d) (1.33, 0.67)

This page is left blank for your working.
Working for the Multiple Choice section will **not** be marked.



**End of Multiple Choice Section**
**Make sure that your answers are entered on the Multiple Choice Answer Sheet**

The Extended Answer Section begins on the next page

## Extended Answer Section

*There are* **four** *questions in this section, each may has a number of parts. Write your answers in the space provided below each part.*

**1.** A drug company claims that a new sleeping pill increases sleeping time by an average of 28 minutes. Assume the increase in sleeping time due to this medicine is normally distributed, with an unknown standard deviation. Researchers suspect the actual increase is less than the company's claim. A sample of 16 data points (see the R output below) was used to test this suspicion.

Carry out the following steps of a hypothesis test. You can use the following R outputs.

```
> sleep = c(22.2, ..., 18.4)
> length(sleep)
[1] 16
> round(mean(sleep),2)
[1] 23.73
> round(sd(sleep),2)
[1] 5.16
```

(a) State the null and alternative hypotheses.



(b) Which testing procedure should be used here? Explain your answer.

(c) Compute the observed test statistic. Show your derivation and round your answer to two decimal places.

(d) Is this a one-sided or two-sided test? What values of test statistics will argue against the null hypotheses? Explain your answer.

(e) At a significance level $\alpha = 0.05$, do we observe sufficient evidence to reject the null hypothesis? Explain your answer. Hint: you can use the following R outputs.

```
> round(qt(c(0.95, 0.975), 14), 3)
[1] 1.761 2.145
> round(qt(c(0.95, 0.975), 15), 3)
[1] 1.753 2.131
> round(qt(c(0.95, 0.975), 16), 3)
[1] 1.746 2.120
```



***Solution:***

(a) The null hypothesis is $H_0 : \mu = 28$, and the alternative hypothesis is $H_1 : \mu < 28$.

(b) The Student T distribution is appropriate here (with 15 degrees of freedom) because the population standard deviation is unknown and the population is normal.

(c) We will use the T-test statistic formula:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{-4.27}{5.16/\sqrt{16}} = -3.310$$

(d) This is a one-sided test because we are interested in whether the population mean is less than 28. Small values of test statistics argue against the null hypotheses.

(e) From the R output, the critical value for T-test with $\alpha = 0.05$ is approximately -1.753 (using symmetry of Student's $t$-distribution), since it's a one-sided test.

Since the observed test statistic is less than the critical value of -1.753, it falls in the rejection region.

Therefore, we reject the null hypothesis and conclude that there is sufficient evidence at the 0.05 significance level to suggest that the mean increase is less than 28 minutes.

**2.** The number of speeding tickets issued by a speed camera near USyd (yes, there's one on Parramatta Road!) over the last five weeks is as follows:

| Week 1 | Week 2 | Week 3 | Week 4 | Week 5 |
|---|---|---|---|---|
| 63 | 60 | 54 | 58 | 65 |

Perform a chi-squared goodness-of-fit test to determine if there is any difference in the number of tickets across these weeks.

(a) State the null and alternative hypotheses.

(b) Set up the table of expected probabilities and the table of expected counts.

(c) Compute the observed test statistic.

(d) What is the degrees of freedom of the chi-squared distribution for test statistic? Explain your answer.

(e) What values of test statistics will argue against the null hypotheses? Explain your answer.

(f) At a significance level $\alpha = 0.05$, do we observe sufficient evidence to reject the null hypothesis? Explain your answer.

```
> qchisq(0.95, 5)
[1] 11.0705
> qchisq(0.95, 4)
[1] 9.487729
> qchisq(0.95, 3)
[1] 7.814728
> qchisq(0.95, 2)
[1] 5.991465
> qchisq(0.95, 1)
[1] 3.841459
```



***Solution:***

(a) $H_0$: The number of tickets is consistent across the five weeks. That is, the number of tickets is equally distributed with probability 1/5.

$H_1$ : The number of tickets is not consistent across the five weeks.

(b) Expected Probability is 0.2 for each week, the table of expected count is

| Week 1 | Week 2 | Week 3 | Week 4 | Week 5 |
|---|---|---|---|---|
| 60 | 60 | 60 | 60 | 60 |

(c)

$$t = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} \approx 1.23$$

(d) The degrees of freedom (df) for a chi-squared goodness-of-fit test is given by: df = Number of categories - 1

In this case, there are 5 weeks (categories), so: df $= 5 - 1 = 4$

(e) For a chi-squared goodness-of-fit test, large values of the test statistic argue against the null hypothesis.

(f) For $\alpha = 0.05$ and 4 degrees of freedom, the critical value from the R output is approximately 9.49. Since the observed test statistic $1.23 < 9.49$, the observed test statistic does not exceed the critical value. Therefore, we fail to reject the null hypothesis. The accident rate appears to be consistent for these months.

**3.** Two samples have been independently drawn from two independent normal populations with unknown variances. The first sample has a size of $n_x = 21$, and the second has a size of $n_y = 17$. From these samples, we obtained sample means of $\bar{x} = 5.2$ and $\bar{y} = 6.1$, with sample standard deviations $s_x = 5.2$ and $s_y = 5.0$, respectively. Using this data, we wish to test whether the difference between the two population means is zero.

(a) State the null and alternative hypotheses.

(b) Which test is appropriate here. Justify your choice.

(c) Suppose we want to carry the classical two-sample T-test, calculate the pooled standard deviation and the estimated standard error of the sample mean difference. Round your answer to two decimal places.

(d) Using a 5% significance level, complete the remaining steps of the classical two-sample tt-test to evaluate the hypotheses stated in part (a). The following R outputs may be useful.

```
> round(qt(c(0.95, 0.975), 36), 2)
[1] 1.69 2.03
> round(qt(c(0.95, 0.975), 37), 2)
```

```
[1] 1.69 2.03
> round(qt(c(0.95, 0.975), 38), 2)
[1] 1.69 2.02
```



***Solution:***

(a) $H_0 : \mu_X - \mu_Y = 0 \;\; vs \;\; H_1 : \mu_X - \mu_Y \neq 0$

(b) We can apply the classical two-sample T-test, as the sample SDs are close. In addition, we have the independence and normal population assumption satisfied.

Applying the Welch test is also appropriate, which does not rely on the additional equal SD assumption.

(c) We have

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{20 \times 5.2^2 + 16 \times 5^2}{21 + 17 - 2} \approx 26.13$$

Hence the pooled standard deviation is $s_p = 5.11$ and the estimated SE of the sample mean difference is

$$SE = s_p\sqrt{\frac{1}{21} + \frac{1}{17}} \approx 1.67.$$

(d) The observed test statistic is

$$t = \frac{\bar{X} - \bar{Y} - 0}{SE} \approx \frac{-0.9}{1.67} \approx -0.539$$

The Student's $t$-distribution has $21 + 17 - 2 = 36$ degrees of freedom. Since we have a two-sided alternative, the critical region of rejection is $|T| > 2.03$ (2.5% areas in both tails). The observed test statistic falls outside of the critical region, so we don't have sufficient evidence to reject $H_0$.

**4.** The relationship between social media advertising spending (`S`) in dollars, website traffic (`V`) in thousands of clicks, and subscription numbers (`I`) for a subscription-service company was analyzed using a multiple regression model. Refer to the model labeled `Model` in the provided R output. Some parts of the output are masked by `#####`. Assuming that all necessary assumptions for applying a linear regression model are satisfied, answer the questions below.

```
# Data set
> data = data.frame(
  Weeks = 1:12,
  S = ...,
  V = ...,
  I = ...
)
> Model = lm(I ~ S + V, data = data)
> summary(Model)

Call:
lm(formula = I ~ S + V, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-53.564 -17.226   6.669  24.167  38.051

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.3196    30.8759   #####   ######
S              .0561     .06706   #####   ######
V              .0367     .01143   #####   ###### *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32.84 on 9 degrees of freedom
Multiple R-squared:  0.9889,        Adjusted R-squared:  0.9865
F-statistic:   402 on 2 and 9 DF,  p-value: 1.58e-09
```

(a) Write down the fitted multiple regression model based on the R output above. Interpret how to use the fitted model to predict changes in subscription numbers in terms of changes in advertising spending and website visits.

(b) After adjusting for the effect of advertising spending, is website traffic significant in explaining subscription numbers? Your answer should include at least the following steps, and the R output provided may be useful:

(*i*) State the null and alternative hypotheses.

(*ii*) What is the observed value of the test statistic? You can round your answer to two decimal places.

(*iii*) What is your conclusion based on a 2% level of significance? Justify your answer and the choice of the R output.

```
> round(qt(c(0.98, 0.985, 0.99, 0.995),9), 3)
[1] 2.398 2.574 2.821 3.250
> round(qt(c(0.98, 0.985, 0.99, 0.995),9), 4)
[1] 2.3984 2.5738 2.8214 3.2498
> round(qt(c(0.98, 0.985, 0.99, 0.995),10), 4)
[1] 2.3593 2.5275 2.7638 3.1693
> round(qt(c(0.98, 0.985, 0.99, 0.995),11), 4)
[1] 2.3281 2.4907 2.7181 3.1058
> round(qt(c(0.98, 0.985, 0.99, 0.995),12), 4)
[1] 2.3027 2.4607 2.6810 3.0545
```

**(cont.)**

***Solution:***

(a) Based on the R output, the fitted multiple regression model is:

$$\widehat{I} = 3.3196 + 0.0561 \cdot S + 0.0367 \cdot V$$

- Holding $V$ constant, a one dollor increase in advertising spending $S$ is associated with an increase of 0.0561 subscriptions.

- Holding $S$ constant, a one thousand increase in website clicks $V$ is associated with an estimated increase of 0.0367 subscriptions.

(b) (*i*) $H_0 : b_V = 0$ Website traffic has no effect in explaining the subscription numbers after adjusting the effect of spending. $H_1 : b_V \neq 0$ Website traffic has an effect in explaining the subscription numbers after adjusting the effect of spending.

(*ii*) The observed test statistic is

$$t = \frac{\widehat{b_V} - b_V}{SE(\widehat{b_V})} = \frac{.0367 - 0}{.01143} \approx 3.21$$

by reading the estimated coefficient and its SE from the R output.

(*iii*) This is a two-sided test with 9 degrees of freedom in Student's $t$-distribution, so the critical region is $|T| > 2.821$ (99% quantile). The observed statistic falls inside of the critical region, so we reject $H_0$, indirectly suggests that website traffic has an effect in explaining the subscription numbers after adjusting the effect of spending.



**End of Extended Answer Section**

## End of Examination

Write your SID here ⟶

The University of Sydney
School of Mathematics and Statistics

**STAT5002 This is a sample exam paper used for showing the format of the exam. The actual exam may cover different topics and have different number of questions in both sections.**

**Code your SID into the columns below each digit, by filling in the appropriate oval.**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | **0** |
| **1** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | **1** |
| **2** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | **2** |
| **3** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | **3** |
| **4** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | **4** |
| **5** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | **5** |
| **6** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | **6** |
| **7** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | **7** |
| **8** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | **8** |
| **9** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | **9** |



Family Name: ..........................................................................

Other Names: ..........................................................................

Seat Number: .................

**Answers** ⟶

| | **a** | **b** | **c** | **d** |
|---|---|---|---|---|
| **Q1** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q2** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q3** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q4** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q5** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q6** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q7** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q8** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q9** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q10** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q11** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q12** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q13** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q14** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q15** | ⬭ | ⬭ | ⬭ | ⬭ |
| **Q16** | ⬭ | ⬭ | ⬭ | ⬭ |

Attempt every question. You will **not** be awarded negative marks for incorrect answers.

Fill in exactly one oval per question.

If you make a mistake, draw a cross (X) through any mistakenly filled in oval(s) and then fill in your intended oval.

An answer which contains two or more filled in (and uncrossed) ovals will be awarded no marks.

## Correct Responses to MC Component of

**STAT5002 This is a sample exam paper used for showing the format of the exam. The actual exam may cover different topics and have different number of questions in both sections.**

*Sample MainExam :*

Q1 $\longrightarrow$ d
Q2 $\longrightarrow$ d
Q3 $\longrightarrow$ d
Q4 $\longrightarrow$ a
Q5 $\longrightarrow$ c
Q6 $\longrightarrow$ b
Q7 $\longrightarrow$ c
Q8 $\longrightarrow$ a
Q9 $\longrightarrow$ a
Q10 $\longrightarrow$ d
Q11 $\longrightarrow$ a
Q12 $\longrightarrow$ a
Q13 $\longrightarrow$ d
Q14 $\longrightarrow$ b
Q15 $\longrightarrow$ c
Q16 $\longrightarrow$ a