Name:_____ **Student Number**:_____
(Last Name) (First Name)

Stats 2B03 3rd Sample Test for Sections 9.1-9.5, 10.1-10.3

(and some previous sections)

Day Class
Duration: 75 Minutes **Instructors:** All Sections
Maximum Mark: 10

This test paper consists of 10 multiple choice questions. Marks will NOT be deducted for wrong answers (i.e., there is no penalty for guessing). QUESTIONS MUST BE ANSWERED ON THE COMPUTER CARD with an HB PENCIL. Answer all questions. You are responsible for ensuring that your copy of this paper is complete. Bring any discrepancy to the attention of your invigilator. Only the McMaster standard Calculator Casio fx-991 is allowed. The formula sheet at the front of this manual will be provided with the tests and exam.

1. A researcher collects the following (x, y) pairs,

x	3	6	8	5	9
y	2	1	?	?	?

fits a regression line, and produces the following 95% prediction interval for the value of y when $x = 3$: $(0.563,\ 4.887)$. Find the first residual.
(a) $-.125$ **(b)** $-.725$ **(c)** $-.325$ **(d)** $.245$ **(e)** $.145$

2. A researcher collects the following (x, y) pairs,

x	1	4	3	6	7
y	4	3	2	5	?

and produces the following Minitab output.

```
---------------------------------------------------------------------------
Analysis of Variance

Source      DF   Adj SS  Adj MS  F-Value  P-Value
Regression   1   0.7018  0.7018     0.23    0.667
  x          1   0.7018  0.7018     0.23    0.667
Error        3   9.2982  3.0994
Total        4  10.0000

Model Summary

      S   R-sq  R-sq(adj)  R-sq(pred)
1.76052  7.02%      0.00%       0.00%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant    3.74     1.74     2.15    0.121
x         -0.175    0.369    -0.48    0.667  1.00

Regression Equation

y = 3.74 - 0.175 x
---------------------------------------------------------------------------
```

Find the missing y-value.

(a) 3 **(b)** 6 **(c)** 7 **(d)** 1 **(e)** 2

3. A researcher collects the following (x, y) pairs,

x	1	4	8	6	7
y	4	3	2	-5	-4

and produces the following Minitab output.

```
---------------------------------------------------------------------------
Regression Analysis: y versus x

Analysis of Variance

Source      DF  Adj SS  Adj MS  F-Value  P-Value
Regression   1   21.95   21.95     1.37    0.326
  x          1   21.95   21.95     1.37    0.326
Error        3   48.05   16.02
Total        4   70.00

Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
4.00216  31.35%      8.47%       0.00%

Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant       ?        ?        ?    0.368
x         -0.844    0.721    -1.17    0.326  1.00

Regression Equation

y = ? - 0.844 x
---------------------------------------------------------------------------
```

Which of the following statements is correct?

(a) There is a 32.6% chance that x and y are correlated.

(b) There is a 32.6% chance of making a Type I error.

(c) x and y are not correlated since $.326 > .05$.

(d) If there is a correlation between x and y then there is a 32.6% chance of falsely concluding that there is no correlation between x and y.

(e) If there is no correlation between x and y then there is a 32.6% chance of obtaining a sample correlation coefficient at least as big (in absolute value) as the observed one.

4. A researcher collects some (x, y) pairs and produces the following Minitab output.

```
---------------------------------------------------------------------
Regression Analysis: y versus x

Analysis of Variance

Source      DF   Adj SS  Adj MS  F-Value  P-Value
Regression   1        ?       ?        ?        ?
  x          1        ?       ?        ?        ?
Error        5    8.115   1.623
Total        6  100.857


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
1.27395  91.95%     90.35%      79.97%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant   4.799    0.922     5.21    0.003
x              ?        ?        ?        ?  1.00


Regression Equation

y = 4.799 - 0.816 x
---------------------------------------------------------------------
```

Is there a significant correlation between y and x? Use $\alpha = .05$.

(a) Yes, because 7.56 is greater than 2.571
(b) Yes, because 7.56 is greater than 2.447
(c) Yes, because 6.76 is greater than 2.571
(d) Yes, because 6.76 is greater than 2.447
(e) Yes, because 5.23 is greater than 2.015

5. A researcher collected the following data on a random sample of 5 patients with hypertension

Y = mean arterial blood pressure (mm Hg)
X_1 = duration of hypertension (years)
X_2 = basal pulse (beats/min)
X_3 = measure of stress

Patient	Y	X_1	X_2	X_3
1	105	5.1	63	33
2	115	3.8	70	14
3	116	8.2	72	10
4	117	5.8	73	99
5	112	7.0	72	95
6	121	9.3	71	10

and produced the following Minitab output:

```
---------------------------------------------------------------------------------------
Source       DF   Adj SS  Adj MS  F-Value  P-Value
Regression    ?        ?       ?        ?    0.261
Error         ?        ?       ?
Total         ?        ?

Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
      ?       ?     54.26%       0.00%

Coefficients

Term          Coef  SE Coef  T-Value  P-Value   VIF
Constant      23.9     34.7     0.69    0.562
X1           0.337    0.932     0.36    0.752  1.33
X2           1.289    0.539     2.39    0.139  1.44
X3         -0.0516   0.0433    -1.19    0.355  1.24
---------------------------------------------------------------------------------------
```

Find the missing F-value in the ANOVA table.

(a) 8.23 **(b)** 6.15 **(c)** 2.98 **(d)** 4.37 **(e)** 9.16

6. Referring to Question #5 above, can we conclude (at the 5% significance level) that mean arterial blood pressure is linearly related to at least one of duration of hypertension, basal pulse, or the measure of stress?

(a) Yes, because the p-value is equal to .261, which is greater than .05.
(b) No, because the p-value is equal to .261, which is greater than .05.
(c) No, because all 3 p-values, $.752,\ .139,$ and .355 are greater than .05.
(d) Yes, because all 3 p-values, $.752,\ .139,$ and .355 are greater than .05.
(e) Yes, because the adjusted R^2 is bigger than 50%.

7. Suppose that a student in this class uses their personalized class data set to do a multiple regression with `height` as the y-variable, and `distance` and `study` as the two x-variables, and produces the Minitb output below.

```
---------------------------------------------------------------------------------------
Regression Analysis: height versus distance, study

Analysis of Variance

Source       DF   Adj SS  Adj MS  F-Value  P-Value
Regression    2    49.94  24.971     0.69    0.502
  distance    1    41.40  41.400     1.15    0.285
  study       1     9.00   8.998     0.25    0.618
Error       181  6526.77  36.060
Total       183  6576.71

Model Summary

      S   R-sq  R-sq(adj)  R-sq(pred)
6.00496  0.76%      0.00%       0.00%

Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant   66.97     5.64    11.87    0.000
distance  -0.968    0.904    -1.07    0.285  1.00
study      0.68     1.35     0.50    0.618  1.00

Regression Equation
height = 66.97 - 0.968 distance + 0.68 study
---------------------------------------------------------------------------------------
```

And suppose that `height` is in fact not linearly related to either of `distance` or `study`. Now suppose that 500 students in this class each repeat the above analysis with their own personalized class data set, and use the corresponding output to test whether or not `height` is linearly related to at least one of `distance` or `study`. Which of the following statements is true.

(a) Because `height` is in fact not linearly related to at least one of `distance` or `study`, none of those students would obtain a p-value less than .01 (in place of the `.502` in the above output)
(b) Approximately 50.2% of those students would fail to reject the null hypothesis.
(c) Approximately 25 of those students would obtain a p-value less than .01 (in place of the `.502` in the above output), and therefore falsely conclude (at the 1% significance level) that `height` is linearly related to at least one of `distance` or `study`.
(d) Approximately 5 of those students would obtain a p-value less than .01 (in place of the `.502` in the above output), and therefore falsely conclude (at the 1% significance level) that `height` is linearly related to at least one of `distance` or `study`.
(e) Approximately 50.2% of those students would reject the null hypothesis.

8. A researcher collects the following data,

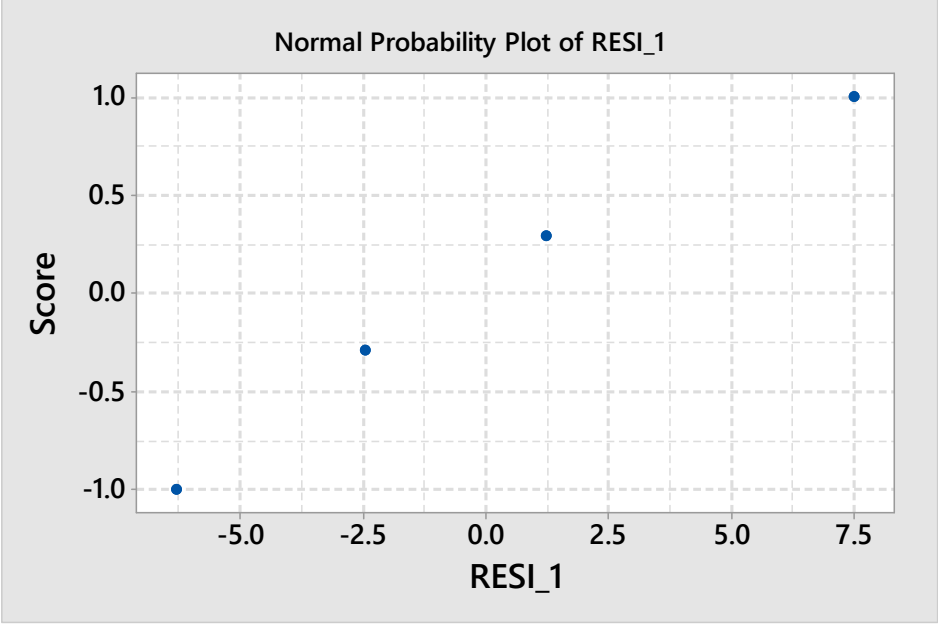
Subject	Y	X_1	X_2
1	118	5.1	63
2	123	3.8	70
3	?	8.2	72
4	108	9.3	71

fits a multiple regression model and produces the following Minitab output and normal probability plot of residuals.

```
Regression Analysis: Y versus X1, X2

Source       DF  Adj SS  Adj MS  F-Value  P-Value
Regression    2   70.03   35.01     0.34    0.771
  X1          1   69.31   69.31     0.67    0.563
  X2          1   25.49   25.49     0.25    0.706
Error         1  102.97  102.97
Total         3  173.00

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant      75      108     0.70    0.612
X1         -2.18     2.66    -0.82    0.563  1.37
X2          0.84     1.68     0.50    0.706  1.37
```



Find the missing y-value.
(a) 125 **(b)** 122 **(c)** 118 **(d)** 115 **(e)** 111

9. A biologist wanted to know if the cowpea weevil has a preference for one type of bean over others as a place to lay eggs. She put equal amounts of three types of seeds into a jar and added adult cowpea weevils. After a few days she observed the following data:

Type of Bean	Number of Eggs
Pinto	167
Cowpea	176
Northern Beans	194

Do these data provide significant evidence of a preference for some types of beans over others? Use $\alpha = .05$.

(a) No, because 2.11 is less than 5.991
(b) No, because 2.11 is less than 7.815
(c) Yes, because 9.81 is greater than 5.991
(d) Yes, because 9.81 is greater than 7.815
(e) Yes, because 194 is much bigger than 167 and 176

10. Marine biologists have noticed that the color of the outermost growth band on a clam tends to be related to the time of the year in which the clam dies. A biologist conducted a small investigation of whether this is true for the species *Protothaca staminea*. She collected a sample of 78 clam shells from this species and classified them according to the month that the clam died (February or March) and the color of the outermost growth band (clear, dark, or unreadable). The data are given in the Minitab output below.

```
-------------------------------------------------------------------------------
Rows: Month   Columns: Color

           1      2      3    All

1          ?     26      ?     44
           ?      ?      ?

2          6      ?      3      ?
           ?      ?      ?

All        ?      ?     12      ?

Cell Contents:      Count
                    Expected count

Chi-Sq = 2.377, DF = 2, P-Value = 0.305
-------------------------------------------------------------------------------
```

What conclusion can we draw from the above output using $\alpha = .05$?

(a) We conclude that month of death and color are dependent since $2.377 > .305$
(b) We conclude that month of death and color are dependent since $.305 > .05$
(c) We cannot reject the hypothesis that month of death and color are independent since $.305 > .05$
(d) We conclude that month of death and color are independent since $.305 > .05$
(e) We cannot reject the hypothesis that month of death and color are independent since $2.377 < 3.841$

11. Marine biologists have noticed that the color of the outermost growth band on a clam tends to be related to the time of the year in which the clam dies. A biologist conducted a small investigation of whether this is true for the species *Protothaca staminea*. She collected a sample of 78 clam shells from this species and classified them according to the month that the clam died (February or March) and the color of the outermost growth band (clear, dark, or unreadable). The data are given in the Minitab output below.

```
------------------------------------------------------------------------------
Rows: Month   Columns: Color

           1      2      3    All

1          ?     26      ?     44
           ?      ?      ?

2          6      ?      3      ?
           ?      ?      ?

All        ?      ?     12      ?

Cell Contents:      Count
                    Expected count

Chi-Sq = 2.377, DF = 2, P-Value = 0.305
------------------------------------------------------------------------------
```

If a chi-square analysis were performed on the above data, what would be the value in cell (1,1) of the expected table?

(a) 5.67 **(b)** 8.46 **(c)** 9.35 **(d)** 7.32 **(e)** 9.16

Answers (3rd Sample Test for Sections 9.1-9.5, 10.1-10.3):

1. b **2.** d **3.** e **4.** a **5.** c **6.** b **7.** d **8.** a **9.** a **10.** c **11.** b