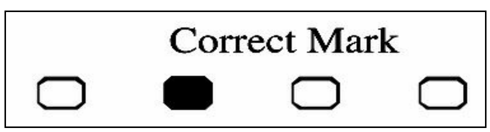
Instructor: Prof. Murdock
Duration: 50 minutes. You must stay in the test room the entire time.
Format: 18 multiple-choice questions with answers recorded on SCANTRON form. Total possible points are 90.
Allowed aids: A non-programmable calculator (and attached aid sheets, which you may detach)

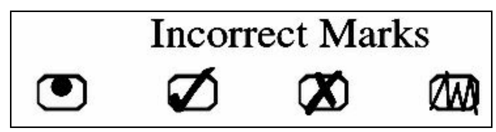
INSTRUCTIONS:
Do NOT write your answers on these test papers; You MAY do scratch work on these pages
ONLY those answers correctly marked on the SCANTRON form can earn positive marks

Correct answers are worth:	**+ 5.00 points**
Incorrect answers are worth:	**0 points**

- Use only a pencil or blue or black ball point pen

- **Pencil strongly recommended**, it can be erased if a mistake is made
- Make dark solid marks that fill the bubble completely





- Select the one best alternative
- Erase completely any marks you want to change
  - Crossing out a marked box is not acceptable and is incorrect

---

1st : Print your **LAST NAME** and **INITIALS** in boxes provided
- Use exact name you are officially registered under
- Darken each letter in the corresponding bracket below each box

2nd : Print your 9 digit **STUDENT NUMBER** in the boxes provided
- Fill in zeros in front of the number if less than 9 digits
- Darken each number in the corresponding bracket below each box

3rd : Print 2 digit **FORM** number in the boxes provided
- Your FORM number is **02**
- Darken each number in the corresponding bracket below each box

4th : Sign your name in the **SIGNATURE** box

► **Questions (1) – (5):** A simple regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ is estimated with n = 26.

(1) For the slope what is the test of statistical significance?

(A) H_0: $\beta_1 = 0$; H_1: $\beta_1 > 0$
(B) H_0: $\beta_1 = 0$; H_1: $\beta_1 \neq 0$
(C) H_0: $\beta_1 \geq 0$; H_1: $\beta_1 = 0$
(D) H_0: $\beta_1 > 0$; H_1: $\beta_1 = 0$
(E) H_0: $\beta_1 \neq 0$; H_1: $\beta_1 = 0$

(2) For that test of statistical significance if t = -1.10 what is the <u>best</u> conclusion?

(A) There is no relationship between y and x
(B) There is no linear relationship between y and x
(C) Any observed relationship between y and x is spurious
(D) There is a negative correlation between y and x but it is not statistically different from 0
(E) There is a statistically significant relationship between y and x at the 5% significance level

(3) Which is an <u>INCORRECT</u> statement about the residuals?

(A) $e_i = \varepsilon_i$
(B) $\sum_{i=1}^{26} e_i = 0$
(C) $e_i = y_i - \hat{y}_i$
(D) $e_i = \hat{y}_i - y_i$
(E) $e_i = y_i - b_0 - b_1 x_i$

(4) If the slope estimate is zero then the constant term estimate will be _______.

(A) 0
(B) 1
(C) β_0
(D) β_1
(E) $\bar{y}$

(5) Increasing the sample size to 200 would cause ____________________.

(A) a decrease in s^2
(B) a decrease in β_0
(C) a decrease in σ^2
(D) no change in the SSR
(E) an increase in the SST

► **Questions (6) – (7):** "For restaurant meals what effect does salt have on tips?" 35 randomly selected customers order a meal (may choose anything on the menu). The researcher secretly and randomly changes the usual salt content. Meals have between 50 to 150 percent of the usual amount of salt. For a simple regression the dependent variable is how much money the customer gives as a tip and the independent variable is the salt content percentage (relative to usual recipe).

(6) By varying the salt content a lot, ______ is reduced.

(A) s_b
(B) σ^2
(C) $E[\varepsilon]$
(D) SSE
(E) s (s_ε)

(7) Considering the described approach to the research question, are these valid criticisms?

#1	The results will be biased because the sample size is too small
#2	The results will be biased because these data are observational
#3	The results will be biased because the model fails to control for other factors that affect the size of the tip

(A) Only criticism #2 is valid
(B) Only criticism #3 is valid
(C) Only criticisms #2 and #3 are valid
(D) Criticisms #1, #2 and #3 are all valid
(E) None of these three criticisms is valid

► **Questions (8) – (9):** For $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i$ with a sample of 243, consider the test of the overall statistical significance of the model.

(8) For α = 0.10, what is the rejection region?

(A) F < 1.94
(B) F < 2.37
(C) F < -2.37
(D) F > 1.94
(E) F > 2.37

(9) A Type I error is obtaining results that are ______________________________________.

(A) statistically significant when all of the slope parameters are zero
(B) statistically significant when any of the slope parameters are zero
(C) not statistically significant when all of the slope parameters are zero
(D) not statistically significant when all of the slope parameters are not zero
(E) not statistically significant when not all of the slope parameters are zero

► **Questions (10) – (11):** For a random sample of 62 Canadians, hours spent watching television per week (tv_hrs_week) and weight in kilograms (weight_kg) are used in a simple regression:

```
weight_kg-hat = 63.07 + 1.68*tv_hrs_week
                        (0.67)
```

(10) To test if the slope is statistically positive, what is the approximate p-value?

(A) p-value < 0.005
(B) 0.005 < p-value < 0.010
(C) 0.010 < p-value < 0.025
(D) 0.025 < p-value < 0.050
(E) 0.050 < p-value < 0.100

(11) We have good evidence that television viewing ________________________________.

(A) is not related to weight
(B) is a significant cause of obesity
(C) is higher among people who weigh more
(D) increases weight by 1.68 kg for each additional hour per week
(E) increases weight by a significant amount holding all other factors constant

► **Questions (12) – (13):** Consider the following multiple regression estimation results.

```
      Source |       SS       df       MS              Number of obs =      82
-------------+------------------------------           F(  2,    79) =    5.81
       Model |  274.637314     2  137.318657           Prob > F      =  0.0044
    Residual |  1868.29754    79  23.6493359           R-squared     =  0.1282
-------------+------------------------------           Adj R-squared =  0.1061
       Total |  2142.93485    81  26.4559858           Root MSE      =  4.8631

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .3167906   .2348559     1.35   0.181    -.1506783    .7842595
          x2 |   .2049877   .1306764     1.57   0.121    -.0551172    .4650927
       _cons |   .1858295   .6019955     0.31   0.758    -1.012412    1.384071
------------------------------------------------------------------------------
```

(12) If x1 and x2 are related to each other then ____________________________________.

(A) there is a spurious correlation
(B) the coefficient estimates are biased
(C) we should drop either x1 or x2 from the model
(D) a scatter diagram of y and x1 may show a negative relationship
(E) we have a violation of the homoscedasticity assumption (heteroscedasticity)

(13) Referring back to the STATA output, is the multiple regression model statistically significant?

(A) Yes, we have overwhelming evidence (significance level < 0.5%)
(B) Yes, at a 1% significance level
(C) Yes, at a 5% significance level
(D) Yes, at a 10% significance level
(E) No, not at any conventional significance level (significance level > 10%)

► **Questions (14) – (15):** For an Ontario university (not U of T) a researcher randomly selects 363 students who have completed their degree. Marks during high school (Best-6, mark out of 100), which affect university admissions, are used to predict university GPA (scale from 0 to 12). For the sample, the average mark on these high school tests is 90 and the standard deviation is 3.

SUMMARY OUTPUT					
Regression Statistics					
Multiple R	0.4883				
R Square	0.2385				
Adjusted R Square	0.2363				
Standard Error	0.8295				
Observations	363				
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	77.78	77.78	113.04	0.0000
Residual	361	248.39	0.69		
Total	362	326.17			
	Coefficients	*Standard Error*	*t Stat*	*P-value*	
Intercept	-5.35	1.31	-4.08	0.0001	
Best-6	0.15	0.01	10.63	0.0000	

(14) Suppose Wendy Lee has a 92 for the high school mark (Best-6). You could be 95% sure that her university GPA would fall in which interval?

(A) (6.82, 10.08)
(B) (7.08, 9.82)
(C) (8.32, 8.58)
(D) (8.35, 8.55)
(E) (8.39, 8.54)

(15) Consider students with an 85 for the high school mark (Best-6). You could be 95% sure that the average university GPA for this group would fall in which interval?

(A) (5.77, 9.03)
(B) (6.03, 8.77)
(C) (7.23, 7.57)
(D) (7.26, 7.54)
(E) (7.38, 7.42)

► **Questions (16) – (18):** "How much does the lighting in classrooms affect students' ability to pay attention to lectures?" A researcher does an experiment by randomly setting lighting in selected classrooms. Lighting is measured in lux: low levels of lux (e.g. 30 lux) are dim lighting and high levels of lux (e.g. 400 lux) are bright lighting. After randomly selected lectures a randomly selected student is asked: "How well did the lecture keep your attention: 0 means not at all and 100 means it was captivating?" For 341 observations the researcher records the lux and the student response.

```
      Source |       SS       df       MS              Number of obs =     341
-------------+------------------------------           F(  1,   339) =   17.77
       Model |  5701.18014     1  5701.18014           Prob > F      =  0.0000
    Residual |  108774.837   339  320.869727           R-squared     =
-------------+------------------------------           Adj R-squared =
       Total |  114476.018   340  336.694169           Root MSE      =  17.913

------------------------------------------------------------
   attention |      Coef.   Std. Err.      t    P>|t|
-------------+----------------------------------------------
         lux |    .036666   .0086985     4.22   0.000
       _cons |   51.26886   2.193176    23.38   0.000
------------------------------------------------------------
```

(16) Do we have sufficient evidence to conclude that each additional unit of lux increases students' attention rating by at least 0.025? (Choose the best answer.)

(A) Yes, we have overwhelming evidence (significance level < 0.5%)
(B) Yes, at a 1% significance level
(C) Yes, at a 5% significance level
(D) Yes, at a 10% significance level
(E) No, not at any conventional significance level (significance level > 10%)

(17) What is the coefficient of determination (R^2)?

(A) 0.0025
(B) 0.0366
(C) 0.0498
(D) 0.9502
(E) 0.9634

(18) What is the 98% confidence interval estimator of the effect of each additional unit of lux on students' attention rating?

(A) (0.014, 0.059)
(B) (0.016, 0.057)
(C) (0.018, 0.055)
(D) (0.020, 0.054)
(E) (0.022, 0.051)