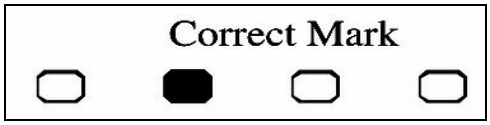
Instructor: Prof. Murdock
Duration: 50 minutes. You must stay in the test room for the entire time.
Format: 18 multiple choice questions with answers recorded on SCANTRON form
Point values: Each question worth 5 points. There are 90 total possible points.
Allowed aids: A non-programmable calculator (and attached aid sheets, which you may detach)

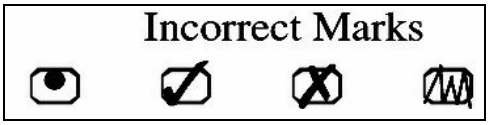
INSTRUCTIONS:

Do NOT write your answers to the multiple choice questions on these test papers
ONLY those answers correctly marked on the SCANTRON form can earn marks
You MAY do scratch work on these pages

- 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





- Erase completely any marks you want to change
  - Crossing out a marked box is not acceptable and is incorrect
- Select the one best alternative
- Correct answers are worth 5 points and incorrect answers are worth 0 points
- If you choose alternative **(E)** "Don't know" you will be awarded partial credit of 1.3 points for being able to recognize that you don't know the answer

---

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 **01**
- Darken each number in the corresponding bracket below each box

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

For the 18 questions, choose the best answer and mark it on the SCANTRON form.

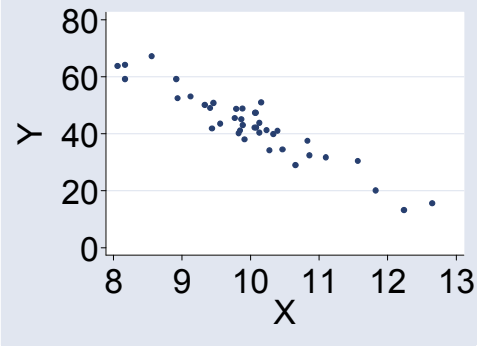
(1) In regression analysis, what is the standard error of estimate?

(A) The estimate of the standard error of the sample mean
(B) The estimate of the standard error of the slope
(C) The standard deviation of the residuals
(D) The standard deviation of the dependent variable
(E) Don't know

(2) Which of the following will be biased if the homoscedasticity assumption is violated?

(A) SSE
(B) R^2
(C) t test statistics
(D) Coefficient estimates
(E) Don't know

► For Question **(3)**: Given the following graph, consider what would happen to the estimated regression line with the introduction of an outlier that is the result of a human error. The position of the outlier is specified: (x, y).



(3) Which of the following outliers would cause a <u>downward bias</u> of the estimated slope?

(A) (6, -10)
(B) (8, 0)
(C) (8, 120)
(D) (13, 140)
(E) Don't know

(4) If you observed all of the variables that affect the dependant variable and included them in a multiple regression model, then what could you say about the SSE?

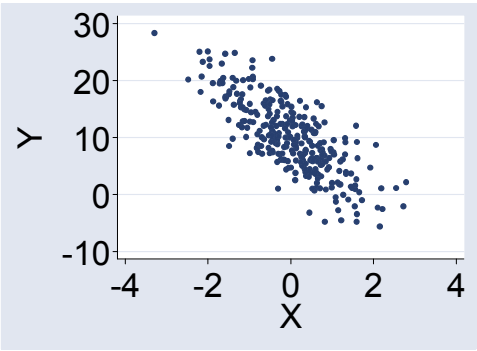
(A) SSE = 0
(B) SSE = SSR
(C) SSE = SST
(D) SSE = SSR - SST
(E) Don't know

(5) Suppose X is an independent variable measured in thousands of dollars and Y is the dependent variable measured in years. Which of the following illustrate the concept of regression towards the mean?

I.	When X increases by 1,000 dollars, Y increases by less than 1 year
II.	When X is 2 standard deviations above its mean, Y is less than 2 standard deviations away from its mean
III.	When X is at its average value, Y is less than its average value

(A) I.
(B) II.
(C) III.
(D) I. and III.
(E) Don't know

► For Question **(6)**: Consider the following graph.



(6) Which of the following is closest to the least squares line?

(A) Y-hat = 10 - 5*X
(B) Y-hat = 10 - 15*X
(C) Y-hat = 30 - 5*X
(D) Y-hat = 30 - 15*X
(E) Don't know

► For Questions **(7) – (10)**: Consider this regression output. Note: Some output has been erased.

```
      Source |       SS       df       MS              Number of obs =     300
-------------+------------------------------           F(  1,   298) =
       Model |   1200.7887     1   1200.7887           Prob > F      =
    Residual |  269043.371   298  902.830104           R-squared     =
-------------+------------------------------           Adj R-squared =
       Total |   270244.16   299  903.826621           Root MSE      =  30.047

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |     -1.002   .8688352
       _cons |   118.7474   1.734772
------------------------------------------------------------------------------
```

(7) What is the Adjusted R^2?

(A) 0.0011
(B) 0.0044
(C) 0.0077
(D) 0.0099
(E) Don't know

(8) What is 95% confidence interval estimate of the slope coefficient on x?

(A) LCL = -3.24 and UCL = 1.24
(B) LCL = -2.71 and UCL = 0.71
(C) LCL = -2.43 and UCL = 0.43
(D) LCL = -1.87 and UCL = -0.13
(E) Don't know

(9) What is the p-value for the test of "statistical significance" of the coefficient on x?

(A) 0.05
(B) 0.15
(C) 0.20
(D) 0.25
(E) Don't know

(10) What is the variance of the variable x?

(A) 0.75
(B) 0.87
(C) 3.02
(D) 4.00
(E) Don't know

(11) Suppose a simple regression is estimated where Y is the dependent variable and X is an explanatory variable. Based on observing an R^2 of 0.03, what can you conclude?

I.	There is a spurious correlation between X and Y
II.	There is no relationship between X and Y
III.	Many other factors aside from X affect Y

(A) I.
(B) II.
(C) III.
(D) I., II., and III.
(E) Don't know

► For Questions **(12) – (13)**: Consider the following cross-tabulation. For these data a simple regression is estimated where Y is the dependent variable and X is the independent variable.

x \ y	1	7	Total
1	55	0	55
3	0	33	33
Total	55	33	88

(12) What is the estimated slope coefficient?

(A) 0
(B) 1
(C) 2
(D) 3
(E) Don't know

(13) What is the standard error of the estimated slope coefficient?

(A) 0.00
(B) 0.05
(C) 0.08
(D) 0.10
(E) Don't know

► For Questions **(14) – (15)**: Consider this regression output. Note: Some output has been erased.

```
      Source |       SS       df       MS
-------------+------------------------------
       Model |   150.85374     5  30.1707481
    Residual |   3332.7987   214  15.5738257
-------------+------------------------------
       Total |  3483.65244   219  15.9070888

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .2617972   .2637235     0.99   0.322
          x2 |  -.7170306   .2658999    -2.70   0.008
          x3 |  -.3089638   .2907126    -1.06   0.289
          x4 |   .0961426   .2716193     0.35   0.724
          x5 |   .0375474   .2864768     0.13   0.896
       _cons |   15.33202     .26859    57.08   0.000
------------------------------------------------------------------------------
```

(14) Is this regression model statistically significant?

(A) Yes, $0 \leq$ p-value ≤ 0.001
(B) Yes, $0.001 <$ p-value ≤ 0.01
(C) Yes, $0.01 <$ p-value ≤ 0.05
(D) No, p-value > 0.05
(E) Don't know

(15) What should we conclude about the relationship between y and x1?

(A) There is no linear relationship: the slope is 0
(B) There is no non-linear relationship
(C) We have insufficient evidence to infer that there is a linear relationship
(D) There is a positive linear relationship: a one unit increase in x1 is associated with a 0.26 unit increase in y
(E) Don't know

(16) A researcher is concerned about the violent atmosphere and negative peer pressure in prisons and would like to measure the effect of incarceration on the number of future crimes. Suppose for a random sample of adults the number of arrests is regressed on the number of years spent incarcerated. The following simple regression results are obtained.

```
ARRESTS-hat = 1.02 + 0.66*YRS_INCAR
              (0.28) (0.54)
```

Which is a correct conclusion?

(A) The effect of incarceration on arrests is understated by the estimated slope
(B) The effect of incarceration on arrests is overstated by the estimated slope
(C) There no relationship between arrests and years spent incarcerated
(D) Each additional year of incarceration results in 0.66 more arrests, but this effect is not statistically significant
(E) Don't know

► For Questions **(17) – (18)**: For a random sample of 63 adults, the number of calories consumed is recorded. In addition, each person's age (in years) and height (in centimeters) are measured. The dependant variable is the natural logarithm of the number of calories.

Variable	Coefficient Estimate (standard error)
AGE	0.003 (0.002)
HEIGHT	0.019 (0.005)

(17) What is the interpretation of the coefficient on height?

(A) It is not statistically different from zero: height does not affect calories consumed
(B) A one centimeter increase in height is associated with 0.019 more calories consumed
(C) A one centimeter increase in height is associated with 1.9 percent more calories consumed
(D) A one percent increase in height is associated with 1.9 more calories consumed
(E) Don't know

(18) At a 5% significance level, what is the rejection region for the test of overall statistical significance of the model of caloric intake?

(A) $F > 3.15$
(B) $F < 3.15$
(C) $F < -3.15$ or $F > 3.15$
(D) $t < -2.00$ or $t > 2.00$
(E) Don't know