While you wait for the start of this test, you may fill in the FRONT AND BACK of the BUBBLE FORM and read this cover. BUT, keep these test papers face up and flat on your desk.

Instructor: Prof. Murdock
Duration: 110 minutes. You MUST STAY for at least 60 minutes
Allowed aids: A non-programmable calculator and the aid sheets provided with this test
Format: This test includes these question papers and a BUBBLE FORM. There are 38 multiple choice questions with point values that may vary from 1 to 3 points each for a total of 69 points. The point value for each question is shown by [1pt], [2pts], [3pts]. Most questions have choices **(A)** – **(E)**. For questions with fewer choices, the correct answer is ALWAYS one of those offered (e.g. if the choices are **(A)** – **(D)**, then **(E)** is NOT a possible correct answer.)

Once the start of the test is announced, you may detach the aid sheets and statistical tables (Standard Normal, Student *t* and *F*) from the end of this test. These question papers and the aid sheets will not be collected. Only the BUBBLE FORM is collected.

You must record your answers on the BUBBLE FORM. In ALL cases what is (or is not) indicated on the BUBBLE FORM is your FINAL ANSWER. Marks are based SOLEY on the BUBBLE FORM, which must be completed before the end of the test is announced.

- On the FRONT of the BUBBLE FORM:
    - Print your 9 (or 10) digit student number in the boxes AND darken each number in the corresponding circles.
    - Print your last name and initial in the boxes AND darken each letter in the corresponding circles.
    - Fill in the upper left region of the form.
    - ******Your FORM CODE is A. Darken the circle with the letter A.******
        - Failing to indicate your FORM CODE means that your answers will be out of sync compared to the solution key used to mark your paper. It is entirely *your responsibility* to properly indicate your FORM CODE.
- On the BACK of the BUBBLE FORM:
    - Write in your name, sign, and record your answers.
- Use a pencil and make dark solid marks that fill the bubble completely.
- Erase completely any marks you want to change; Crossing out a marked box is incorrect.
- Choose the best answer for each question. If more than one answer is selected that question earns 0 points.
- For questions with numeric answers that require rounding, round your final answer to be consistent with the choices offered. Use standard rounding rules.

▶ **Questions (1) – (9):** Recall the publically available data for all ON public sector employees with salaries of $100,000 or more (http://www.fin.gov.on.ca/en/publications/salarydisclosure/pssd/). Consider the 15,583 university employees in the disclosure of 2013 salaries. There are 34 Ontario universities in that disclosure. Some have few employees listed (e.g. University of Sudbury) while others have over one thousand (e.g. York University). Regression #1 includes indicator (dummy) variables for each of the nine universities with at least 750 listed employees: the omitted category is the remaining 25 universities. Regression #2 includes indicator (dummy) variables for universities with at least 250 listed employees: the omitted category is the remaining 19 universities. Salaries are measured in $1,000's of dollars.

REGRESSION #1:

```
      Source |       SS       df       MS              Number of obs =   15583
-------------+------------------------------           F(  9, 15573) =   44.89
       Model |  573262.542     9   63695.838           Prob > F      =  0.0000
    Residual |    22095175 15573  1418.81301           R-squared     =  0.0253
-------------+------------------------------           Adj R-squared =  0.0247
       Total |  22668437.6 15582  1454.78357           Root MSE      =  37.667

------------------------------------------------------------------------------
 salary_1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      guelph |   3.437123   1.466413     2.34   0.019     .5627833    6.311463
    mcmaster |   13.76919   1.327661    10.37   0.000     11.16682    16.37156
      ottawa |   4.027757    1.23554     3.26   0.001     1.605954     6.44956
      queens |   9.581655   1.373348     6.98   0.000     6.889734    12.27358
     ryerson |   4.443582   1.373924     3.23   0.001     1.750531    7.136634
        uoft |   17.13051   .9251845    18.52   0.000     15.31704    18.94397
    waterloo |   7.767967   1.278097     6.08   0.000     5.262749    10.27319
     western |   5.760823   1.259841     4.57   0.000     3.291387    8.230259
        york |   8.273656   1.129976     7.32   0.000     6.058771    10.48854
       _cons |   136.1973   .6227887   218.69   0.000     134.9765     137.418
------------------------------------------------------------------------------
```

REGRESSION #2:

```
      Source |       SS       df       MS              Number of obs =   15583
-------------+------------------------------           F( 15, 15567) =   29.16
       Model |  619448.089    15  41296.5393           Prob > F      =  0.0000
    Residual |  22048989.5 15567  1416.39298           R-squared     =  0.0273
-------------+------------------------------           Adj R-squared =  0.0264
       Total |  22668437.6 15582  1454.78357           Root MSE      =  37.635

------------------------------------------------------------------------------
 salary_1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       brock |   6.427058   2.085241     3.08   0.002     2.339743    10.51437
    carleton |  -3.365421   1.948717    -1.73   0.084    -7.185133    .4542909
      guelph |   5.114239   1.878237     2.72   0.006     1.432676    8.795803
    lakehead |  -2.443387   2.633932    -0.93   0.354    -7.606201    2.719426
  laurentian |   3.249077    2.43601     1.33   0.182    -1.525785     8.02394
    mcmaster |    15.4463   1.772217     8.72   0.000     11.97255    18.92006
      ottawa |   5.704874   1.704416     3.35   0.001      2.36402    9.045727
      queens |   11.25877   1.806639     6.23   0.000     7.717549    14.79999
     ryerson |   6.120699   1.807076     3.39   0.001     2.578619    9.662779
        uoft |   18.80762   1.495187    12.58   0.000     15.87688    21.73836
    waterloo |   9.445084    1.73546     5.44   0.000      6.04338    12.84679
     western |    7.43794   1.722083     4.32   0.000     4.062457    10.81342
     wilfrid |   4.140295   2.162565     1.91   0.056    -.0985828    8.379174
     windsor |   4.919408   2.133302     2.31   0.021     .7378877    9.100929
        york |   9.950772   1.629646     6.11   0.000     6.756477    13.14507
       _cons |   134.5202   1.329767   101.16   0.000     131.9137    137.1267
------------------------------------------------------------------------------
```

(1) [1pt] The dummy variable for the University of Toronto is uoft. For Regression #1, what does the coefficient on uoft mean?

(A) Compared to the 25 universities with fewer than 750 employees listed in the disclosure, U of T employees on average have salaries that are $17,131 higher
(B) Compared to Guelph, McMaster, Ottawa, Queens, Ryerson, Waterloo, Western and York, U of T employees on average have salaries that are $17,131 higher

(2) [2pts] How much of the variation in salaries across the ON university employees listed in the 2013 disclosure is explained by which ON university an employee works for?

(A) almost none of the variation
(B) less than a quarter but more than 5 percent of the variation
(C) less than half but more than one quarter of the variation
(D) more than half but less than 95 percent of the variation
(E) nearly all of the variation

(3) [2pts] What is the standard deviation of salary for university employees listed in the 2013 disclosure?

(A) $1,335
(B) $35,242
(C) $36,798
(D) $37,667
(E) $38,142

(4) [2pts] Consider a dummy variable other that is = 1 if an employee works for one of the 25 universities with fewer than 750 employees listed and is = 0 otherwise. For Regression #1, a mathematically equivalent approach would have been to include the dummy variable other and make the University of Toronto the omitted category. If U of T were the omitted category in Regression #1, what would the coefficient on the McMaster dummy be?

(A) -17.13051
(B) -13.76919
(C) -3.36132
(D) 13.76919
(E) 30.89969

(5) [2pts] Comparing Regression #2 with Regression #1, why are the coefficients on the dummy variables that appear in both regressions all higher in Regression #2 compared to Regression #1?

(A) because Regression #1 has a smaller value of k
(B) because overall the 19 universities with fewer than 250 listed employees have lower average salaries than the 25 universities with fewer than 750 listed employees
(C) because average salaries at Brock, Carleton, Lakehead, Laurentian, Wilfrid Laurier, and Windsor, are each lower than at each of the 9 universities included in Regression #1
(D) All of the above

(6) [2pts] For Regression #2, if salary were measured in tens of thousands of dollars (instead of thousands of dollars), which of these statements are true?

(A) The value of the SST would become 2266843.76
(B) The value of the Root MSE would become 376.35
(C) The value of the intercept would become 1345.202
(D) The value of the coefficient on brock would become 0.6427058
(E) The value of the R-squared and the Adjusted R-squared would become 0.273 and 0.264

(7) [1pt] On average what is the difference in salaries for employees at Laurentian versus Lakehead?

(A) Mean salaries at Laurentian are about $2,400 higher than at Lakehead
(B) Mean salaries at Laurentian are about $3,250 higher than at Lakehead
(C) Mean salaries at Laurentian are about $5,700 higher than at Lakehead

(8) [2pts] Using Regression #2, what is the residual for a U of T employee who has a salary of $131,431?

(A) -21.9
(B) -15.7
(C) -3.1
(D) 3.1
(E) 15.7

(9) [2pts] Consider re-running Regression #1 for a random sample of 200 university employees. Is the regression below statistically significant overall?

```
      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  9,   190) =    1.87
       Model |  24329.7577     9  2703.30641           Prob > F      =  0.0583
    Residual |    274331.1   190   1443.8479           R-squared     =  0.0815
-------------+------------------------------           Adj R-squared =  0.0380
       Total |  298660.858   199  1500.80833           Root MSE      =  37.998

------------------------------------------------------------------------------
 salary_1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      guelph |   2.192202   13.87491     0.16   0.875    -25.17644    29.56085
    mcmaster |   18.08286   10.59714     1.71   0.090    -2.820285    38.98601
      ottawa |    8.70768   11.06011     0.79   0.432     -13.1087    30.52406
      queens |   16.63529   13.28421     1.25   0.212    -9.568196    42.83877
     ryerson |   27.64895   10.21166     2.71   0.007     7.506174    47.79173
        uoft |   26.89136   8.634328     3.11   0.002       9.8599    43.92281
    waterloo |   2.100756   11.06011     0.19   0.850    -19.71562    23.91713
     western |   25.64764   11.96456     2.14   0.033     2.047208    49.24808
        york |   15.34928   10.39603     1.48   0.141    -5.157187    35.85575
       _cons |   130.0389   5.664407    22.96   0.000     118.8657    141.2121
------------------------------------------------------------------------------
```

(A) No, not at any conventional significance level
(B) Yes, but only at a 10% significance level
(C) Yes, at a 5% significance level but not a 1% significance level
(D) Yes, at a 1% significance level but not a 0.1% significance level
(E) Yes, at 0.1% significance level

▶ **Questions (10) – (17):** Recall women's downhill ski racing in the Winter Olympics. Instead of the 2002 Salt Lake City games, consider those from the 2014 Sochi games. The dependent variable is the finish time measured in seconds. The x variable is start order (1st, 2nd , … 35th) of the skiers. The second regression also includes that variable squared.

REGRESSION #1:

```
      Source |       SS       df       MS              Number of obs =      35
-------------+------------------------------           F(  1,    33) =   23.90
       Model |  63.7083777     1  63.7083777           Prob > F      =  0.0000
    Residual |  87.9725269    33  2.66583415           R-squared     =  0.4200
-------------+------------------------------           Adj R-squared =  0.4024
       Total |  151.680905    34  4.46120308           Root MSE      =  1.6327

------------------------------------------------------------------------------
     seconds |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 start_order |   .1093066   .0223596     4.89   0.000     .0638156    .1547976
       _cons |   101.4167   .5568143   182.14   0.000     100.2839    102.5496
------------------------------------------------------------------------------
```

REGRESSION #2:

```
      Source |       SS       df       MS              Number of obs =      35
-------------+------------------------------           F(  2,    32) =   34.14
       Model |   103.27807     2  51.6390349           Prob > F      =  0.0000
    Residual |  48.4028348    32  1.51258859           R-squared     =  0.6809
-------------+------------------------------           Adj R-squared =  0.6609
       Total |  151.680905    34  4.46120308           Root MSE      =  1.2299

------------------------------------------------------------------------------
     seconds |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 start_order |   -.234679   .0693311    -3.38   0.002    -.3759018   -.0934562
start_order_sq |   .0080585   .0015756     5.11   0.000     .0048492    .0112678
       _cons |   103.8592   .6355807   163.41   0.000     102.5646    105.1538
------------------------------------------------------------------------------
```

(10) [1pt] For Regression #1, if finish time were measured in minutes what would the coefficient on start_order be?

(A) 0.0018
(B) 0.1093
(C) 6.5584

(11) [1pt] For Regression #1, if finish time were measured in minutes what would the standard error for the coefficient on start_order be?

(A) 0.0004
(B) 0.0224
(C) 1.3416

(12) [1pt] For Regression #1, if finish time were measured in minutes what would the Root MSE be?

(A) 0.0272
(B) 1.6327
(C) 97.962

(13) [2pts] For these 2014 data, is there sufficient evidence to suggest that a quadratic term is necessary to address a violation of the linearity assumption?

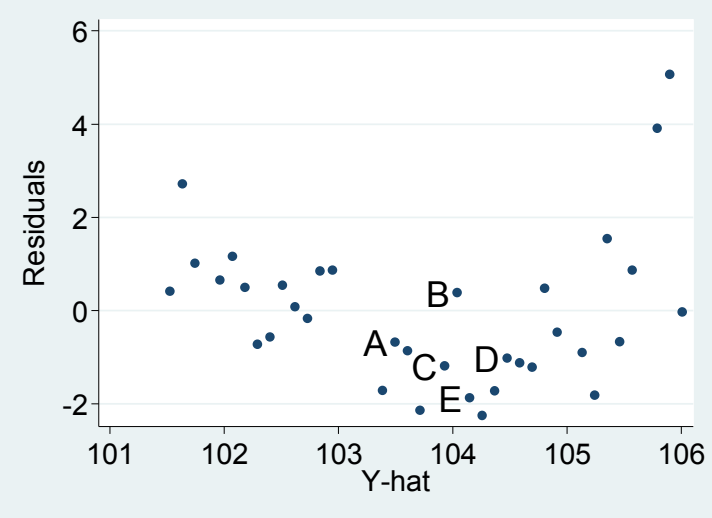
(A) No, not at any conventional significance level
(B) Yes, but only at a 10% significance level
(C) Yes, at a 5% significance level but not a 1% significance level
(D) Yes, at a 1% significance level but not a 0.1% significance level
(E) Yes, at 0.1% significance level

(14) [2pts] For Regression #2, what is the point estimate of the slope when start order is equal to 23?

(A) -1.1347
(B) -0.2347
(C) -0.2266
(D) -0.0493
(E) 0.1360

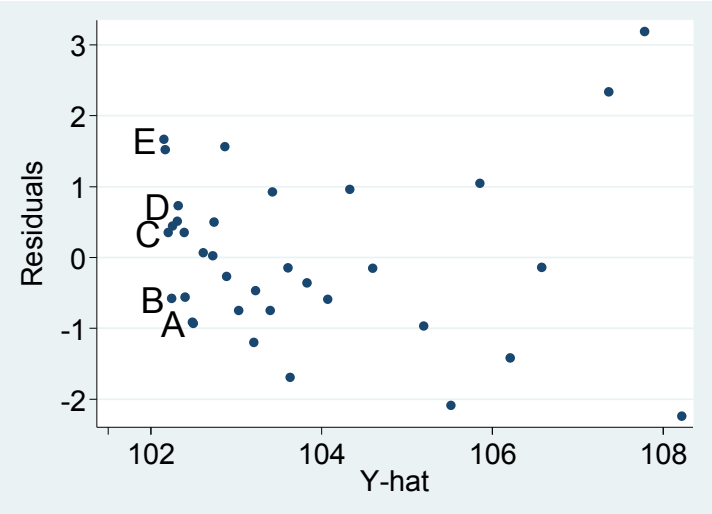
(15) [2pts] Consider the diagnostic plot to the right for Regression #1. Larisa Yurkiw (Canadian) skied 28th and finished with a time of 103.46 seconds. Which point corresponds to Larisa? (*Note:* The labels are immediately to the left of each observation.)

(A) Point A
(B) Point B
(C) Point C
(D) Point D
(E) Point E



(16) [2pts] Consider the diagnostic plot to the right for Regression #2. Stacey Cook (American) skied 10th and finished with a time of 103.05 seconds. Which point corresponds to Stacey? (*Note:* The labels are immediately to the left of each observation.)

(A) Point A
(B) Point B
(C) Point C
(D) Point D
(E) Point E



(17) [2pts] The order in which skiers race is *not* randomly assigned to each racer. Ski order is associated with a skier's skill. Page 805 of the textbook explains "Changing snow conditions can affect finish times, and in fact the top seeds can choose their starting positions and try to guess when the conditions will be best. Skiers expect conditions to improve and then, as the day wears on, to deteriorate." Skiing skill is a lurking (unobserved/omitted/confounding) variable in both Regression #1 and #2. Imagine the Olympic committee decided to allocate start order in a completely random manner, how should you expect the results for the regressions (Regression #1 and #2) to differ?

(A) The constant terms would likely be smaller
(B) The R-squared values would likely be smaller
(C) The quadratic term would likely not be necessary
(D) There would no longer be any statistically significant relationship between start order and finish time
(E) All of the above

▶ **Questions (18) – (19):** Consider this STATA regression output. Note some output has be intentionally erased.

```
      Source |       SS       df       MS              Number of obs =    3000
-------------+------------------------------           F(  3,  2996) =   13.88
       Model |  968562.389     3   322854.13           Prob > F      =  0.0000
    Residual |  69666417.6  2996  23253.1434           R-squared     =  0.0137
-------------+------------------------------           Adj R-squared =  0.0127
       Total |    70634980  2999  23552.8443           Root MSE      =  152.49

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |    .526466   .2786454                     -.0198896    1.072822
           w |  -8.261005   1.380535    -5.98   0.000     -10.9679   -5.554112
           z |   4.252782                      0.122
       _cons |    12.0056   2.785398     4.31   0.000     6.544112    17.46708
------------------------------------------------------------------------------
```

(18) [2pts] What is the value missing under `P>|t|` for the coefficient on x?

(A) 0.03
(B) 0.06
(C) 0.08
(D) 0.12
(E) 0.16

(19) [2pts] What is the value missing under `Std. Err.` for the coefficient on z?

(A) 2.50
(B) 2.75
(C) 3.00
(D) 3.25
(E) 3.50

▶ **Questions (20) – (28):** Recall the readings and study materials for "Asiaphoria Meets Regression to the Mean," *NBER Working Paper 20573*, Oct. 2014, by Lant Pritchett and Larry Summers. These results use the more recent PWT 8.1 data.[1]

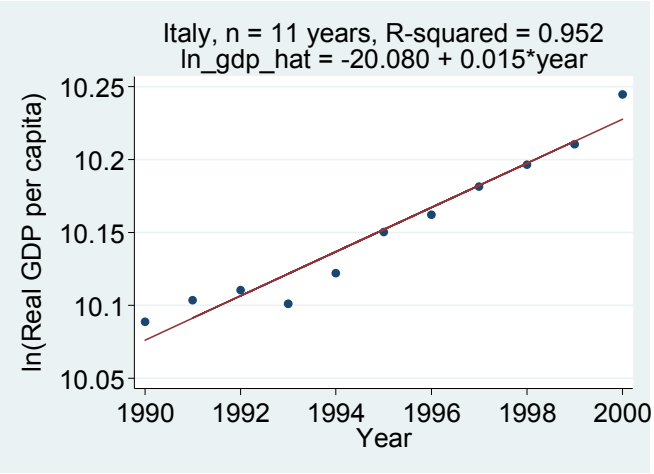


Table 1: Little persistence in cross-national growth rates across decades

Period 1	Period 2	Regression Coefficient	R-squared	N
PANEL A: *Adjacent decades*				
1950 – 60	1960 – 70	0.3375783	0.1236	66
1960 – 70	1970 – 80	0.4084345	0.1234	108
1970 – 80	1980 – 90	0.3225473	0.1138	142
1980 – 90	1990 – 00	0.2884994	0.1304	142
1990 – 00	2000 – 10	0.2051206	0.0562	142
PANEL B: *Two decades apart*				
1950 – 60	1980 – 90	-0.0475639	0.0020	66
1960 – 70	1990 – 00	0.1580633	0.0234	108
1970 – 80	2000 – 10	-0.0148128	0.0005	142
Source: Calculations based on PWT 8.1.				

(20) [2pts] The Penn World Tables report "Real GDP at constant 2005 national prices (in millions US$, 2005)" for each country in each year. In the 1990's, on average the Italian economy grew by about ____ per year.

(A) 1.5 percent
(B) 0.015 percent
(C) 1.5 million US$, 2005
(D) 0.015 million US$, 2005
(E) Italy's annual growth during that decade is not statistically different from zero

(21) [2pts] How many regressions like the one shown for Italy had to be estimated to create the raw data necessary to estimate the *first regression reported in Panel A* in Table 1?

(A) 2
(B) 4
(C) 66
(D) 132
(E) more than 200

(22) [1pt] How many regressions like the one shown for Italy had to be estimated to create the raw data necessary to estimate the *third regression in Panel B compared to the first regression in Panel B* in Table 1?

(A) the same number of first-stage regressions were necessary for both
(B) the third row regression required about twice as many first-stage regressions
(C) the third row regression required about four times as many first-stage regressions

[1] Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" forthcoming *American Economic Review*, available for download at www.ggdc.net/pwt. (DOI: 10.15141/S5NP4S, Retrieved June 8, 2015.)

(23) [1pt] If reliable data were available for more countries, how would that impact Table 1?

(A) The reported value for N would increase
(B) The number of rows in Table 1 would increase
(C) The reported values for the R-squared would increase
(D) The reported values for the R-squared would decrease

(24) [1pt] When data become available for 2010 - 2020, how would that impact Table 1?

(A) The reported value for N would increase
(B) The number of rows in Table 1 would increase
(C) The reported values for the R-squared would increase
(D) The reported values for the R-squared would decrease

(25) [2pts] Is the *first regression reported in Panel A* of Table 1 statistically significant overall?

(A) No, not at any conventional significance level
(B) Yes, but only at a 10% significance level
(C) Yes, at a 5% significance level but not a 1% significance level
(D) Yes, at a 1% significance level

(26) [2pts] To test if the *last regression reported in Panel A* of Table 1 is statistically significant overall, what is the value of the test statistic?

(A) 6.9
(B) 7.8
(C) 8.3
(D) 9.1
(E) 10.4

(27) [2pts] Which *could* be the value of the standard error of the slope coefficient for *first row results in Panel B*?

(A) -0.131
(B) -0.019
(C) 0.019
(D) 0.131

(28) [2pts] Which of these are a correct conclusion based on the *third regression in Panel B*?

(A) A country's growth rate in the 1970's is a terrible predictor of its growth rate in the 2000's
(B) Countries that had the strongest growth in the 1970's on average had the weakest growth in the 2000's
(C) The third regression, unlike the second regression in Panel B, shows regression to the mean because the regression coefficient is negative
(D) On average countries with growth rates that were 1 percentage point higher in the 1970's had growth rates that were 1.5 percentage points lower in the 2000's
(E) All of the above

▶ **Questions (29) – (34):** Recall predicting males' percent body fat ("Fitting Percentage of Body Fat to Simple Body Measurements," *Journal of Statistics Education,* 1996 (http://www.amstat.org/publications/jse/v4n1/datasets.johnson.html)). The variables are: percent body fat (pct_body_fat), height in cm (height_cm), abdominal circumference in cm (abdomen_cm), chest circumference in cm (chest_cm), neck circumference in cm (neck_cm), and age in years (age).

Table of descriptive statistics for these variables:

Variable	Obs	Mean	Std. Dev.	Min	Max
pct_body_fat	247	19.24777	8.105507	3.7	47.5
height_cm	247	178.6689	6.617156	162.56	197.485
abdomen_cm	247	92.4915	10.08156	70.4	126.2
chest_cm	247	100.8227	8.045327	83.4	128.3
neck_cm	247	37.98138	2.274223	31.1	43.9
age	247	44.98381	12.69369	22	81

Here is the correlation matrix for these variables:

	pct_bo~t	height~m	abdome~m	chest_cm	neck_cm	age
pct_body_fat	1.0000					
height_cm	-0.0642	1.0000				
abdomen_cm	0.8175	0.1667	1.0000			
chest_cm	0.6899	0.2072	0.9074	1.0000		
neck_cm	0.4725	0.3110	0.7209	0.7639	1.0000	
age	0.2873	-0.2596	0.2356	0.1744	0.1109	1.0000

Here is are the regression results once each variable has been *standardized*, which is why variables start with "s_":

Source	SS	df	MS		
Model	179.045703	5	35.8091407	Number of obs	= 247
Residual	66.9542981	241	.277818664	F(5, 241)	= 128.89
Total	246.000001	246	1.00000001	Prob > F	= 0.0000
				R-squared	= 0.7278
				Adj R-squared	= 0.7222
				Root MSE	= .52709

s_pct_body~t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
s_height_cm	-.1592175	.0372177	-4.28	0.000	-.2325309	-.085904
s_abdomen_cm	1.072114	.0819149	13.09	0.000	.9107531	1.233474
s_chest_cm	-.1474113	.0865263	-1.70	0.090	-.3178556	.023033
s_neck_cm	-.1421832	.0540621	-2.63	0.009	-.2486778	-.0356885
s_age	.0348432	.0364641	0.96	0.340	-.0369859	.1066723
_cons	-2.03e-09	.0335376	-0.00	1.000	-.0660643	.0660643

(29) [2pts] What is the value of s_height_cm for the tallest male in these data?

(A) 2.73
(B) 2.84
(C) 2.95
(D) 3.06
(E) 3.17

(30) [2pts] For the *standardized variables*, how would the table of descriptive statistics look different?

(A) It would be exactly the same
(B) The means would all be one
(C) The maximum values would all be larger
(D) The standard deviations would all be zero
(E) The minimum values would all be negative

(31) [2pts] For the *standardized variables*, how would the correlation matrix look different?

(A) It would be exactly the same
(B) All of the values would be 1.000
(C) All of the values would be either 1.000 or -1.000
(D) The signs of the correlations would be the same but numbers would be larger in absolute value
(E) The signs of the correlations would be the same but numbers would be smaller in absolute value

(32) [2pts] On average, males who are 1 s.d. taller have percent body fat that is about ___ s.d.'s lower.

(A) 0.06
(B) 0.16
(C) 0.26
(D) 0.36
(E) 0.46

(33) [2pts] On average, after controlling for height, chest circumference, neck circumference, and age, males with an abdominal circumference that is 10.1 cm higher have percent body fat that is about ___ higher.

(A) 0.8 percentage points
(B) 1.1 percentage points
(C) 8.1 percentage points
(D) 0.8 standard deviations
(E) 1.1 standard deviations

(34) [3pts] Which of these statements are TRUE?

(A) Males with a larger neck circumference tend to have lower percent body fat
(B) Males with a larger chest circumference tend to have lower percent body fat
(C) There is no statistically significant relationship between age and percent body fat
(D) There is no statistically significant relationship between height and percent body fat
(E) All of the above

▶ **Questions (35) – (36):** A random sample of movies are each categorized in one of three genres: comedy, action or drama. The dependent variable is gross revenues in millions of USD (gross). One of the x variables is the budget for making the film in millions of USD (budget). Consider these OLS results:

*gross-hat = -10.14 + 1.77*budget - 0.63*budget*comedy - 0.86*budget*drama + 32.12*comedy + 9.63*drama*

(35) [2pts] If you estimated the regression *gross-hat = b0 + b1*budget* using ONLY the observations for movies in the drama genre, what would the value of *b1* be?

(A) -0.91
(B) -0.86
(C) 0.86
(D) 0.91
(E) 1.77

(36) [2pts] For *gross-hat = b0 + b1*budget + b2*budget*comedy + b3*budget*action + b4*comedy + b5*action*, what would the value of *b2* be?

(A) -0.63
(B) -0.23
(C) 0.23
(D) 0.63
(E) 0.86

(37) [2pts] Consider testing $H_0: \beta_1 = 1$ versus $H_1: \beta_1 < 1$ in a simple regression. For a random sample with 30 observations, the OLS slope coefficient is 0.85 with a standard error of 0.11. What should you conclude? There is ___.

(A) sufficient evidence at a 1% significance level to infer the population slope is below 1
(B) sufficient evidence at a 5% (but not 1%) significance level to infer the population slope is below 1
(C) sufficient evidence at a 10% (but not 5%) significance level to infer the population slope is below 1
(D) some evidence that the population slope is below 1 but it does not meet any usual burden of proof
(E) no evidence that the population slope is below 1: our data directly contradict the research hypothesis

(38) [2pts] Consider testing $H_0: \beta_3 = -1$ versus $H_1: \beta_3 > -1$ in a multiple regression with three explanatory variables. For a random sample with 124 observations, the OLS slope coefficient on X_3 is -1.51 with a standard error of 0.65. What should you conclude? There is ___.

(A) sufficient evidence at a 1% significance level to infer the population slope on X_3 is above -1
(B) sufficient evidence at a 5% (but not 1%) significance level to infer the population slope on X_3 is above -1
(C) sufficient evidence at a 10% (but not 5%) significance level to infer the population slope on X_3 is above -1
(D) some evidence that the population slope on X_3 is above -1 but it does not meet any usual burden of proof
(E) no evidence that the population slope on X_3 is above -1: our data directly contradict the research hypothesis

REMEMBER, you must record your FORM CODE and answers on your BUBBLE FORM.