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

(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)**

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

(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? **(C)**

(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? **(B)**

(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? **(D)**

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

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

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

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

(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)**

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

(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? **(E)**

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

(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.) **(D)**

(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.) **(D)**

(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? **(B)**

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

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

(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)**

(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? **(D)**

(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? **(B)**

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

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

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

(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? **(C)**

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

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

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

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

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

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

(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. **(E)**

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

(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? **(D)**

(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? **(C)**

(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 ___. **(C)**

(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 ___. **(E)**