**NAME: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**PubH 6310 Summer 2021 Mid-Term Exam Part 1**

# General Instructions

This exam is take-home and open book. Part 1 is worth 10 points and Part 2 is worth 10 points. **You must work on the midterm independently without consultation with other students.** You may use your class notes, Stata or the internet for reference. Submit this file with your answers, named something like *yournamemidtermpart1docx* in the appropriate submission folder on Blackboard. Both parts are due **Wednesday June 30 at 5 PM.**

If you have any questions, please contact [lku@gwu.edu](mailto:lku@gwu.edu), not Josh. If a question needs to be clarified, the instructor may email all the students, so be on the lookout for emails.

# Integrity Assurance

Both parts of this midterm were done by me, without assistance or consultation from anyone other than the instructor of this course. My work will conform with the [GW Code of Academic Integrity](https://studentconduct.gwu.edu/sites/g/files/zaxdzs3451/f/downloads/160912%20Code%20of%20Academic%20Integrity.pdf). Problems identified may be subject to sanctions, including grade reductions, failure or more serious sanctions.

Type or sign your name above and date

# Part 1 (10 Points).

Q1. (1 pt) You have a variable called “marry” which is not labelled. You want to add these labels:

1 (married), 2 (single), 3 (divorced), 4 (widowed) Which choice below will label the values correctly? (Mark X next to one)

|  |
| --- |
| a. \_\_ label values marry 1= married 2=single 3=divorced 4=widowed |
| b. \_\_label define marlabel 1=married 2=single 3=divorced 4=widowed  label values marry marlabel |
| c. \_\_ label var marry 1=“married” 2=“single” 3=“divorced” 4= “widowed” |
| d. \_X\_label define marlabel 1 “married” 2 “single” 3 “divorced” 4 “widowed”  label values marry marlabel |

Q2. (1 pt) You want to run a logistic regression model seeing if there is a relationship between the probability that a woman currently smokes cigarettes (yes/no) with her age (in years), her education (in years) and her marital status (as described in Q1). How many binary dummy variables do you need to represent marital status in the model?

Number of dummy variables \_\_3\_\_

Q3. (1 pt) You want to investigate if the percent of a state’s working age (18 to 64) population that is uninsured in 2019 is related to whether the state expanded Medicaid eligibility for adults (to 138% of poverty or higher) by 2018. What kind of statistical test would be suitable for this? (Mark X next to one)

1. \_\_\_ chi-square test
2. \_\_\_ t-test
3. \_\_\_ one-way analysis of variance, with Scheffe
4. \_\_\_ logistic regression

Q4. (1 pt) You type the following Stata command and get the resulting error message:

**. mean povlev by raceethx**

variable by not found r(111);

Which of the following would fix the problem? (Mark X next to one)

|  |  |
| --- | --- |
| a. \_\_\_ mean povlev, over(raceethx) | c. \_\_X\_ mean povlev, by(raceethx) |
| b. \_\_\_ sum povlev, detail(raceethx) | d. \_\_\_ mean povlev over(raceethx) |

Q5. (1 pt) You have data about 900 nursing homes in 2020. You want to see if there is an association between the percent of nursing home residents who had deaths in 2020 related to COVID-19 (COVIDMORT) and the number of citations for inadequate infection control and prevention each nursing home had in 2019 (INFECTCIT, range from 0 to 5). Write Stata code that would let you test this relationship.

Q6. (1 pt) You now want to run the test from Q5 separately by for-profit/non-profit status of the facility (PROFIT, 1 = for profit, 0 – non-profit). How would you write the command(s)?

Q7. You are reading a survey of doctor’s practices that includes the following results for a regression model about the relationship of the percent of their diabetic patients who have controlled their disease (based on a hemoglobin A1c level below 7%) and whether the practice has attained Patient-Centered Medical Home (PCMH) recognition and the number of doctors in each practice?

Coeff SE 95% Conf Intvl

PCMH status 2.7 1.1

# doctors -1.3 (-1.7, 0.4)

1. (1 pt) Is PCMH status significant at p < .05? \_\_X\_ Yes \_\_\_ No
2. (1 pt) Is the number of doctors significant at p < .05? \_\_ Yes \_\_X\_\_ No

Q8 Review the table below.

. tab inscov18 region [aw=perwt18f], row col cell

HEALTH INSURANCE |

COVERAGE | CENSUS REGION AS OF 12/31/18

INDICATOR 2018 | Northeast Midwest South West | Total

-----------------+--------------------------------------------+----------

1. ANY PRIVATE | 2,218.929 2,713.365 4,578.842 2,929.754 | 12,440.89

| 17.84 21.81 36.80 23.55 | 100.00

| 77.00 78.68 72.98 72.80 | 74.81

| 13.34 16.32 27.54 17.62 | 74.81

-----------------+--------------------------------------------+----------

2. PUBLIC ONLY | 521.41198 489.61014 835.1543 744.01283 | 2,590.189

| 20.13 18.90 32.24 28.72 | 100.00

| 18.09 14.20 13.31 18.49 | 15.58

| 3.14 2.94 5.02 4.47 | 15.58

-----------------+--------------------------------------------+----------

3. UNINSURED | 141.44486 245.70608 860.00586 350.76416 | 1,597.921

| 8.85 15.38 53.82 21.95 | 100.00

| 4.91 7.12 13.71 8.72 | 9.61

| 0.85 1.48 5.17 2.11 | 9.61

-----------------+--------------------------------------------+----------

Total | 2,881.785 3,448.681 6,274.002 4,024.531 | 16,629

| 17.33 20.74 37.73 24.20 | 100.00

| 100.00 100.00 100.00 100.00 | 100.00

| 17.33 20.74 37.73 24.20 | 100.00

1. (1/2 pt) What percent of those in the South are uninsured? \_13.71%\_\_\_\_\_
2. (1/2 pt) What percent of privately insured people are in the West? \_\_23.55%\_\_\_\_

Q9. (1 pt) There are two variables about graduate students’ time use: **hrschool** = average number of hours spent on school each week and **hrjob** = average number of hours spent on their jobs each week. (These two variables do not overlap and there are no missing data.) You want to create a new variable **hschjob,** which:

= 1 if the number of hours for school and for the job are each less than or equal to 20 hours per week,

= 2 if school is more than 20 hours but job is 20 or less,

= 3 if school is 20 or less, but job is over 20, and

= 4 if both school and job are each more than 20 hours per week.

Write Stata code to create **hrschjob** in the space below.

* **gen hschjob=1 if hrschool <=20 & hrjob <=20**
* **replace hschjob=2 if hrschool >20 & hrjob <20**
* **replace hschjob=3 if hrschool <20 & hrjob >20**
* **replace hschjob=4 if hrschool**

Q10. There are no points associated with this, but which category do you fall into? We’re just curious.