**Sarah Pickersgill**

**Stata Assignment 2**

**Use the Stata dataset vipcls to answer Question 1 below. In Questions 2 and 3, you will create a new analysis dataset vipcls1. In Question 4, please use the newly created dataset.**

**Please turn in answers or output corresponding to those items that are bolded below:**

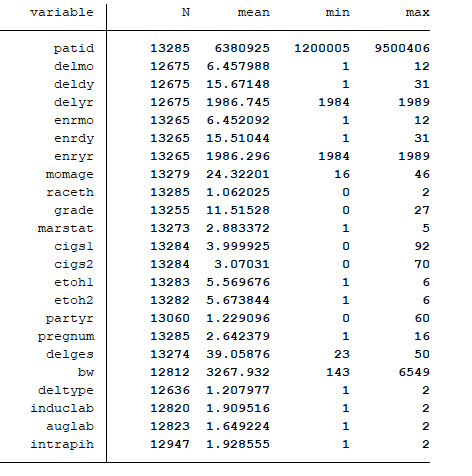
1. **Create a Word Document Table with the following information for each variable in the vipcls file:** number of missing values, specific values (and their frequencies) that seem implausible (i.e. those values that you will want to exclude from further analyses), and number of valid observations.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| variable | Patid | delmo | deldy | delyr | enrmo | enrdy | enryr | momage |
| Number of missing values | 0 | 610 | 610 | 610 | 20 | 20 | 20 | 4 |
| Implausible values  **Value(freq)** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 (2) \*study cut-off was 16 |
| Number of valid observations | 13,285 | 12,675 | 12,675 | 12,675 | 13,265 | 13,265 | 13,265 | 13,279 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| variable | raceth | grade | marstat | cigs1 | cigs2 | etoh1 | etoh2 | partyr |
| Number of missing values | 0 | 30 | 12 | 1 | 1 | 2 | 3 | 223 |
| Implausible values  **Value(freq)** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 200 (2) |
| Number of valid observations | 13,285 | 13,255 | 13,273 | 13,284 | 13,284 | 13,283 | 13,282 | 13,060 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| variable | pregnum | delges | bw | deltype | induclab | auglab | intrapih |
| Number of missing values | 0 | 0 | 473 | 649 | 465 | 462 | 338 |
| Implausible values  **Value (freq)** | 0 | 70 (1), 59(1), 57(1), 56(1), 53(1), 52(1), 51(5) | 0 | 0 | 0 | 0 | 0 |
| Number of valid observations | 13,285 | 13,274 | 12,812 | 12,636 | 12,820 | 12,823 | 12,947 |

1. Write a Stata do file called “analysis.do” which sets missing and implausible values as determined in Question 1 to the Stata missing value code (.). Add variable labels. Create a permanent Stata dataset, named vipcls1, incorporating these changes. **Use summarize to produce output regarding the N, MIN, MAX, and MEAN of all variables in this new dataset.**



1. Use the appropriate STATA commands (label define) to set up label definitions for the new Stata dataset vipcls1. Add these commands to the “analysis.do” do file and **copy this do file into your homework document to turn in.**

version 15

cd "C:\Users\Sarah\Epi510\Stata"

use vipcls

/// Labeling all variables ///

label variable patid "Patient ID"

label variable delmo "Delivery Month"

label variable deldy "Delivery Day"

label variable delyr "Delivery Year"

label variable enrmo "Enrollment Month"

label variable enrdy "Enrollment Day"

label variable enryr "Enrollment Year"

label variable momage "Mother's Age"

label variable raceth "Mother's Race/Ethnicity"

label variable grade "Mother's Education (years)"

label variable marstat "Mother's marital status"

label variable cigs1 "Mother's smoking, 1st trimester (cigs/days)"

label variable cigs2 "Mother's smoking, 2nd trimester (cigs/days)"

label variable etoh1 "Mother's alcohol intake, 1st trimester"

label variable etoh2 "Mother's alcohol intake, 2nd trimester"

label variable partyr "Number of sexual partners last year"

label variable pregnum "Number of pregnancies"

label variable delges "Gestational age at delivery"

label variable bw "Birth weight"

label variable deltype "Delivery Method"

label variable induclab "Induction of labor"

label variable auglab "Augmentation of labor"

label variable intrapih "Gestational hypertension"

///Label definitions///

label define l\_raceth 0 white 1 hispanic 2 black

label values raceth l\_raceth

label define l\_marstat 1 married 2 separated 3 divorced 4 widowed 5 "never married

label values marstat l\_marstat

label define l\_alcohol 1 "every day" 2 "3-5/week" 3 "one/week" 4 "<one/week" 5 "<one/month" 6 never

label values etoh1 l\_alcohol

label values etoh2 l\_alcohol

label define l\_deltype 1 vaginal 2 cesarean

label values deltype l\_deltype

label define l\_yesno 1 yes 2 no

label values induclab l\_yesno

label values auglab l\_yesno

label values intrapih l\_yesno

///Changing implausible values to -1 ///

recode momage 15 = -1

recode partyr 200 = -1

recode delges (51/max = -1)

///Missing Data: changing all -1 values to (.) ///

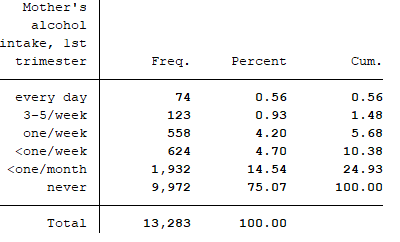
recode \_all (-1 = .)

///Summary table for all variables///

tabstat patid delmo deldy delyr enrmo enrdy enryr momage raceth gra marstat cigs1 cigs2 etoh1 etoh2 partyr pregnum delges bw deltype induclab auglab intrapih, stats(n mean min max) columns(statistics)

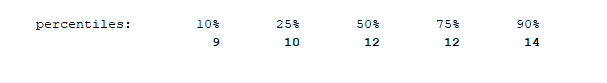
save vicpls1, replace

1. Use Stata commands summarize, codebook, tabstat, and tabulate as appropriate to determine the following:
   1. **What is the distribution of alcohol use in the 1st Trimester (create a table)? How many mothers drank daily in their 1st Trimester?**



74 mothers drank daily in their 1st Trimester

* 1. **What are the values of the interquartile range (25 and 75 percentiles) of mother’s education level?**

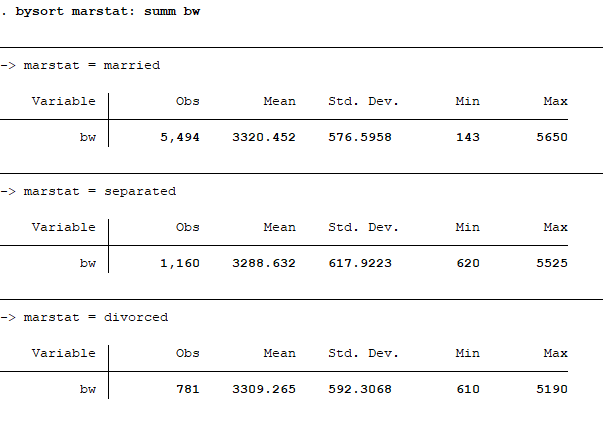


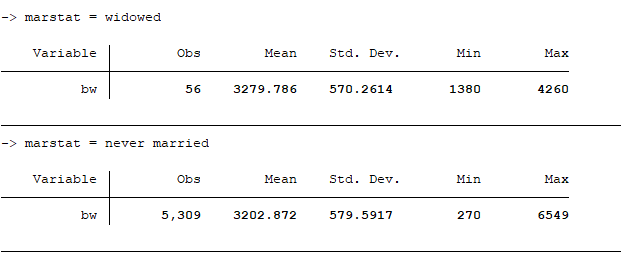
25th percentile = 10 years

75th percentile = 12 years

IQR = 2 years

* 1. **What are the mean, standard deviation, and sample sizes for the birth weight of babies in each of the five mother’s marital status groups?**





Be sure that your output is labeled and identified appropriately.