

## Biostatistics 203A – Fall 2023

### Mid-Quarter Project Submission Template

Student Name: Yuhui Wang  
Student ID Number: 606332401  
Data Subset Number: 50

#### **Methods:**

Three datasets: w2007, w2011, and famidsubset were used in the analysis. FAMID with subset number 50 were extracted from “famidsubset” and used to filter observations from w2007 and w2009 using merge command by “FAMID” in SAS Studio.

For table 1 (Demographic and Employment Characteristics in 2007), w2007 and w2011 are firstly merged by FAMID and those appear in both w2007 and w2011 are marked as “Yes” in variable completed\_2011, and those appear in w2007 but not in w2011 are marked as “No” in variable completed\_2011. This dataset is named w2007\_w2011. Afterwards, variables in the table are formatted based on explanations in YDS\_Formats; then, each characteristic (except job\_security and job\_satisfaction) is tabled by 2011 assessment completed status using “proc freq” (missing values are still included in table) based on dataset w2007\_w2011. For job\_security and job\_satisfaction, the used data is w2007\_w2011 with current job status = “Yes”.

For table 2 (Demographic Descriptive Statistics in 2007), firstly generate variable birth date formatted by time function using birth month and birth year. Then, use “01OCT2007” to minus the birth date by function “intck” and divided by 12 to get the age in 2007. After that, build descriptive statistics for Age using modified data and proc means command. After age, filter the data with no missing variable “E5HI17”; then, table the descriptive statistics. For mental\_health, when every variable is not missing, use do loop to reverse the certain items and add up all items to get the mental\_health\_score. Then table the descriptive statistics when no missing mental\_health\_score.

For table 3 (Financial Stress in 2007 and 2011), firstly merge w2007 and w2009 when FAMID both appear. Then based on this data, build descriptive statistics when there is no missing for variables both in 2007 and 2009.

For table 4 (Employment-Related Concerns about the Future in 2007 and 2011), table the frequency for each variable including missing values after formatting.

#### **Results:**

For my data, there are 535 respondents finishing both 2007 and 2011 assessment, and there are 119 respondents only finishing 2007 assessment.

For those respondents, people who completed 2011 assessment have more annual household income, and they tend to have a more stable life. For example, they are most likely employed, married, having

children. Education level is distributed like normal but skew left, meaning higher education level people are fewer. It turns out that people with good and stable life in 2007 are more likely to complete 2011 assessment. Also, people who completed 2011 assessment have more annual household income than those not.

For respondents who completed both 2007 and 2011 assessments, it turns out that all of them are more concerned about work or career plans in the future, including ability, overqualification, openings, relocation, illness, caring about relatives, and transportation. However, their financial stress is remaining the same.

*Some questions you may want to address:*

- Among the sample of respondents who completed the 2007 assessment, were there demographic or employment differences between those that did and did not go on to complete the 2011 assessment?

Answer: Yes, people who completed 2011 assessment have more annual household income, and they tend to have a more stable life. For example, they are most likely employed, married, having children.

- Among the subsample with assessments completed at both time points, does anything appear to have changed between 2007 and 2011?

Answer: Yes, people are more concerned about work or career plans in the future, including ability, overqualification, openings, relocation, illness, caring about relatives, and transportation.

### **Tables:**

**Table 1.** Demographic and employment frequencies for the sample in 2007 stratified by whether or not the respondent completed an assessment in 2011

<i>Demographic and Employment Characteristics in 2007</i>	Completed an assessment in 2011?			
	No (N = 119)		Yes (N = 535)	
	n	%	n	%
Gender	119	18.2	535	81.8
Female	61	9.33	313	47.86
Male	58	8.87	222	33.94
Missing	0	0	0	0
Highest Level of Education	119	18.2	535	81.8
High School or Less	37	5.66	104	15.9
Technical or Vocational	14	2.14	67	10.24
Some College	43	6.57	161	24.62
Bachelor's Degree	13	1.99	133	20.34
Graduate Degree	6	0.92	53	8.10
Missing	6	0.92	17	2.6

Currently Employed?	119	18.20	535	81.80
No	16	2.45	74	11.31
Yes	100	15.29	458	70.03
Missing	3	0.46	3	0.46
Currently Married or Cohabiting in an Intimate Relationship?	119	18.20	535	81.8
No	36	5.5	141	21.56
Yes	80	12.23	386	59.02
Missing	3	0.46	8	1.22
Any Children?	119	18.2	535	81.8
No	25	3.82	155	23.7
Yes	91	13.91	377	57.65
Missing	3	0.46	3	0.46
Job Security*	100	17.92	458	82.08
Not at all secure	5	0.9	17	3.05
Somewhat secure	22	3.94	106	19
Secure	33	5.91	168	30.11
Very Secure	40	7.17	162	29.03
Missing	0	0	5	0.9
Job Satisfaction*	100	17.92	458	82.08
Extremely or very dissatisfied	10	1.79	42	7.53
Somewhat dissatisfied	12	2.15	47	8.42
Somewhat satisfied	37	6.63	140	25.09
Extremely or very satisfied	40	7.17	226	40.5
Missing	1	0.18	3	0.54
*Frequencies and percentages only calculated among those respondents that indicated having a job in 2007				

**Table 2.** Demographic descriptive statistics for the sample in 2007 stratified by whether or not the respondent completed an assessment in 2011

<i>Demographic Descriptive Statistics in 2007</i>	Completed an assessment in 2011?					
	No			Yes		
	N	Mean	SD	N	Mean	SD
Age in Years	117	34.69	0.52	531	34.72	0.58
Annual Household Income	93	59780.29	38709.22	490	74305	78219.02
Mental Health Total Score	115	40.53	10.54	517	38.31	8.85

**Table 3.** Financial stress in 2007 and 2011 for the subsample with assessments completed at both time points

<i>Financial Stress</i>	2007	2011
-------------------------	------	------

	N	Mean	SD	N	Mean	SD
How much stress have you felt in meeting financial obligations?	535	4.27	1.83	535	4.31	1.87
How difficult is it for you to pay your bills on time?	533	3.12	1.91	533	3.22	1.98

**Table 4.** Employment-related concerns about the future in 2007 and 2011 for the subsample with assessments completed at both time points

<i>Employment-Related Concerns About the Future</i>	Assessment Time Point (N = 535)			
	2007		2011	
	n	%	n	%
<b>Are you concerned that any of the following might interfere with your work or career plans in the future?</b>				
Lack of ability to get training or degree	67	12.52	91	17.01
Lack of money to complete education or get started in my chosen career field	140	26.17	136	25.42
I am considered "overqualified"	20	3.74	46	8.6
Lack of openings in my field	86	16.07	156	29.16
Relocation is difficult or impossible	84	15.7	103	19.25
Illness, accident, or disability	64	11.96	108	20.19
Caring for a sick parent or relative	81	15.14	107	20
Transportation problems - difficulty in getting to or from work	34	6.36	48	8.97

Code Part:

```
libname mydata "~/nonshare";
```

```
data w2007;
```

```
  set mydata.w2007;
```

```
run;
```

```
data w2011;
```

```
  set mydata.w2011;
```

```
run;
```

```
data famid;
```

```
  set mydata.famidsubset;
```

```
  if subsetnumber=50;
```

```
run;
```

```
data w2007_subset;
```

```
  merge w2007(in=a) famid(in=b);
```

```
  by famid;
```

```
  if a and b;
```

```
run;
```

```
data w2011_subset;
  merge w2011(in=a) famid(in=b);
  by famid;
  if a and b;
run;
```

```
data w2007_w2011;
  merge w2007_subset(in=a) w2011_subset(in=b);
  by famid;
  completed_2011 = "No";
  if a then do;
    completed_2011 = "No";
    if b then completed_2011 = "Yes";
  output;
end;
run;
```

```
*table 1;
*gender;
proc format;
value gender 0 = "Female"
             1 = "Male"
             . = "Missing Values";
run;
```

```
proc freq data=w2007_w2011;
  format W1C542C gender.;
  tables W1C542C*completed_2011 / missing;
run;
```

```
*education;
proc format;
value highest_edu low-2 = "High School or Less"
                  3 = "Technical or Vocational"
                  4-5 = "Some College"
                  6 = "Bachelor's Degree"
                  7-high = "Graduate Degree"
                  . = "Missing Values";
run;
```

```
proc freq data=w2007_w2011;
  format C1ED17 highest_edu.;
  tables C1ED17*completed_2011 / missing;
run;
```

```

*employ;
proc format;
value employ 1 = "No"
              2 = "Yes"
              . = "Missing Values";

run;

proc freq data=w2007_w2011;
    format D617 employ.;
    tables D617*completed_2011 / missing;
run;

*marry;
proc format;
value marry 1 = "No"
             2 = "Yes"
             . = "Missing Values";

run;

proc freq data=w2007_w2011;
    format F1017 marry.;
    tables F1017*completed_2011 / missing;
run;

*children;
proc format;
value children 1 = "No"
                2 = "Yes"
                . = "Missing Values";

run;

proc freq data=w2007_w2011;
    format F517 children.;
    tables F517*completed_2011 / missing;
run;

data job;
set w2007_w2011;
if D617=2;
run;

*job_security;

```

```

proc format;
value job_sercurity 1 = "Not at all secure"
                    2 = "Somewhat secure"
                    3 = "Secure"
                    4 = "Very secure"
                    . = "Missing Values";

run;

proc freq data=job;
    format D1617 job_sercurity.;
    tables D1617*completed_2011 / missing;
run;

*job_satisfaction;
proc format;
value job_satisfaction 1-2 = "Extremely or very dissatisfied"
                       3 = "Somewhat dissatisfied"
                       4 = "Somewhat satisfied"
                       5-6 = "Extremely or very satisfied"
                       . = "Missing Values";

run;

proc freq data=job;
    format D1817 job_satisfaction.;
    tables D1817*completed_2011 / missing;
run;

*//////////////////////////;

*table 2;
*age;
data w2007_w2011_age;
    set w2007_w2011;
    age_at_2007 = .;
    if not missing(BIRTMO19) and not missing(BIRTYR19) then do;
        birth_date = mdy(BIRTMO19, 1, BIRTYR19);
        age_at_2007 = intck('month', birth_date, '01OCT2007'd) / 12;
        output;
    end;
    format age_at_2007 6.2;
run;

proc means data=w2007_w2011_age n nmiss mean std;
    class completed_2011;

```

```

var age_at_2007;
run;

*income;
data w2007_w2011_income;
  set w2007_w2011;
  if not missing(E5H17);
run;

proc means data=w2007_w2011_income nmiss n mean std;
  var E5H17;
  class completed_2011;
run;

*mental health;
data w2007_w2011_mh;
  set w2007_w2011;
  Mental_Health_Total_Score = .;
  if not missing(H13A17) and not missing(H13B17) and not missing(H13C17) and not missing(H13D17)
and not missing(H13E17) and not missing(H13F17) and not missing(H13G17) and not missing(H13H17)
and not missing(H13I17) and not missing(H13J17) and not missing(H13K17) and not missing(H13L17)
and not missing(H13M17) and not missing(H13N17) and not missing(H13O17) then do;
    H13A17_r = 6 - H13A17;
    H13D17_r = 6 - H13D17;
    H13F17_r = 6 - H13F17;
    H13I17_r = 6 - H13I17;
    H13N17_r = 6 - H13N17;
    H13O17_r = 6 - H13O17;

    Mental_Health_Total_Score = H13A17_r + H13B17 + H13C17 + H13D17_r + H13E17 + H13F17_r +
H13G17 + H13H17 + H13I17_r + H13J17 + H13K17 + H13L17 + H13M17 + H13N17_r + H13O17_r;
  output;
end;
  if not missing(Mental_Health_Total_Score);
run;

proc means data=w2007_w2011_mh n nmiss mean std;
  var Mental_Health_Total_Score;
  class completed_2011;
run;

*//////////////////////////;

*table 3;

```



```

data combine;
    merge w2007_subset(in=a) w2011_subset(in=b);
    by famid;
    if a and b;
run;

*stress;
data combine_stress;
    set combine;
    if not missing(E117);
    if not missing(E2STRE19);
run;

proc means data = combine_stress n nmiss mean std;
    var E117 E2STRE19;
    label E117 = "2007 Stress Level"
          E2STRE19 = "2011 Stress Level";
run;

*difficult for bills;
data combine_bill;
    set combine;
    if not missing(E217);
    if not missing(E3BILL19);
run;

proc means data = combine_bill n nmiss mean std;
    var E217 E3BILL19;
    label E217 = "2007 bill difficult Level"
          E3BILL19 = "2011 bill difficult Level";
run;

*//////////////////////////;
*table4;
proc format;
value yes 1="Yes"
          0="No";
run;

proc freq data=combine;
    format D3CNRA17 D4CNRA19 D3CNRB17 D4CNRB19 D3CNRC17 D4CNRD19 D3CNRE17
D4CNRE19 D3CNRF17 D4CNRF19 D3CNRD17 D4CNRI19 D3CNRI17 D4CNRJ19 D3CNRJ17 D4CNRK19 yes.;
    *ability;
    tables D3CNRA17 / missing;

```

```
label D3CNRA17 = "2007";
tables D4CNRA19 / missing;
label D4CNRA19 = "2011";
*money;
tables D3CNRB17 / missing;
label D3CNRB17 = "2007";
tables D4CNRB19 / missing;
label D4CNRB19 = "2011";
*overqualified;
tables D3CNRC17 / missing;
label D3CNRC17 = "2007";
tables D4CNRD19 / missing;
label D4CNRD19 = "2011";
*opening;
tables D3CNRE17 / missing;
label D3CNRE17 = "2007";
tables D4CNRE19 / missing;
label D4CNRE19 = "2011";
*relocation;
tables D3CNRF17 / missing;
label D3CNRF17 = "2007";
tables D4CNRF19 / missing;
label D4CNRF19 = "2011";
*illness;
tables D3CNRD17 / missing;
label D3CNRD17 = "2007";
tables D4CNRI19 / missing;
label D4CNRI19 = "2011";
*relative care;
tables D3CNRI17 / missing;
label D3CNRI17 = "2007";
tables D4CNRJ19 / missing;
label D4CNRJ19 = "2011";
*transportation;
tables D3CNRJ17 / missing;
label D3CNRJ17 = "2007";
tables D4CNRK19 / missing;
label D4CNRK19 = "2011";
```

```
run;
```