

Questions 1-3

learndis.xlsx.

Overview of Learning Disabilities in Children Study

__1. (20 points total)

- __a) By any means you like, import learndis.xlsx into R Studio or Stata. **Tip** - While in Excel, take care that the columns are saved in the appropriate formats (numeric or text).
- __b) Produce a description of your imported dataset
In Stata the command is **describe**
- __c) Save your imported data to a permanent Stata or R dataset.

```
. * Q1A: Import learn.xlsx
. import excel "/Users/cbigelow/Desktop/learndis.xlsx", sheet("Sheet1") firstrow

. * Q1B: Description of imported dataset
. describe
```

Contains data from /Users/cbigelow/Desktop/learndis.dta

```
obs:      105
vars:      6
size:     945
```

variable name	storage type	display format	value label	variable label
grade	byte	%10.0g		Grade Level
gender	byte	%10.0g	genderf	Gender
placemen	byte	%20.0g	placemenf	Type of Placement
readcomp	int	%10.0g		Reading Comprehension
mathcomp	int	%10.0g		Math Comprehension
iq	int	%10.0g		Intellectual Ability

```
. * Q1C: Save imported dataset
. save "/Users/cbigelow/Desktop/learndis.dta"
file /Users/cbigelow/Desktop/learndis.dta saved
```

__2. (10 points total)

__a) By any means you like, create a studyid variable that you name **studyid** and that has

__2. (20 points total)

__a) For all variables: Create variable labels

__b) For discrete variables ONLY: Create variable value labels.

__c) For all variables, as needed: Assign missing value codes.

Dear class: I will accept lots of answers here, because the excel file that you started with utilized blanks for missing values which were then assigned the missing value code “.” upon import. Regarding how to determine number of missing values, there are lots of ways to do this. I did a little sleuthing to see if there might be a simple command that would produce a tabulation of missing values. And I found one! So again, I will accept lots of answers here.

__d) Produce again a description of your imported dataset

In Stata the command is **describe**

```
. * Q2A: For all variables, create variable labels
. label variable grade "Grade Level"
. label variable gender "Gender"
. label variable placemen "Type of Placement"
. label variable readcomp "Reading Comprehension"
. label variable mathcomp "Math Comprehension"
. label variable iq "Intellectual Ability"

.* Q2B: For discrete variables, create variable value labels
. label define genderf 0 "male" 1 "female"
. label values genderf genderf
. label define placemenf 0 "Part-time" 1 "Full-time Segregated"
. label values placemenf placemenf
. label list
```

genderf:

0 male
1 female

placemenf:

0 Part-time
1 Full-time Segregated

```
* Q2C: For all variables, as needed: Assign missing value codes
. misstable summarize
```

				Obs<.		
Variable	Obs=.	Obs>.	Obs<.	Unique values	Min	Max
readcomp	29		76	35	22	107
mathcomp	11		94	43	61	121

__3. (20 points total)

- __a) Create a grouped variable that you name **quart_iq** that has values 1, 2, 3, and 4 according to quartile of value of **iq**.
- __b) Label your variable **quart_iq**
- __c) Attach variable value labels to the values 1, 2, 3, and 4 of **quart_iq**

```
. * Q3A: Create quart_iq that has values 1, 2, 3, or 4 according to value of iq
. . xtile quart_iq=iq, nq(4)
. fre quart_iq
```

quart_iq -- 4 quantiles of iq

		Freq.	Percent	Valid	Cum.
Valid	1	31	29.52	29.52	29.52
	2	22	20.95	20.95	50.48
	3	26	24.76	24.76	75.24
	4	26	24.76	24.76	100.00
	Total	105	100.00	100.00	

```
. * Q3A ANOTHER SOLUTION: Brute force
. tabstat iq, statistics(min q max)
```

variable	min	p25	p50	p75	max
iq	51	74	80	89	105

```
. generate quart_iq=iq
. recode quart_iq (min/74.0=1) (74.01/80.0=2) (80.1/89.0=3) (89.1/max=4)
(quart_iq: 105 changes made)
```

```
. fre quart_iq
```

quart_iq

		Freq.	Percent	Valid	Cum.
Valid	1	31	29.52	29.52	29.52
	2	22	20.95	20.95	50.48
	3	26	24.76	24.76	75.24
	4	26	24.76	24.76	100.00
	Total	105	100.00	100.00	

```
. * Q3B: Label your variable quart_iq
. label variable quart_iq "quart_iq: Quartile IQ"
```

```
. * Q3C: Attach variable labels to the values 1, 2, 3, and 4 of quart_iq
. label define quart_iqf 1 "Q1: 51-74" 2 "Q2: 74-80" 3 "Q3: 80-89" 4 "Q4: 89-105"
. label values quart_iq quart_iqf
. fre quart_iq
```

quart_iq -- 4 quantiles of iq

		Freq.	Percent	Valid	Cum.
Valid	1 Q1: 51-74	31	29.52	29.52	29.52
	2 Q2: 74-80	22	20.95	20.95	50.48
	3 Q3: 80-89	26	24.76	24.76	75.24
	4 Q4: 89-105	26	24.76	24.76	100.00
	Total	105	100.00	100.00	

Questions 4-5

gss1000.xlsx.

Overview of General Social Survey (GSS)

__4. (20 points total)

The variable **fepol** contains responses to the question “Female not suited for politics” and is coded 1=yes and 0=no. This is potentially confusing since a value of 1 is saying that the respondent believes females are not suited for politics.

__a) Create a more straightforward variable that you name **fepol_yes** and that is a reverse coding of **fepol**.

__b) Label this variable “Females suited for politics”

__c) Assign value labels of “Yes” to the value 1 and “No” to the value 0.

```
. * Q4: Preliminary - Look at distribution of fepol
. fre fepol
```

fepol -- Females not suited for politics (yes=1 no=0)? (recoded)

		Freq.	Percent	Valid	Cum.
Valid	0 0. No	496	49.60	78.86	78.86
	1 1. Yes	133	13.30	21.14	100.00
	Total	629	62.90	100.00	
Missing	.i	342	34.20		
	.n	29	2.90		
	Total	371	37.10		
Total		1000	100.00		

```
. Q4A: Create fepol_yes that is a reverse coding of fepol
. generate fepol_yes=fepol
(371 missing values generated)

. recode fepol_yes (0=1) (1=0)
(fepol_yes: 629 changes made)

. Q4B: Label this variable
. label variable fepol_yes "Females SUITED for politics"

. Q4C: Assign labels of "yes" to the value 1 and "no" to the value of 0
. label define fepolyesf 1 "Yes, suited" 0 "Not suited"
. label values fepol_yes fepolyesf

. * Check.
. tab2 fepol fepol_yes

-> tabulation of fepol by fepol_yes
```

Females not suited for politics (yes=1 no=0)? (recoded)	Females SUITED for politics		Total
	Not suite	Yes, suit	
0. No	0	496	496
1. Yes	133	0	133
Total	133	496	629

__5. (20 points total)

The variable **socbar** contains responses to the question "Spend evening at bar" and is coded 1=never, 2=once a year, 3=sev times a year, 4=once a month, 5=sev times a mnth, 6=sev times a week, 7=almost daily. There are missing values.

__5a) Create a new variable **socbar4** that is a grouping of the values of **socbar** as follows:

IF socbar =	THEN code socbar4 =	Assign socbar4 value label
missing	1	Unknown
1	2	Never
2 or 3 or 4	3	At most 1x/month
5 or 6 or 7	4	At least several x/month

__5b) Label your new variable.

__5c) Label your new variable values.

```
. * Q5: Preliminary - Look at distribution of socbar
. fre socbar
```

```
socbar -- spend evening at bar
```

		Freq.	Percent	Valid	Cum.
Valid	1 never	321	32.10	48.86	48.86
	2 once a year	97	9.70	14.76	63.62
	3 sev times a year	74	7.40	11.26	74.89
	4 once a month	67	6.70	10.20	85.08
	5 sev times a mnth	58	5.80	8.83	93.91
	6 sev times a week	34	3.40	5.18	99.09
	7 Almost daily	6	0.60	0.91	100.00
	Total	657	65.70	100.00	
Missing	.d Don't Know	1	0.10		
	.i Inapplicable	342	34.20		
	Total	343	34.30		
Total		1000	100.00		

```
. * Q5A: Create a new variable socbar4
```

```
. generate socbar4=socbar
```

```
(343 missing values generated)
```

```
. recode socbar4 (.i=1) (1=2) (2/4=3) (5/7=4)
```

```
(socbar4: 925 changes made)
```

```
. * Q5B: Label your new variable
```

```
. label variable socbar4 "Spend evening bar, grouped"
```

```
. * Q5C: Label your new variable values.
```

```
. label define socbar4f 1 "Inapplicable" 2 "Never" 3 "At most 1x/mo" 4 "At least several/mo"
```

```
. label values socbar4 socbar4f
```

```
. numlabel, add
```

```
. tab2 socbar socbar4, missing
```

```
-> tabulation of socbar by socbar4
```

spend evening at bar	1. Inappl	2. Never	3. At mos	4. At lea	.d	Total
1. never	0	321	0	0	0	321
2. once a year	0	0	97	0	0	97
3. sev times a year	0	0	74	0	0	74
4. once a month	0	0	67	0	0	67
5. sev times a mnth	0	0	0	58	0	58
6. sev times a week	0	0	0	34	0	34
7. Almost daily	0	0	0	6	0	6
.d. Don't Know	0	0	0	0	1	1
.i. Inapplicable	342	0	0	0	0	342
Total	342	321	238	98	1	1,000