

Topic: Excellence with Excel and Data
Curation!

 , **sas**, & More!



Welcome to Today's Lecture!

Today's Lecture

- **MS Excel:** Not just calculation! Data documentation!
- **Defaults** in Excel
- Making tabular data documentation to improve your analytic efficiency!

Codebook and Frequencies

SEQN - Respondent sequence number

Variable Name: SEQN
SAS Label: Respondent sequence number
English Text: Respondent sequence number.
Target: Both males and females 1 YEARS - 150 YEARS

OHDEXSTS - Overall Oral Health Exam Status

Variable Name: OHDEXSTS
SAS Label: Overall Oral Health Exam Status
English Text: Overall Oral Health Exam Status
Target: Both males and females 1 YEARS - 150 YEARS

Code or Value	Value Description	Count	Cumulative	Skip to Item
1	Complete	13271	13271	
2	Partial	7	13278	
3	Not Done	494	13772	
.	Missing	0	13772	

OHDDDESTS - Dentition Status Code

TABLE OF CONTENTS

- Component Description
- Eligible Sample
- Protocol and Procedure
- Quality Assurance & Quality Control
- Data Processing and Editing
- Analytic Notes
- Codebook
 - SEQN - Respondent sequence number
 - OHDEXSTS - Overall Oral Health Exam Status
 - OHDDDESTS - Dentition Status Code
 - OHXIMP - Dental Implant: yes / no?
 - OHX01TC - Tooth Count: #1
 - OHX02TC - Tooth Count: #2
 - OHX03TC - Tooth Count: #3
 - OHX04TC - Tooth Count: #4
 - OHX05TC - Tooth Count: #5
 - OHX06TC - Tooth Count: #6
 - OHX07TC - Tooth Count: #7
 - OHX08TC - Tooth Count: #8
 - OHX09TC - Tooth Count: #9
 - OHX10TC - Tooth Count: #10
 - OHX11TC - Tooth Count: #11



	A	B	C	D	E	F
	Order	Source Dataset	Native Variable Name	New Variable Name	Native Description	Values
1	1	All datasets	SEQN	SEQN	Respondent sequence number. Target: "both males and females 0 YEARS - 150 YEARS" ????	
2	2	P_DEMO	DMDYRUSZ	DMDYRUSZ	Length of time the participant has been in the US.	USYRS
3	3	R		DMDYRUSZ2	Recode of DMDYRUSZ	USYRS
4	4	P_OHXDEN	OHXIMP	OHXIMP	Do you have a tooth replaced with a surgical implant?	YNU
5	5	R		OHXIMP2	Recode of OHXIMP	YNU
6	6	R		IN_OHX	Flag for being in the OHXDEN dataset	
7	7	P_SMQ	SMQ020	SMQ020	These next questions are about cigarette smoking and other tobacco use. Have you smoked at least 100 cigarettes in your entire life?	YNU2
8	8	R		SMQ020_2	Recode of SMQ020	YNU2
9	9	R		IN_SMOK	Flag for being in the SMQ dataset	
10	10	P_OHQ	OHQ845	OHQ845	Overall, how would you rate the health of your teeth and gums?	RATE
11	11	R		OHQ845_2	Recode of OHQ845	RATE
12	12	R		IN_OHQS	Flag for being in the OHQ dataset	



Excel for Data Documentation

- Data documentation in public health often comes in PDF format, which **I do not like.**
 - Not easy to modify (add your own documentation)
 - Not easy to sort or rearrange the information
 - Hard to find what you're looking for
- By contrast, data documentation in MS Excel format is **awesome!**

Resources

Example making NHANES data dictionary using Excel:

<https://dethwench.com/nhanes-datasets-for-use-in-data-analysis/>

More extensive lesson on how to make an Excel data dictionary in Monika's LinkedIn Learning course, "Data Curation Foundations": <https://buff.ly/3310B14>

I also use these data dictionaries in my LinkedIn Learning courses on statistical software

R: <https://buff.ly/3qXTSP8>

SAS: <https://buff.ly/3Njfrky>



Why Data Documentation in Excel?

1. Many of the “business rules” we make when doing data analytics are based on variables or tables (e.g., listy things)
2. Tabular format forces you to be neat, make policy
3. Tables can easily be imported into a program (e.g., R, MS Word for a mail merge function, etc.)
4. If there is something complicated you have to explain that doesn't fit in tabular format, you can always do that in MS Word.
5. You can neatly add more information (more columns) to add more information in an organized manner.
6. You can just document what you want (only the variables you want, only the picklist levels you need, etc.)
7. Sorting, searching, and other functions are easy.
8. Easy to share with other teams who can then continue to modify for themselves.



Questions?

Facilitator: Monika M. Wahi, MPH, CPH

Email:

dethwench@gmail.com

Blog:

<https://dethwench.com/blog/>

LinkedIn:

<https://www.linkedin.com/in/dethwench/>

Follow our
company page
to make sure you don't
miss our
new and exciting
data science
educational events
and resources!

<https://www.linkedin.com/company/dethwench-professional-services/>