

**Assignment 3 - Data Analysis**

Author : Gayi Komi Selassi

ID : RA2422021010003

Program : MSc Epidemiology &amp; Biostatistics

Portfolio: [Click Here](#)Access the GitHub Repository: [Click Here](#)**1. Information about the final dataset :*****The CONTENTS Procedure***

<i>Data Set Name</i>	DATA_SP.DATA_MAIN_CLEAN	<i>Observations</i>	405
<i>Member Type</i>	DATA	<i>Variables</i>	30
<i>Engine</i>	V9	<i>Indexes</i>	0
<i>Created</i>	04/09/2025 11:19:49	<i>Observation Length</i>	656
<i>Last Modified</i>	04/09/2025 11:19:49	<i>Deleted Observations</i>	0
<i>Protection</i>		<i>Compressed</i>	NO
<i>Data Set Type</i>		<i>Sorted</i>	NO
<i>Label</i>			
<i>Data Representation</i>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<i>Encoding</i>	utf-8 Unicode (UTF-8)		

***Engine/Host Dependent Information***

<i>Data Set Page Size</i>	131072
<i>Number of Data Set Pages</i>	3
<i>First Data Page</i>	1
<i>Max Obs per Page</i>	199
<i>Obs in First Data Page</i>	190
<i>Number of Data Set Repairs</i>	0
<i>Filename</i>	/home/u64176007/sas_pratical_assignment/assignment_5_Summer_project/2. data_SP/data_main_clean.sas7bdat
<i>Release Created</i>	9.0401M8
<i>Host Created</i>	Linux
<i>Inode Number</i>	2353426605
<i>Access Permission</i>	rw-r--r--
<i>Owner Name</i>	u64176007
<i>File Size</i>	512KB
<i>File Size (bytes)</i>	524288

***Alphabetic List of Variables and Attributes***

#	Variable	Type	Len	Format	Informat
27	age	Num	8		
28	arm	Char	32		
2	bsl_assess_complete	Char	2	\$2.	\$2.
5	cgpart	Char	5	\$5.	\$5.
7	ecogps	Char	2	\$2.	\$2.
12	fact_bl	Char	7	\$7.	\$7.
16	fact_chg12	Char	8	\$8.	\$8.
19	fact_chg24	Char	8	\$8.	\$8.
18	hads_anx_chg12	Char	3	\$3.	\$3.

**Summer project Data Preparation: Final dataset informations****The CONTENTS Procedure**

<i>Alphabetic List of Variables and Attributes</i>					
#	Variable	Type	Len	Format	Informat
17	hads_dep_chg12	Char	3	\$3.	\$3.
13	hads_pt_anx_bl	Char	2	\$2.	\$2.
14	hads_pt_dep_bl	Char	2	\$2.	\$2.
15	n12	Char	2	\$2.	\$2.
25	n12_grp	Char	18	\$18.	\$18.
1	patref	Num	8	BEST12.	BEST32.
8	ptethncty	Char	24	\$24.	\$24.
9	pthighedu	Char	31	\$31.	\$31.
10	ptmargstat	Char	24	\$24.	\$24.
23	ptpqpt11_grp	Char	67	\$67.	\$67.
24	ptpqpt12_grp	Char	5	\$5.	\$5.
20	ptpqpt1_grp	Char	128	\$128.	\$128.
21	ptpqpt7_grp	Char	50	\$50.	\$50.
22	ptpqpt9e_grp	Char	28	\$28.	\$28.
11	ptrelgn	Char	14	\$14.	\$14.
3	race1	Char	48	\$48.	\$48.
4	sex	Char	3	\$3.	\$3.
26	sitetype	Char	11	\$11.	\$11.
6	tumrtyp	Char	44	\$44.	\$44.
29	wk12_dead	Char	32		
30	wk24_dead	Char	32		

**Summer project Data Preparation: Final dataset informations****2. Check frequency table of categorical variable and confirm recoding :*****The FREQ Procedure***

<i>race1</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>Asian</i>	15	3.70	15	3.70
<i>Black or African American</i>	46	11.36	61	15.06
<i>Other</i>	41	10.12	102	25.19
<i>White</i>	303	74.81	405	100.00

<i>ptrelgn</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>Catholic</i>	123	31.78	123	31.78
<i>Jewish</i>	20	5.17	143	36.95
<i>None</i>	28	7.24	171	44.19
<i>Other</i>	124	32.04	295	76.23
<i>Protestant</i>	92	23.77	387	100.00
<i>Frequency Missing = 18</i>				

<i>pthighedu</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>8th grade or less</i>	10	2.60	10	2.60
<i>9-11th grade</i>	27	7.01	37	9.61
<i>Advanced degree</i>	61	15.84	98	25.45
<i>Associate degree/some college</i>	64	16.62	162	42.08
<i>Bachelor's degree</i>	65	16.88	227	58.96
<i>High school graduate/GED</i>	117	30.39	344	89.35
<i>NA</i>	14	3.64	358	92.99
<i>Other</i>	5	1.30	363	94.29
<i>Vocational/technical school</i>	22	5.71	385	100.00
<i>Frequency Missing = 20</i>				

<i>ptmargstat</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>Divorced</i>	54	13.92	54	13.92
<i>Domestic partnership</i>	9	2.32	63	16.24
<i>Married</i>	227	58.51	290	74.74
<i>NA</i>	14	3.61	304	78.35
<i>Never married</i>	34	8.76	338	87.11
<i>Separated</i>	8	2.06	346	89.18
<i>Widowed</i>	42	10.82	388	100.00
<i>Frequency Missing = 17</i>				

<i>arm</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>1</i>	202	49.88	202	49.88
<i>2</i>	203	50.12	405	100.00