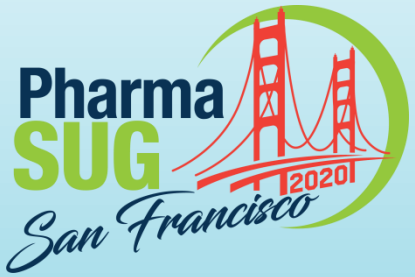
Demographic Table and Subgroup Summary Macro %TABLEN

Jeffrey Meyers, Mayo Clinic:

Statistical Programmer Analyst III within Mayo Clinic's Cancer Center Statistics team.

Research focus on gastrointestinal and respiratory cancers as well as meta-analysis.

10 years of SAS programming experience focusing on macros, graphics, SQL and reports.



Demographic Table and Subgroup Summary Macro %TABLEN

Jeffrey Meyers, Mayo Clinic

%TABLEN

- Creates a data summary table in a format commonly used for manuscript baseline characteristics summaries
- Summarizes five different variable types: continuous, discrete, date, survival and logistic regression/binomial
- Three different methods of grouping for comparisons
- Outputs the same style table to RTF, HTML, PDF, EXCEL, and PowerPoint
- Many options for flexibility
- Contains error checking, documentation, and cleans up after itself

%TABLEN Output

	Arm			
	A (N=49)	B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	
Randomization Date				0.2076[1]
N	49	55	104	
Median	05/12/2010	04/19/2010	04/27/2010	
Range	01/14/2010, 08/11/2010	06/23/2009, 08/11/2010	06/23/2009, 08/11/2010	
Gender, n (%)				0.6547[2]
Female	19 (38.8%)	19 (34.5%)	38 (36.5%)	
Male	30 (61.2%)	36 (65.5%)	66 (63.5%)	
Smoking Status, n (%)				0.4379[2]
Current	16 (32.7%)	15 (27.3%)	31 (29.8%)	
Former	9 (18.4%)	16 (29.1%)	25 (24.0%)	
Never	24 (49.0%)	24 (43.6%)	48 (46.2%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

Macro Required Parameters

- DATA: user input dataset to summarize
- VAR: space delimited list of variables to summarize
- TYPE: designates which type of each variable listed in VAR
  - 1: Continuous variable
  - 2: Discrete variable
  - 3: Date variable
  - 4: Survival variable
  - 5: Logistic regression/binomial variable
- The last used TYPE value carries forward to avoid unnecessary repeating of values
  - Example: VAR=age gender smoking_stat race
  - TYPE=1 2 2 2 is equivalent to TYPE=1 2 as the last 2 will carry forward
- Any combination of variable types is allowed

Macro Features

- Optional Grouping parameters:
  - BY: a variable to split distributions into multiple columns and allows p-value comparisons
  - COLBY (Column-BY): Nests the same set of columns specified by VAR and BY within each level of another variable in multiple columns
  - ROWBY(Row-BY): Same as COLBY but in rows instead of columns
  - WHERE: subsets the input DATA within the macro
- Optional Style parameters for fonts, borders, spacing, shading
- Can be wrapped in ODS tags or output to single file with OUTDOC

BY Example

```
%TABLEN(DATA=example,
    BY=arm,
    BYORDER=2 1,
    BYLABEL=Treatment Arm,
    SHOWTOTAL=0,
    VAR=age date_on sex smoke_st,
    TYPE=1 3 2);
```

- BYORDER: reorders the BY variable values
  - Default is unformatted values
  - Numbers correspond to original order (1=A, 2=B)
- BYLABEL: changes BY variable label
- SHOWTOTAL: turns off TOTAL column

	Treatment Arm B (N=55)	Treatment Arm A (N=49)	P-value
Age			0.4116[1]
N	55	49	
Mean (SD)	62.4 (8.73)	60.9 (11.34)	
Median	62.0	60.0	
Range	37.0, 81.0	37.0, 81.0	
Randomization Date			0.2076[1]
N	55	49	
Median	04/19/2010	05/12/2010	
Range	06/23/2009, 08/11/2010	01/14/2010, 08/11/2010	
Gender, n (%)			0.6547[2]
Female	19 (34.5%)	19 (38.8%)	
Male	36 (65.5%)	30 (61.2%)	
Smoking Status, n (%)			0.4379[2]
Current	15 (27.3%)	16 (32.7%)	
Former	16 (29.1%)	9 (18.4%)	
Never	24 (43.6%)	24 (49.0%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

COLBY Example

```
%TABLEN(DATA=example,
    COLBY=sex,
    COLBYORDER=2 1,
    COLBYLABEL=Gender, …);
```

- COLBYORDER: reorders the COLBY variable values
  - Default is unformatted values
  - Numbers correspond to original order (1=Female, 2=Male)
- COLBYLABEL: changes COLBY variable label
- BY example table is repeated for each gender

	Gender					
	Male			Female		
	Treatment Arm			Treatment Arm		
	B (N=36)	A (N=30)	P-value	B (N=19)	A (N=19)	P-value
Age			0.8064[1]			0.3348[1]
N	36	30		19	19	
Mean (SD)	61.4 (8.79)	60.7 (11.20)		64.4 (8.49)	61.2 (11.84)	
Median	61.0	60.0		65.0	61.0	
Range	37.0, 81.0	37.0, 80.0		45.0, 78.0	43.0, 81.0	
Randomization Date			0.7040[1]			0.0285[1]
N	36	30		19	19	
Median	04/25/2010	04/24/2010		04/02/2010	06/08/2010	
Range	07/01/2009, 08/10/2010	01/15/2010, 07/29/2010		06/23/2009, 08/11/2010	01/14/2010, 08/11/2010	
Smoking Status, n (%)			0.4803[2]			0.7415[2]
Current	10 (27.8%)	12 (40.0%)		5 (26.3%)	4 (21.1%)	
Former	14 (38.9%)	8 (26.7%)		2 (10.5%)	1 (5.3%)	
Never	12 (33.3%)	10 (33.3%)		12 (63.2%)	14 (73.7%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

ROWBY Example

```
%TABLEN(DATA=example,
    ROWBY=sex,
    ROWBYORDER=2 1,
    ROWBYLABEL=Gender, …);
```

- ROWBYORDER: reorders the ROWBY variable values
  - Default is unformatted values
  - Numbers correspond to original order (1=Female, 2=Male)
- ROWBYLABEL: changes ROWBY variable label
- BY example table is repeated for each gender
- Each value of ROWBY has its own header with counts

Gender		Treatment Arm B	Treatment Arm A	P-value
Male		(N=36)	(N=30)	
	Age			0.8064[1]
	N	36	30	
	Mean (SD)	61.4 (8.79)	60.7 (11.20)	
	Median	61.0	60.0	
	Range	37.0, 81.0	37.0, 80.0	
	Randomization Date			0.7040[1]
	N	36	30	
	Median	04/25/2010	04/24/2010	
	Range	07/01/2009, 08/10/2010	01/15/2010, 07/29/2010	
	Smoking Status, n (%)			0.4803[2]
	Current	10 (27.8%)	12 (40.0%)	
	Former	14 (38.9%)	8 (26.7%)	
	Never	12 (33.3%)	10 (33.3%)	
Female		(N=19)	(N=19)	
	Age			0.3348[1]
	N	19	19	
	Mean (SD)	64.4 (8.49)	61.2 (11.84)	
	Median	65.0	61.0	
	Range	45.0, 78.0	43.0, 81.0	
	Randomization Date			0.0285[1]
	N	19	19	
	Median	04/02/2010	06/08/2010	
	Range	06/23/2009, 08/11/2010	01/14/2010, 08/11/2010	
	Smoking Status, n (%)			0.7415[2]
	Current	5 (26.3%)	4 (21.1%)	
	Former	2 (10.5%)	1 (5.3%)	
	Never	12 (63.2%)	14 (73.7%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

Numeric and Date Type Example

```
%TABLEN(DATA=example,
    VAR=age date_on, TYPE=1 3,
CONTDISPLAY=n_nmiss mean_sd median_iqr,
DATEDISPLAY=n median range);
```

- CONTDISPLAY: controls which statistics are shown and the order for continuous variables
  - Keywords: N, NMISS, MEAN, SD, MEDIAN, IQR, RANGE along with combinations N_NMISS, MEAN_SD, MEDIAN_IQR, MEDIAN_RANGE
- DATEDISPLAY: same functionality and keywords as CONTDISPLAY but for date variables
- Significant digits can be set for each statistic separately for each variable
- Date format can be set with DATEFMT

	Total (N=104)
Age	
N (Missing)	104 (0)
Mean (SD)	61.7 (10.02)
Median (IQR)	61.0 (55.5, 69.0)
Randomization Date	
N	104
Median	04/27/2010
Range	06/23/2009, 08/11/2010

Discrete Type Example

```
%TABLEN(DATA=example,
    VAR=sex smoke_st, TYPE=2,
    DIS_DISPLAY=N,
    DIS_ORDER=2 1|FREQD);
```

- DIS_DISPLAY: Determines if counts, percentages or both are shown
- DIS_ORDER: Allows the user to change the order of each discrete variable.
  - Default order is unformatted values.
  - Options are to use a numbered list corresponding to the default order, or the keywords FREQA or FREQD (frequency ascending/descending)
- Percentage digits can be controlled
- Missing values can be included as valid values

	Total (N=104)
Gender, n	
Male	66
Female	38
Smoking Status, n	
Never	48
Current	31
Former	25

Survival Type Example

```
%TABLEN(DATA=example,
    VAR=fu_time pg_time, TYPE=4,
    SURV_STAT=fu_stat pg_stat,
    CEN_VL=1, TDIVISOR=30.44 30.44,
    TIMELIST=6|3, TIME_UNITS=months);
```

- SURV_STAT: assigns event status variable for the time variables included in VAR
- CEN_VL: designates the censor value
- TDIVISOR: transforms time values
- TIMELIST: specifies times for event-free rates
- TIME_UNITS: adds a unit to time-point estimates

	Total (N=104)
Followup#Time	
Events/N	95/104
Median (95% CI)	5.91 (4.24 - 8.44)
6 months Est (95% CI)	0.49 (0.40 - 0.60)
Time#to#Progression	
Events/N	86/104
Median (95% CI)	3.48 (2.20 - 3.75)
3 months Est (95% CI)	0.54 (0.45 - 0.65)

Survival Type Example 2

```
%TABLEN(DATA=example, …,
    BY=arm,
    REFERENCE=B,
    SURVDISPLAY=events_n median hr);
```

- Using a BY variable enables p-values and hazard ratios
- REFERENCE: designates reference value for hazard ratios
- SURVDISPLAY: determines shown statistics and their order
  - Keywords: EVENTS, N, EVENTS_N, MEDIAN, HR, TIMELIST, COXPVAL

	Arm			
	A (N=49)	B (N=55)	Total (N=104)	P-value
Followup#Time				0.3613[1]
Events/N	44/49	51/55	95/104	
Median (95% CI)	4.24 (3.61 - 8.15)	8.28 (4.50 - 10.48)	5.91 (4.24 - 8.44)	
Hazard Ratio (95% CI)	1.21 (0.80 - 1.81)	Reference		
Time#to#Progression				0.2582[1]
Events/N	42/49	44/55	86/104	
Median (95% CI)	2.00 (1.84 - 4.50)	3.68 (3.52 - 5.32)	3.48 (2.20 - 3.75)	
Hazard Ratio (95% CI)	1.28 (0.83 - 1.97)	Reference		

[1]Logrank p-value;

Logistic Type Example

```
%TABLEN(DATA=example,
    VAR=gd3, TYPE=5,
    LOG_EVENT=1,
    LOG_DISPLAY=events_n binrate oddsratio);
```

- LOG_EVENT: determines the event for the binomial variable listed in VAR list
- LOG_DISPLAY: determines the statistics shown for logistic regression variables
  - Keywords: EVENTS, N, EVENTS_N, BINRATE, ODDSRATIO, and BINRATE (binomial success rate)
- REFERENCE: determines reference group for odds ratios

	Arm			
	A (N=49)	B (N=55)	Total (N=104)	P-value
Grade 3+ AE (Event=1)				0.4218[1]
Events/N	41/49	49/55	90/104	
Event Rate (95% CI)	0.84 (0.73 - 0.94)	0.89 (0.81 - 0.97)	0.87 (0.80 - 0.93)	
Odds Ratio (95% CI)	0.63 (0.20 - 1.96)	Reference		

[1]Type-3 Wald p-value;

P-values

- Multiple p-values available for each variable type
- Are coded as numeric values that are different for each variable type. 1 is default for all:
  - Numeric/Date: 1=Kruskal-Wallis. Other p-values include Wilcoxon, ANOVA F, and T-tests
  - Discrete: 1=Chi-square. Other p-values include Fisher's exact and Cochran-Armitage
  - Survival: 1=Logrank. Other p-values include Wilcoxon, type-3 Score/Wald/Likelihood-ratio
  - Logistic: 1=type-3 Wald. Other p-values include Chi-square and Fisher's exact
- Last value listed in PVALS is carried forward for rest of VAR list
- Value of 0 indicates no p-value to be shown
- Significant digits can be changed

Style Modifiers (Borders)

BORDERDISPLAY=1

	Arm A (N=49)	Arm B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	
Gender, n (%)				0.6547[2]
Female	19 (38.8%)	19 (34.5%)	38 (36.5%)	
Male	30 (61.2%)	36 (65.5%)	66 (63.5%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

BORDERDISPLAY=2

	Arm A (N=49)	Arm B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	
Gender, n (%)				0.6547[2]
Female	19 (38.8%)	19 (34.5%)	38 (36.5%)	
Male	30 (61.2%)	36 (65.5%)	66 (63.5%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

BORDERDISPLAY=3

	Arm A (N=49)	Arm B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	
Gender, n (%)				0.6547[2]
Female	19 (38.8%)	19 (34.5%)	38 (36.5%)	
Male	30 (61.2%)	36 (65.5%)	66 (63.5%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

BORDERDISPLAY=4

	Arm A (N=49)	Arm B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	
Gender, n (%)				0.6547[2]
Female	19 (38.8%)	19 (34.5%)	38 (36.5%)	
Male	30 (61.2%)	36 (65.5%)	66 (63.5%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

Style Modifiers (Spacing)

SPLIT=space

	Arm A (N=49)	Arm B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N (Missing)	49 (0)	55 (0)	104 (0)	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median (Range)	60.0 (37.0, 81.0)	62.0 (37.0, 81.0)	61.0 (37.0, 81.0)	
Randomization Date				0.0409[2]
N	49	55	104	
Median	05/12/2010	04/19/2010	04/27/2010	
Range	01/14/2010, 08/11/2010	06/23/2009, 08/11/2010	06/23/2009, 08/11/2010	

[1]Kruskal-Wallis p-value; [2]ANOVA F-test p-value;

SPLIT=line

	Arm A (N=49)	Arm B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	
Gender, n (%)				0.6547[2]
Female	19 (38.8%)	19 (34.5%)	38 (36.5%)	
Male	30 (61.2%)	36 (65.5%)	66 (63.5%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

SPLIT=none

	Arm A (N=49)	Arm B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	
Gender, n (%)				0.6547[2]
Female	19 (38.8%)	19 (34.5%)	38 (36.5%)	
Male	30 (61.2%)	36 (65.5%)	66 (63.5%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

Style Modifiers (Shading)

SHADING=0

	Arm A (N=49)	Arm B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	
Gender, n (%)				0.6547[2]
Female	19 (38.8%)	19 (34.5%)	38 (36.5%)	
Male	30 (61.2%)	36 (65.5%)	66 (63.5%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

SHADING=1

	Arm A (N=49)	Arm B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N (Missing)	49 (0)	55 (0)	104 (0)	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median (Range)	60.0 (37.0, 81.0)	62.0 (37.0, 81.0)	61.0 (37.0, 81.0)	
Randomization Date				0.0409[2]
N	49	55	104	
Median	05/12/2010	04/19/2010	04/27/2010	
Range	01/14/2010, 08/11/2010	06/23/2009, 08/11/2010	06/23/2009, 08/11/2010	

[1]Kruskal-Wallis p-value; [2]ANOVA F-test p-value;

SHADING=2

	Arm A (N=49)	Arm B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	
Gender, n (%)				0.6547[2]
Female	19 (38.8%)	19 (34.5%)	38 (36.5%)	
Male	30 (61.2%)	36 (65.5%)	66 (63.5%)	

[1]Kruskal-Wallis p-value; [2]Chi-Square p-value;

Outputting to Different Destinations

RTF

	Arm			
	A (N=49)	B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	

[1]Kruskal-Wallis p-value;

PDF

	Arm			
	A (N=49)	B (N=55)	Total (N=104)	P-value
Age				0.4116 [1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	

[1]Kruskal-Wallis p-value;

EXCEL

	A	B	C	D	E
1					
2		Arm			
3		A (N=49)	B (N=55)	Total (N=104)	P-value
4	**Age**				0.4116[1]
5	N	49	55	104	
6	Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
7	Median	60	62	61	
8	Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	
9	[1]Kruskal-Wallis p-value;				

SAS Studio

	Arm			
	A (N=49)	B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	

[1]Kruskal-Wallis p-value;

PowerPoint

	Arm			
	A (N=49)	B (N=55)	Total (N=104)	P-value
Age				0.4116[1]
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	

[1]Kruskal-Wallis p-value;

Listing

	Arm			
	A (N=49)	B (N=55)	Total (N=104)	P-value
Age				0.4116*
N	49	55	104	
Mean (SD)	60.9 (11.34)	62.4 (8.73)	61.7 (10.02)	
Median	60.0	62.0	61.0	
Range	37.0, 81.0	37.0, 81.0	37.0, 81.0	

*Kruskal-Wallis p-value

Conclusion

- The TABLEN macro is a powerful tool for producing demographic summary tables in publication quality format
- Many options for modifying the table style to meet the user's needs
- Flexibility and versatility in outputting to different ODS destinations
- The macro is available for download on the SAS Communities page

Name: Jeffrey Meyers
Organization: Mayo Clinic
E-mail: meyers.jeffrey@mayo.edu / jpmeyers.spa@gmail.com