Sensitivity of Variable Type in Statistical Analysis using Selective SAS/STAT® Procedures

Sakthivel Sivam
Jairaj Ramaiah
Quartesian Clinical Research

Quartesian
The Data Company

Agenda

- Introduction
- Variable data types
- Comparative Analyses
- Results & Conclusion

Introduction

- Variable data type is one of the key component for Statistical procedures in SAS.
- Is there any impact of variable data type in SAS/STAT procedures?
- What happens if we use character variable instead of numeric?
- When can we use character & numeric variables interchangeably and when not to?

Variable types

- Data Types
  - Character is a sequence of characters
  - Numeric contains only numbers, and sometimes a decimal point, a minus sign, or both.
- Variable Types
  - Quantitative
    - Discrete
    - Continuous
  - Qualitative
    - Nominal
    - Ordinal

Variable	Value
Numeric	Numbers
Character	Numbers
Character	Text

Quartesian
The Data Company

Comparative Analyses

Proc Freq

	Treatment	trtn	Gender	gendern	Response	respn	Count
1	Active	1	female	2	Better	1	16
2	Active	1	female	2	Same	0	11
3	Placebo	2	female	2	Better	1	5
4	Placebo	2	female	2	Same	0	20
5	Active	1	male	1	Better	1	12
6	Active	1	male	1	Same	0	16
7	Placebo	2	male	1	Better	1	7
8	Placebo	2	male	1	Same	0	19

Comparative Analyses
Proc Freq

```
proc freq data=Migraine;
   tables Gender*Treatment*Response
          / cmh;
   weight Count;
run;
```

```
proc freq data=Migraine;
   tables Gendern*trtn*respn / cmh;
   weight Count;
run;
```

Quartesian
The Data Company

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	3.3132	1.4456	7.5934
	Logit	3.2941	1.4182	7.6515
Cohort (Col1 Risk)	Mantel-Haenszel	2.1636	1.2336	3.7948
	Logit	2.1059	1.1951	3.7108
Cohort (Col2 Risk)	Mantel-Haenszel	0.6420	0.4705	0.8761
	Logit	0.6613	0.4852	0.9013

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	0.3018	0.1317	0.6917
	Logit	0.3036	0.1307	0.7051
Cohort (Col1 Risk)	Mantel-Haenszel	0.6420	0.4705	0.8761
	Logit	0.6613	0.4852	0.9013
Cohort (Col2 Risk)	Mantel-Haenszel	2.1636	1.2336	3.7948
	Logit	2.1059	1.1951	3.7108

Comparative Analyses - Mixed Effect Model

	PATID	trtn	atpt	atptn	CONC	DQTCF
1	A1091001 10011001	2	0	0	0	-11.38
2	A1091001 10011001	2	0	0	0	-13.05
3	A1091001 10011001	2	0	0	0	0.62
4	A1091001 10011001	2	0	0	0	0.62
5	A1091001 10011001	2	0	0	0	-13.71
6	A1091001 10011001	2	0	0	0	-6.38
7	A1091001 10011001	2	0.5	0.5	0	-4.38
8	A1091001 10011001	2	0.5	0.5	0	-6.71
9	A1091001 10011001	2	1	1	0	0.62
10	A1091001 10011001	2	1	1	0	-4.38
11	A1091001 10011001	2	2	2	0	3.62
12	A1091001 10011001	2	2	2	0	-0.05
13	A1091001 10011001	2	3	3	0	1.95
14	A1091001 10011001	2	3	3	0	-5.38

Comparative Analyses - Mixed Effect Model

```
PROC MIXED DATA=model_data1;
ods select SolutionF estimates;
ods output estimates = estimate SolutionF=SF;

 class atpt trtn patid;
 MODEL dqtcf = conc trtn  atpt/ solution cl alpha=0.1;
 RANDOM intercept conc/ subject=patid type=cs ;

run;
```

```
PROC MIXED DATA=model_data1 ;
ods select SolutionF estimates;
ods output estimates = estimate SolutionF=SF;

 class atptn trtn patid;
 MODEL dqtcf = conc trtn atptn/ solution cl alpha=0.1;
 RANDOM intercept conc/ subject=patid type=cs ;

run;
```

Quartesian
The Data Company

Solution of fixed effect estimates - atpt

Effect	atpt	trtn	Estimate
Intercept			-4.8703
CONC			6.296E-6
trtn		1	1.7426
trtn		2	0
atpt	0		-0.3156
atpt	0.5		3.1545
atpt	1		5.1331
atpt	1.5		1.1809
atpt	12		0.06143
atpt	2		4.4085
atpt	2.5		6.3607
atpt	24		-1.5721
atpt	3		3.0117
atpt	3.5		4.7301
atpt	4		3.0197
atpt	5		1.0664
atpt	6		-1.3416
atpt	7		-4.3657
atpt	8		-1.5731
atpt	9		0

Solution of fixed effect estimates - atptn

Effect	atptn	trtn	Estimate
Intercept			-6.4423
CONC			6.296E-6
trtn		1	1.7426
trtn		2	0
atptn	0		1.2565
atptn	0.5		4.7266
atptn	1		6.7052
atptn	1.5		2.7530
atptn	2		5.9806
atptn	2.5		7.9328
atptn	3		4.5837
atptn	3.5		6.3022
atptn	4		4.5917
atptn	5		2.6384
atptn	6		0.2305
atptn	7		-2.7937
atptn	8		-0.00106
atptn	9		1.5721
atptn	12		1.6335
atptn	24		0

Quartesian
The Data Company

Comparative Analyses
PROC GLM

	patno	trt	trtn	type	typen	hgbch
1	1	ACT	1	CERVICAL	1	1.7
2	3	ACT	1	CERVICAL	1	-0.2
3	6	ACT	1	CERVICAL	1	1.7
4	7	ACT	1	CERVICAL	1	2.3
5	10	ACT	1	CERVICAL	1	2.7
6	12	ACT	1	CERVICAL	1	0.4
7	13	ACT	1	CERVICAL	1	1.3
8	15	ACT	1	CERVICAL	1	0.6
9	22	ACT	1	PROSTATE	2	2.7
10	24	ACT	1	PROSTATE	2	1.6
11	26	ACT	1	PROSTATE	2	2.5
12	28	ACT	1	PROSTATE	2	0.5
13	29	ACT	1	PROSTATE	2	2.6
14	31	ACT	1	PROSTATE	2	3.7
15	34	ACT	1	PROSTATE	2	2.7
16	36	ACT	1	PROSTATE	2	1.3
17	42	ACT	1	COLORECTAL	3	-0.3
18	45	ACT	1	COLORECTAL	3	1.9
19	46	ACT	1	COLORECTAL	3	1.7
20	47	ACT	1	COLORECTAL	3	0.5
21	49	ACT	1	COLORECTAL	3	2.1
22	51	ACT	1	COLORECTAL	3	-0.4
23	52	ACT	1	COLORECTAL	3	0.1

Comparative Analyses
PROC GLM – ANOVA & LS Means

```
proc glm data = model_data2;
class trt type;
model hgbch = trt type trt*type / ss3;
lsmeans type / pdiff stderr t lines;
run;
quit;
```

```
proc glm data = model_data2;
class trtn typen;
model hgbch = trtn typen trtn*typen /ss3;
lsmeans typen / pdiff stderr t lines;
run;
quit;
```

Dependent Variable: hgbch

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	15.29604167	3.05920833	2.38	0.0543
Error	42	53.88375000	1.28294643		
Corrected Total	47	69.17979167			

R-Square	Coeff Var	Root MSE	hgbch Mean
0.221106	101.6229	1.132672	1.114583

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	1	5.26687500	5.26687500	4.11	0.0491
type	2	9.11291667	4.55645833	3.55	0.0376
trt*type	2	0.91625000	0.45812500	0.36	0.7018

Dependent Variable: hgbch

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	15.29604167	3.05920833	2.38	0.0543
Error	42	53.88375000	1.28294643		
Corrected Total	47	69.17979167			

R-Square	Coeff Var	Root MSE	hgbch Mean
0.221106	101.6229	1.132672	1.114583

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trtn	1	5.26687500	5.26687500	4.11	0.0491
typen	2	9.11291667	4.55645833	3.55	0.0376
trtn*typen	2	0.91625000	0.45812500	0.36	0.7018

The GLM Procedure
Least Squares Means

type	hgbch LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
CERVICAL	1.05625000	0.28316806	0.0006	1
COLORECTAL	0.61250000	0.28316806	0.0363	2
PROSTATE	1.67500000	0.28316806	<.0001	3

Least Squares Means for Effect type
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: hgbch

i/j	1	2	3
1		1.1081 0.2741	-1.5451 0.1298
2	-1.1081 0.2741		-2.6532 0.0112
3	1.545098 0.1298	2.653198 0.0112	

The GLM Procedure
Least Squares Means

typen	hgbch LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
1	1.05625000	0.28316806	0.0006	1
2	1.67500000	0.28316806	<.0001	2
3	0.61250000	0.28316806	0.0363	3

Least Squares Means for Effect typen
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: hgbch

i/j	1	2	3
1		-1.5451 0.1298	1.1081 0.2741
2	1.545098 0.1298		2.653198 0.0112
3	-1.1081 0.2741	-2.6532 0.0112	

Quartesian
The Data Company

Comparative Analyses

Logistic Regression

	pat	x	group	groupn	relapse	relapsen
1	1	3	ACT	1	NO	0
2	2	3	ACT	1	YES	1
3	4	3	ACT	1	YES	1
4	6	6	ACT	1	YES	1
5	7	15	ACT	1	NO	0
6	10	6	ACT	1	YES	1
7	11	6	ACT	1	YES	1
8	14	6	ACT	1	YES	1
9	15	15	ACT	1	NO	0
10	17	15	ACT	1	NO	0
11	20	12	ACT	1	NO	0
12	21	18	ACT	1	NO	0
13	22	6	ACT	1	YES	1
14	25	15	ACT	1	NO	0
15	26	6	ACT	1	YES	1
16	28	15	ACT	1	NO	0
17	29	12	ACT	1	YES	1
18	32	9	ACT	1	NO	0

Quartesian
The Data Company

Comparative Analyses
Logistic Regression

```
proc logistic data=model_data3;
class group/ param=ref;
model relapse(event='YES')= group x;
oddsratio group;
oddsratio x;
run;
```

```
proc logistic data=model_data3;
class groupn / param=ref;
model relapsen(eventn='1')= groupn x;
oddsratio groupn;
oddsratio x;
run;
```

Comparative Analyses

Logistic Regression

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	2.6135	0.7149	13.3662	0.0003
group	ACT	1	-1.1191	0.4669	5.7446	0.0165
x		1	-0.1998	0.0560	12.7187	0.0004

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	1.4944	0.5569	7.2007	0.0073
groupn	1	1	1.1191	0.4669	5.7446	0.0165
x		1	-0.1998	0.0560	12.7187	0.0004

Odds Ratio Estimates and Wald Confidence Intervals

Label	Estimate	95% Confidence Limits	
group ACT vs PBO	0.327	0.131	0.815
x	0.819	0.734	0.914

Odds Ratio Estimates and Wald Confidence Intervals

Label	Estimate	95% Confidence Limits	
groupn 1 vs 2	3.062	1.226	7.646
x	0.819	0.734	0.914

Quartesian
The Data Company

Comparative Analyses
Proc PHREG

	LogWBC	Platelet	plateletn	'Survival Time'	'0=Alive 1=Dead'
1	3.6628	1	YES	1.25	1
2	3.9868	1	YES	1.25	1
3	3.8751	1	YES	2	1
4	3.8062	0	NO	2	1
5	3.7243	0	NO	2	1
6	4.4757	1	YES	3	1
7	4.9542	1	YES	5	1
8	3.7324	1	YES	5	1
9	3.5441	1	YES	6	1
10	3.5441	0	NO	6	1
11	3.5185	1	YES	6	1
12	3.9294	1	YES	6	1
13	3.3617	1	YES	7	1
14	3.7324	0	NO	7	1
15	3.7243	1	YES	7	1
16	3.7993	1	YES	9	1
17	3.8808	1	YES	11	1
18	3.7709	1	YES	11	1
19	3.7993	1	YES	11	1
20	3.8865	1	YES	11	1

Comparative Analyses
Proc PHREG

```
proc phreg data=Myeloma;
class Platelet LogWBC;
model Time*VStatus(0)=  Platelet  LogWBC;
run;
```

```
proc phreg data=Myeloma;
class plateletn LogWBC ;
model Time*VStatus(0)=  Plateletn LogWBC;
run;
```

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
Platelet	1	4.5912	0.0321
LogWBC	44	533.2162	<.0001

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
plateletn	1	4.5912	0.0321
LogWBC	44	533.2162	<.0001

Comparative Analyses
Proc Lifetest

	subj	time	censor	stratan	strata
1	1	6	1	1	White
2	2	44	1	2	Black
3	4	21	0	3	Brown
4	4	14	1	1	White
5	5	62	1	2	Black

Quartesian
The Data Company

Comparative Analyses
Proc Lifetest

```
proc lifetest data = model_data5;
time time*censor(0);
strata stratan;
run;
```

```
proc lifetest data = model_data5;
time time*censor(0);
strata strata;
run;
```

The LIFETEST Procedure

Stratum 1: stratan = 1

Product-Limit Survival Estimates

time	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	2
6.0000	0.5000	0.5000	0.3536	1	1
14.0000	0	1.0000		2	0

Summary Statistics for Time Variable time

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval Transform	[Lower	Upper)
75	14.0000	LOGLOG	6.0000	14.0000
50	10.0000	LOGLOG	6.0000	14.0000
25	6.0000	LOGLOG	6.0000	14.0000

Mean	Standard Error
10.0000	4.0000

The LIFETEST Procedure

Stratum 1: strata = Black

Product-Limit Survival Estimates

time	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	2
44.0000	0.5000	0.5000	0.3536	1	1
62.0000	0	1.0000		2	0

Summary Statistics for Time Variable time

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval Transform	[Lower	Upper)
75	62.0000	LOGLOG	44.0000	62.0000
50	53.0000	LOGLOG	44.0000	62.0000
25	44.0000	LOGLOG	44.0000	62.0000

Mean	Standard Error
53.0000	9.0000

Conclusion

- When the value order for character variable and numeric variable are same, There is no change observed in the results

Character	Char Value Order	Numeric	Num Value order
Low	**2**	1	**1**
Normal	**3**	2	**2**
High	**1**	3	**3**

- SAS/STAT procedures are not dependent on variable data type but dependent on the value order.
- It’s recommended to use numeric data type of any categorical variable in statistical modelling rather than character variable.


Q.
A.

THANK YOU

sakthivel.sivam@quartesian.com

jairaj.ramaiah@quartesian.com