

СТАТИСТИЧЕСКИ СОФТУЕР

Упражнение 5 : Използване на SAS за анализ на данни при повторени във времето изследвания.

Дългосрочният статистически анализ на данни се прилага тогава, когато състоянието на всеки обект, участващ в дадено изследване, се проследява два или повече пъти. Интересуваме се от промяната на състоянието на обекта във времето или зависимостта на резултатите от определени променливи. В повечето случаи разпределението на наблюдаваните случайни величини е неизвестно, а извадките включват малък брой на изследваните единици.

Дългосрочният статистически анализ се прилага в следните случаи:

- ▶ 1) Когато една хомогенна група от индивиди ($k = 1, 2, 3, \dots, n$ обекта) се наблюдава многократно през $j = 1, 2, 3, \dots, b$ времеви интервали;
 - ▶ 2) Когато една хомогенна група от индивиди ($k = 1, 2, 3, \dots, n$ обекта) се наблюдава многократно през $j = 1, 2, 3, \dots, b$ времеви интервали, като за всеки индивид се изследва определена точка няколко пъти във времето (например: различни съдове в тялото, ляво и дясно око);
 - ▶ 3) Когато една хомогенна група от индивиди ($k = 1, 2, 3, \dots, n$ обекта) се наблюдава многократно при $i = 1, 2, 3, \dots, a$ условия. Всеки индивид се наблюдава при всяко условие ($i = 1, 2, 3, \dots, a$) през $j = 1, 2, 3, \dots, b$ времеви интервали;
 - ▶ 4) Когато повече от една хомогенна група от индивиди ($k = 1, 2, 3, \dots, n$, като например: мъжки и женски групи, подложени на различен тип лечение, $i = 1, 2, 3, \dots, a$) се наблюдават многократно през $j = 1, 2, 3, \dots, b$ времеви интервали.
-



СТАТИСТИЧЕСКИ СОФТУЕР

Упражнение 5 : Използване на SAS за анализ на данни при повторени във времето изследвания.

```
data clinic;  
  input ID  
        VisitDate $  
        Dx $  
        HR  
        SBP  
        DBP;  
datalines;  
101 10/21/05 GIProblems 68 120 80  
101 02/25/06 Cold 68 122 84  
255 09/01/05 RoutineVisit 76 188 100  
255 12/18/05 RoutineVisit 74 180 95  
255 02/01/06 HeartProblems 79 210 110  
255 04/01/06 HeartProblems 72 180 88  
303 10/10/06 RoutineVisit 72 138 84  
409 09/01/05 Injury 88 142 92  
409 10/02/05 RoutineVisit 72 136 90  
409 12/15/06 RoutineVisit 68 130 84  
712 04/06/06 Infection 58 118 70  
712 04/15/06 Infection 56 118 72
```

```
proc print data=clinic;  
run;
```

Obs	ID	VisitDate	Dx	HR	SBP	DBP
1	101	10/21/05	GIProble	68	120	80
2	101	02/25/06	Cold	68	122	84
3	255	09/01/05	RoutineV	76	188	100
4	255	12/18/05	RoutineV	74	180	95
5	255	02/01/06	HeartPro	79	210	110
6	255	04/01/06	HeartPro	72	180	88
7	303	10/10/06	RoutineV	72	138	84
8	409	09/01/05	Injury	88	142	92
9	409	10/02/05	RoutineV	72	136	90
10	409	12/15/06	RoutineV	68	130	84
11	712	04/06/06	Infectio	58	118	70
12	712	04/15/06	Infectio	56	118	72

	ID	VisitDate	Dx	HR	SBP	DBP
1	101	10/21/05	GIProble	68	120	80
2	101	02/25/06	Cold	68	122	84
3	255	09/01/05	RoutineV	76	188	100
4	255	12/18/05	RoutineV	74	180	95
5	255	02/01/06	HeartPro	79	210	110
6	255	04/01/06	HeartPro	72	180	88
7	303	10/10/06	RoutineV	72	138	84
8	409	09/01/05	Injury	88	142	92
9	409	10/02/05	RoutineV	72	136	90
10	409	12/15/06	RoutineV	68	130	84
11	712	04/06/06	Infectio	58	118	70
12	712	04/15/06	Infectio	56	118	72

```
712 04/15/06 Infection 56 118 72
```

```
;
```

СТАТИСТИЧЕСКИ СОФТУЕР

Упражнение 5 : Използване на SAS за анализ на данни при повторени във времето изследвания.

Създаване на първа и последна променлива:

- ▶ **proc sort** data=clinic out=clinic;
- ▶ by ID VisitDate;
- ▶ **run;**
- ▶ **data** last;
- ▶ set clinic;
- ▶ by ID;
- ▶ put ID= VisitDate= First.ID= Last.ID=;
- ▶ if last.ID;
- ▶ **run;**

	ID	VisitDate	Dx	HR	SBP	DBP
1	101	10/21/05	GIProble	68	120	80
2	255	12/18/05	RoutineV	74	180	95
3	303	10/10/06	RoutineV	72	138	84
4	409	12/15/06	RoutineV	68	130	84
5	712	04/15/06	Infectio	56	118	72



СТАТИСТИЧЕСКИ СОФТУЕР

Упражнение 5 : Използване на SAS за анализ на данни при повторени във времето изследвания.

- ▶ Определяне на броя на посещенията на пациент, използвайки „DATA step“
- ▶ **data count;**
- ▶ **set clinic;**
- ▶ **by ID;**
- ▶ ***Initialize counter at first visit;**
- ▶ **if first.ID then N_visits = 0;**
- ▶ ***Increment the visit counter;**
- ▶ **N_visits + 1;**
- ▶ ***Output an observation at the last visit;**
- ▶ **if last.ID then output;**
- ▶ **run;**

	ID	VisitDate	Dx	HR	SBP	DBP	N_visits
1	101	02/25/06	Cold	68	122	84	2
2	255	04/01/06	HeartPro	72	180	88	4
3	303	10/10/06	RoutineV	72	138	84	1
4	409	12/15/06	RoutineV	68	130	84	3
5	712	04/15/06	Infectio	56	118	72	2

СТАТИСТИЧЕСКИ СОФТУЕР

Упражнение 5 : Използване на SAS за анализ на данни при повторени във времето изследвания.

Използване на PROC FREQ за определяне на броя на наблюденията по групи (BY group)

```
proc freq data=clinic noprint;  
tables ID / out=counts;  
run;
```

	ID	COUNT	PERCENT
1	101	2	16.6666666667
2	255	4	33.3333333333
3	303	1	8.33333333333
4	409	3	25
5	712	2	16.6666666667



СТАТИСТИЧЕСКИ СОФТУЕР

Упражнение 5 : Използване на **SAS** за анализ на данни при повторени във времето изследвания.

Използване на **RENAME=** и **DROP=** за определяне на броя на наблюденията:

- ▶ **proc freq** data=clinic noprint;
- ▶ tables ID / out=counts (rename=(count = N_Visits)
- ▶ drop=percent);
- ▶ **run;**
- ▶ **data** clinic;
- ▶ merge clinic counts;
- ▶ by ID;
- ▶ **run;**

	ID	N_Visits
1	101	2
2	255	4
3	303	1
4	409	3
5	712	2



СТАТИСТИЧЕСКИ СОФТУЕР

Упражнение 5 : Използване на SAS за анализ на данни при повторени във времето изследвания.

Използване на PROC MEANS за определяне на броя на наблюденията по групи (BY group):

- ▶ **proc means** data=clinic nway noprint;
- ▶ class ID;
- ▶ output out=counts(rename=(**_freq_** = **N_Visits**))
- ▶ drop= **_type_**);
- ▶ **run**;

	ID	N_Visits	_STAT_	HR	SBP	DBP
1	101	2	N	2	2	2
2	101	2	MIN	68	120	80
3	101	2	MAX	68	122	84
4	101	2	MEAN	68	121	82
5	101	2	STD	0	1.4142135624	2.8284271247
6	255	4	N	4	4	4
7	255	4	MIN	72	180	88
8	255	4	MAX	79	210	110
9	255	4	MEAN	75.25	189.5	98.25
10	255	4	STD	2.9860788112	14.177446879	9.2511260576
11	303	1	N	1	1	1
12	303	1	MIN	72	138	84
13	303	1	MAX	72	138	84
14	303	1	MEAN	72	138	84
15	303	1	STD	.	.	.
16	409	3	N	3	3	3
17	409	3	MIN	68	130	84

СТАТИСТИЧЕСКИ СОФТУЕР

Упражнение 5 : Използване на **SAS** за анализ на данни при повторени във времето изследвания.

Използване на „**RETAIN**” функция за запомняне на предишни стойност на променливата:

▶ data hypertension;

▶ set clinic;

▶ by ID;

▶ retain HighBP;

▶ if first.ID then HighBP = 'No ';

▶ if SBP gt **140** then HighBP = 'Yes';

▶ if last.ID then output;

▶ run;

	ID	VisitDate	Dx	HR	SBP	DBP	HighBP
1	101	02/25/06	Cold	68	122	84	No
2	255	04/01/06	HeartPro	72	180	88	Yes
3	303	10/10/06	RoutineV	72	138	84	No
4	409	12/15/06	RoutineV	68	130	84	Yes
5	712	04/15/06	Infectio	56	118	72	No



СТАТИСТИЧЕСКИ СОФТУЕР

**Упражнение 5 : Използване на SAS за анализ на данни при повторени
във времето изследвания.**

- ▶ data difference;
- ▶ set clinic;
- ▶ by ID;
- ▶ if first.ID and last.ID then delete;
- ▶ Diff_HR=HR- lag(HR);
- ▶ Diff_SBP=SBP-lag(SBP);
- ▶ Diff_DBP=DBP-lag(DBP);
- ▶ if not first.ID then output;
- ▶ run;



СТАТИСТИЧЕСКИ СОФТУЕР

**Упражнение 5 : Използване на SAS за анализ на данни при повторени
във времето изследвания.**

- ▶ data difference;
 - ▶ set clinic;
 - ▶ by ID;
 - ▶ if first.ID and last.ID then delete;
 - ▶ if first.ID or last.ID then do;
 - ▶ Diff_HR=HR- lag(HR);
 - ▶ Diff_SBP=SBP-lag(SBP);
 - ▶ Diff_DBP=DBP-lag(DBP);
 - ▶ end;
 - ▶ if last.ID then output;
 - ▶ run;
-



СТАТИСТИЧЕСКИ СОФТУЕР

Упражнение 5 : Използване на SAS за анализ на данни при повторени във времето изследвания.

- ▶ **data first_last;**
- ▶ **set clinic;**
- ▶ **by ID;**
- ▶ **if first.ID and last.ID then delete;**
- ▶ **retain first_HR first_SBP first_DBP;**
- ▶ **if first.ID then do;**
- ▶ **first_HR=HR;**
- ▶ **first_SBP=SBP;**
- ▶ **first_DBP=DBP;**
- ▶ **end;**
- ▶ **if last.ID then do;**
- ▶ **diff_HR=HR-first_HR;**
- ▶ **diff_SBP=SBP-first_SBP;**
- ▶ **diff_DBP=DBP-first_DBP;**
- ▶ **end;**
- ▶ **drop first_;;**

- ▶ **run;**