

## محاسبه رگرسیون Ordinal در نرم افزار SAS

به منظور بررسی و مطالعه در مورد متغیر های پیش بین موثر بر ابعاد یک متغیر گسسته از رگرسیون های ترتیبی استفاده می شود. این روش رگرسیونی مبتنی بر روش حداکثر درستنمایی است. این روش در مقابل روش هایی چون تحلیل تشخیصی که روشی صرفاً کمی است بسیار کامل بوده و امروزه در تمام حوزه هایی که شانس رخ داد تصادفی هر یک از ابعاد یک متغیر وابسته را بخواهند برآورد کنند مورد استفاده قرار می گیرد. به عنوان مثال یک شرکت بازار یابی می خواهد در مورد متغیر های پیش بین موثر بر میزان علاقه مندی افراد به محصولات یک شرکت تولید خودرو معادله رگرسیونی را برآورد نماید. به این منظور با یک طیف لیکرت میزان علاقه مندی افراد را به محصولات این شرکت در قالب چهار گروه مورد بررسی قرار می دهد. یا میزان علاقه مندی افراد به سه اندازه تپلت ها تحت تاثیر چه متغیر هایی است؟ یا اینکه یک موسسه آموزشی علاقه مند است تا مطالعه کند متغیر های اصلی روی میزان مندی دانشجویان به ثبت نام در قالب سه طیف بسیار علاقه مند، علاقه مند و تا حدی چه متغیر هایی هستند. میزان تحصیلات والدین، دولتی یا خصوصی بودن موسسه قبلی و نمره معدل دوره قبلی فرد مورد توجه قرار گرفته و به عنوان متغیر مستقل مورد مطالعه قرار گرفتند. در ادامه با تعریف داده ای فرضی این مثال در نرم افزار SAS حل می شود.

ابتدا در محیط برنامه نویسی SAS دستور زیر را وارد می کنیم:

```
proc freq data = "آدرس فایل داده";  
tables apply;  
tables pared;  
tables public;  
run;
```

با این دستور می توان آماره توصیفی داده های وارد شده را مشاهده نمود. به عنوان مثال ۲۲۰ نفر از افراد مورد مطالعه علاقه تا حدی داشته و ۴۰ نفر بسیار مشتاق بوده اند. ۳۳۷ نفر تحصیلات والدینشان غیر مرتبط با رشته فرد متقاضی و ۶۳ نفر تحصیلات والدینشان مرتبط بوده است. ۳۴۳ نفر قبلاً در موسسه دولتی و ۵۷ نفر در موسسه خصوصی دوره قبلی خود را طی کرده اند.

The FREQ Procedure

APPLY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	220	55.00	220	55.00
1	140	35.00	360	90.00
2	40	10.00	400	100.00

PARED	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	337	84.25	337	84.25
1	63	15.75	400	100.00

  

PUBLIC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	343	85.75	343	85.75
1	57	14.25	400	100.00

در صورتیکه بخواهیم میانگین متغیرها را نیز داشته باشیم دستور زیر را وارد می‌کنیم.

```
proc means data = " آدرس فایل داده ";
var gpa;
run;
```

نکته اصلی اینجا است که اجرای این مدل نیازمند این است که در هر سلول حداقل تعدادی از افراد قرار گیرند. به این منظور از دستور زیر استفاده می‌شود.

```
proc freq data = " آدرس فایل داده ";
tables apply*pared / nopercnt norow nocol missprint;
tables apply*public / nopercnt norow nocol missprint;
run;
```

اگر در خروجی سلولی خالی نبوده و یا مقدار آن کوچک نباشد می‌توان این روش را اجر نمود. دستور اصلی برای اجرای این برنامه به شرح زیر است:

```
proc logistic data = " آدرس فایل داده " desc;
model نام متغیر وابسته = نام های متغیر های مستقل با یک فاصله بین
run;
```

خروجی نرم افزار به شرح زیر خواهد بود:

```
The LOGISTIC Procedure
Number of Observations Read      400
Number of Observations Used      400
Response Profile

Ordered Value      APPLY      Total
Frequency

1      2      40
2      1      140
3      0      220
```

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
4.8446	3	0.1835

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	745.205	727.025
SC	753.188	746.982
-2 Log L	741.205	717.025

جهت مقایسه بین مدل های مختلف این آماره ها مورد استفاده قرار می گیرند. به عبارت دیگر مقادیر این آماره ها در تست بین مدل ها و کارایی آنها مورد استفاده قرار می گیرد.

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	24.1804	3	<.0001
Score	23.4804	3	<.0001
Wald	24.3337	3	<.0001

آماره کای اسکویر برای تست حداکثر درستنمایی معنی دار بوده و نشان می دهد در تابع حداقل یک متغیر مستقل وجود دارد که توانایی پیش بینی ابعاد متغیر وابسته را دارد.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 2	1	-4.2983	0.8092	28.2189	<.0001
Intercept 1	1	-2.2029	0.7844	7.8869	0.0050
PARED	1	1.0478	0.2684	15.2350	<.0001
PUBLIC	1	-0.0585	0.2886	0.0411	0.8393
GPA	1	0.6156	0.2626	5.4963	0.0191

در این قسمت مشخص می شود که دو متغیر مستقل معدل و تحصیلات والدین دارای نقش معنی داری بر سطوح متغیر وابسته هستند. سطح معنی داری این متغیر های مستقل با استفاده از آماره والد محاسبه می شود.

### Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
PARED	2.851	1.685	4.826
PUBLIC	0.943	0.536	1.661
GPA	1.851	1.106	3.096

نتایج نشان می دهد در صورتیکه میانگین معدل یک واحد افزایش یابد احتمال اینکه فردی از گروه علاقه مندی تا حدی به گروه بالاتر رود ۱.۸۵ واحد افزایش می یابد. در مورد متغیر تحصیلات والدین نیز این امر صادق است و در صورتیکه والدین تحصیلات مرتبط با رشته تحصیلی دانشجو داشته باشند این میزان ۲.۸۵ واحد افزایش می یابد. به عبارت دیگر در بین متغیرهای مستقل متغیرهای تحصیلات والدین و نمره معدل دارای نقش های معنی دار به لحاظ آماری هستند و نوع موسسه قبلی فرد در این بین نقشی ندارد. در این مقاله سعی بر معرفی روش رگرسیون ترتیبی در نرم افزار SAS بود.

www.AcademiaCafe.com



*AcademiaCafe*

a non-profit website to promote education