

با سلام و احترام،

ادامه فصل بیستم و ششم: آزمون ناپارامتریک

آزمون دو جمله‌ای

در برخی از فرضیه‌ها، با متغیرهایی سر و کار داریم که دو وجهی هستند. یعنی دو مقوله یا طبقه دارند و بنا داریم تا نسبت این مقوله‌ها را با همدیگر و با توجه به یک نسبت فرضی مقایسه کنیم. نسبت افراد قبول شده در کنکور سراسری و نسبت افرادی که رد شده‌اند، مقایسه نسبت روزهای بارانی به غیربارانی و... از جمله مواردی هستند که می‌توان به کمک آزمون ناپارامتری دو جمله‌ای، آزمون فرضیه کرد. آزمون دو جمله‌ای مشخص می‌کند که آیا نسبت مشاهده شده با نسبت مفروض فرق دارد یا نه؟

محاسبه چند توزیع

یک شرکت مخابراتی حدود **27%** مشتریان خود را برای پارازیت‌های هر ماه از دست می‌دهد. به منظور رفع این معضل، مدیریت می‌خواهد بداند که آیا این درصد در سرتاسر گروه‌های مشتریان تغییر می‌نماید. با استفاده از آزمون دو جمله‌ای تعیین کنید که آیا نرخ واحد از پارازیت، به قدر کافی چهار نوع اصلی مشتریان را توصیف می‌نماید.

آماده کردن داده‌ها

به منظور اجرای آزمون، ابتدا باید فایل را با **Customer Category** تقسیم نمایید.

۱. برای تقسیم فایل، مسیر **Data > Split File** را از منوی اصلی **Data Editor** برگزینید.
۲. گزینه **Customer Category** را انتخاب کنید.
۳. گزینه **Customer Category** را به عنوان متغیری که گروه‌ها بر آن بنا می‌شوند (**Groups Based**) انتخاب نمایید.
۴. **Ok** را کلیک کنید.

آغاز تحلیل



۱. برای شروع تحلیل، مسیر **Analyze > Nonparametric Tests > Legacy Dialogs > Binomial** را از منوی اصلی برگزینید.
۲. **Churn Within last month** را به عنوان متغیر آزمون برگزینید.
۳. **0.27** را به عنوان نسبت آزمون وارد کنید.

چون که آزمون دو جمله‌ای، نسبت مشاهده شده را با یک نسبت مفروض مقایسه می‌کند، بنابراین لازم است که این نسبت مفروض را بدانیم. نسبت مفروض، همان نسبتی از پاسخگویان است که با توجه به فرضیه تحقیق، فرض شده این نسبت باید در میان پاسخگویان وجود داشته باشد تا فرضیه ما تأیید شود و در صورت عدم وقوع چنین نسبتی از نمونه در نسبت مشاهده شده، فرضیه رد می‌شود. البته گاهی پژوهشگران، این نسبت را با فرض اینکه هیچ گونه ارجحیتی برای نسبت‌های مشاهده شده و مفروض وجود ندارد به صورت برابر و معادل **0.5** در نظر می‌گیرند. در مثال مورد نظر ما این نسبت معادل **0.27** در نظر گرفته شده است.

نهایتاً موقع تفسیر نتایج این آزمون باید توجه کرد که اولاً تفاوت وجود دارد یا نه؟ که این کار با مشاهده معنی داری تفاوت در سطح کوچکتر از **0.05** انجام می‌شود. سپس باید پرسید که آیا این تفاوت در راستای فرض هست یا برخلاف آن؟ که این کار هم با مقایسه نسبت مشاهده شده و نسبت مفروض امکان‌پذیر است.

۴. دکمه **options** را کلیک کنید.
۵. گزینه **Descriptive** را فعال کنید.
۶. دکمه **Continue** را کلیک کنید.
۷. **Ok** را در کادر محاوره **Binomial Test** کلیک کنید.

جدول آماره‌های توصیفی

چون **Churn Within Last month** یک متغیر دو مقوله‌ای است، میانگین، نسبت پارازیت را با هر نوع مشتری بیان می‌نماید، ضرب این نسبت‌ها در **100**، همین داده‌ها را بر حسب درصد بیان می‌نماید.

Customer category		N	Mean	Std. Deviation	Minimum	Maximum
Basic service	Churn within last month	266	31	464	0	1
E-service	Churn within last month	217	27	446	0	1
Plus service	Churn within last month	281	16	364	0	1
Total service	Churn within last month	236	37	485	0	1

برای مثال، درصد پارازیت برای مشتریانی که فقط با سرویس اصلی مشترک شده‌اند برابر **31%** می‌باشد. به طور مشابهی، مشتریانی که با سرویس‌های الکترونیکی اشتراک یافته‌اند نرخ پارازیتی معادل **27%** در ماه گذشته

گزارش شده است. حدود ۲۸۰ مشتری از سرویس‌های ویژه برای راحتی بیشتر استفاده می‌نمایند. در این مورد فقط ۱۶٪ اغتشاش وجود دارد. درصد اغتشاش مربوط به مشتریانی که با همه سرویس‌ها مشترک شده‌اند ۳۷٪ می‌باشد، به عبارتی ۱۰٪ بالاتر از میانگین مشتریان مربوط به ماه قبل.

جدول آزمون دو جمله‌ای

هر پانل جدول آزمون دو جمله‌ای یک آزمون دو جمله‌ای را نشان می‌دهد. مثلاً، پانل اول آزمون فرضیه صفری که بیان می‌کند نسبت پارازیت برای کاربران Basic Service همان نسبت پارازیت در نمونه کل می‌باشد را نشان می‌دهد (a).

از ۲۶۶ مشتری Basic Service، ۸۳ پارازیت در ماه گذشته وجود دارد. ستون Observed Prop. در اینجا نشان می‌دهد که این ۸۳ مشتری، ۳۱٪ برای گروه Basic Service در این نمونه در نظر گرفته شده‌اند (b). نسبت آزمون ۰.۲۷ حکایت از آن دارد که باید انتظار داشته باشیم $0.27 * 266$ ، یا حدود ۷۲ مشتری را در نتیجه پارازیت از دست بدهیم (c).

Customer category	a	b	Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (1-tailed)
Basic service	Churn within last month	Group 1	Yes	83	.31	.27	.070 ^a
		Group 2	No	183	.69		
		Total		266	1.00		
E-service	Churn within last month	Group 1	Yes	59	.27	.27	.505 ^a
		Group 2	No	158	.73		
		Total		217	1.00		
Plus service	Churn within last month	Group 1	No	237	.84	.27	.000 ^a
		Group 2	Yes	44	.16		
		Total		281	1.00		
Total service	Churn within last month	Group 1	Yes	88	.37	.27	.000 ^a
		Group 2	No	148	.63		
		Total		236	1.00		

استفاده از نقطه برش برای تعیین نمونه‌ها

در اینجا مسئله شرکت مخابراتی (تعویض شرکت توسط مشتریان) را ادامه می‌دهیم.

از دستورالعمل Binomial Test برای محاسبه دینامیکی و بررسی نسبت هر گروه پارازیت که در زیر مقدار میانه می‌افتد استفاده می‌کنید.

آماده کردن داده‌ها

برای اجرای آزمون، باید ابتدا فایل را با **Churn Within Last month** تقسیم نمایید.

۱. برای تقسیم فایل، مسیر **Data > Split File** را از منوی اصلی **Data Editor** انتخاب کنید.
۲. دکمه **Reset** را کلیک کنید تا تنظیمات پیش فرض بازخوانی شوند.
۳. گزینه **Compare groups** را انتخاب نمایید.
۴. **Churn Within Last month** را به عنوان متغیری که گروه‌ها بر آن قرار می‌گیرند انتخاب کنید.
۵. **Ok** را کلیک کنید.

شروع تحلیل

۱. برای شروع تحلیل، مسیر **Analyze > Nonparametric Tests > Legacy Dialogs > Binomial** را از منوی اصلی برگزینید.
۲. دکمه **Reset** را کلیک کنید تا تنظیمات پیش فرض بازخوانی شوند.
۳. **Household income in thousands** را به عنوان متغیر آزمون برگزینید.
۴. **47** را به عنوان نقطه برش تایپ کنید تا داده‌ها در دو دسته تعریف گردد.
۵. دکمه **options** را کلیک کنید.
۶. گزینه **Quartiles** را فعال نمایید تا خروجی به صورت چارکی نمایش داده شود.
۷. دکمه **Continue** را کلیک کنید.
۸. **Ok** را در کادر محاوره **Binomial Test** کلیک کنید.

چارک‌های درآمد با گروه پرازیت

جدول توصیفی چارک‌های هر گروه پرازیت را نمایش می‌دهد. معمولاً مشتری‌هایی که در ماه گذشته پرازیت داشته‌اند کمترین درآمدها را داشته‌اند.

جدول آزمون دو جمله‌ای با مقادیر **Churn Within Last month** تقسیم می‌شوند. آزمون اول فقط آن مشتریانی در نمونه که در ماه گذشته پرازیت نداشته‌اند انتخاب می‌نماید (a). با این گروه فایل تقسیمی اول، نقطه

برش دو گروه را ایجاد کرده است. گروه اول شامل مشتریانی است که پارازیت نداشته‌اند و عایدی آنها کمتر یا مساوی میانه مربوط به نمونه کل می‌باشد (b).

Churn within last month		N	Percentiles		
			25th	50th (Median)	75th
No	Household income in thousands	726	30.0000	49.0000	89.2500
Yes	Household income in thousands	274	26.7500	41.0000	70.0000

در این داده‌ها فقط نیمی از آنها دارای درآمد متوسط یا کمتر از آن هستند. همانطور که انتظار می‌رود، اختلاف در نسبت‌ها معنی‌دار نیست (چون sign بزرگتر از 0.05 می‌باشد) (c). با این وجود 274 مشتری که ماه گذشته پارازیت داشته‌اند، نسبت درآمدها در میانه یا پایین‌تر از آن به طور معنی‌داری از مقدار فرضیه صفر اختلاف دارد

Churn within last month		Category		N	Observed Prop.	Test Prop.	Asymp. Sig. (2-tailed)
No	Household income in thousands	Group 1	<= 47	345	48	.50	.194
		Group 2	> 47	381	52		
		Total		726	1.00		
Yes	Household income in thousands	Group 1	<= 47	160	58	.50	.007
		Group 2	> 47	114	42		
		Total		274	1.00		

a. Based on Z Approximation.



ادامه فصل 26 را در مرجع کاربردی SPSS 20 (46) دنبال نمایید.



Telegram.me/iepnu
کانال تخصصی مهندسی صنایع دانشگاه پیام نور