

با سلام و احترام،

ادامه فصل بیستم و ششم: آزمون ناپارامتریک

آزمون تک نمونه‌ای کلموگوروف - اسمیرنوف

دستورالعمل **One - Sample Kolpnogorov - Smirnov** برای بررسی این فرضیه صفر که یک نمونه دارای یک توزیع خاص است به کار می‌رود. این کار با یافتن بزرگترین اختلاف (مقدار مطلق می‌باشد) بین دو تابع توزیع تراسمی (CDFs) که یکی مستقیماً از داده‌ها به دست می‌آید؛ و دیگری از تئوری ریاضی انجام می‌شود. همانطور که می‌دانیم فرمول **x2** (خی دو) تک متغیره وقتی صادق است که فراوانی‌های مورد انتظار بیش از ۲۰ درصد خانه‌های جدول کمتر از ۵ نباشد. بنابراین وقتی که فراوانی‌های مورد انتظار بیش از ۲۰ درصد خانه‌ها کمتر از ۵ باشد، دیگر فرمول **x2** صادق نیست. این مسئله اغلب در صورتی اتفاق می‌افتد که حداقل حجم نمونه کوچکتر از ۵۰ باشد. به سخن دیگر، موقعی که حجم نمونه کوچکتر از ۵۰ باشد، امکان اینکه فراوانی‌های مورد انتظار بیش از ۲۰ درصد خانه‌های جدول کمتر از ۵ باشد، وجود دارد. در چنین حالتی به جای آزمون **x2**، از آزمون تک نمونه‌ای کولموگوروف - اسمیرنوف استفاده می‌شود.

این آزمون به مقایسه تابع توزیع تجمعی مشاهده شده با تابع توزیع تجمعی نظری (مورد انتظار) در یک متغیر ترتیبی می‌پردازد. به عبارتی، در این آزمون توزیع یک صفت در یک نمونه با توزیعی که برای آن جامعه مفروض است، مقایسه می‌شود.

این آزمون ناپارامتری و بدون توزیع است. اما باید توزیع متغیر در جامعه برای هر یک از رتبه‌های مقیاس رتبه‌ای در جامعه به طور نسبی در نظر گرفته شود که آن را نسبت مورد انتظار می‌نامند.

چهار توزیع تئوری در دسترس است **Normal, Uniform, Poisson** و **Exponential** به طور اختیاری، آماره‌های توصیفی و یا چارک‌های متغیر آزمون می‌تواند نمایش داده شود.

تعیین میزان انطباق

یک تحلیل‌گر بیمه می‌خواهد تعداد اتومبیل‌های تصادفی را نسبت به تعداد راننده مدل نماید. او به طور تصادفی از راننده‌های در یک منطقه خاص داده‌برداری کرده و از آزمون کلموگوروف - اسمیرنوف برای بررسی اینکه تعداد تصادفات تابع توزیع پواسن هستند استفاده می‌نماید.

آدرس: میدان

۱. مسیر **Analyze > Nonparametric Tests > Legacy Dialogs > 1 – Sample K – S** را از منوی اصلی برگزینید.

۲. گزینه **Number of accidents 5 years** را به عنوان متغیر آزمون انتخاب نمایید.

۳. **Normal** را غیر فعال، و **Poisson** را به عنوان توزیع آزمون فعال نمایید.

۴. **Ok** را کلیک کنید.

توزیع بر اساس ۵۰۰ نمونه راننده می باشد که به طور تصادفی انتخاب شده اند (N). توزیع پواسن فقط با یک پارامتر (میانگین) شناخته می شود. در این نمونه راننده ها به طور متوسط، حدود 1.72 تصادف در پنج سال گذشته داشته اند. سه سطر بعدی در مقوله عمومی **Most Extreme Differences** قرار می گیرند. این اختلاف ها، بزرگترین نقاط تفاوت مثبت و منفی بین **CDFs** تجربی و نظری می باشد (a).



		Number of accidents past 5 years
N		500
Poisson Parameter ^{a,b}	Mean	1.72
Most Extreme Differences	Absolute	.065
	Positive	.065
	Negative	-.041
Kolmogorov-Smirnov Z		1.460
Asymp. Sig. (2-tailed)		.028

a. Test distribution is Poisson.

b. Calculated from data

اولین مقدار اختلاف، با عنوان **Absolute**، مقدار مطلق بزرگترین دو مقدار اختلاف موجود در زیر آن می باشد، لازم خواهد داشت این مقدار، آماره آزمون را محاسبه نماید.

تفاوت مطلق نشان دهنده بیشترین تفاوت مطلق بین تابع توزیع تجمعی مشاهده شده و مفروض است. مقدار مطلق همیشه از دو مقدار تفاوت مثبت و منفی بیشتر است. مقدار آماره آزمون کولموگروف – اسمیرنوف (Z) نیز از همین مقدار تفاوت مطلق محاسبه شود.

اختلاف **Positive** نقطه در تابع توزیع تراکم (CDF) عملی است که از **CDF** تئوری با بیشترین مقدار تجاوز می نماید.

تفاوت مثبت نقطه‌ای است که در آن مقدار تابع توزیع تجمعی مشاهده شده از مقدار تابع توزیع تجمعی مورد انتظار بیشتر است.

در نقطه مقابل، اختلاف **Negative** نقطه در **CDF** تئوری است که از **CDF** عملی با بیشترین مقدار تجاوز می‌نماید.

تفاوت منفی نقطه‌ای است که در آن مقدار تابع توزیع تجمعی مشاهده شده از مقدار تابع توزیع تجمعی مورد انتظار کمتر است.

آماره آزمون **Z** حاصلضرب جذر (ریشه دوم) اندازه نمونه و بزرگترین اختلاف مطلق بین **CDF** های عملی و تئوری می‌باشد.

برخلاف بیشتر آزمون‌های آماری، سطح معنی‌داری حاصله در اینجا (**Asymp. Sig. (2tailed)**) نتیجه خوبی نیست. احتمال آماره **Z** در زیر **0.05** به این معنی است که توزیع پواسن با یک پارامتر **1.72** نشان از انطباق مناسب با تعداد تصادف‌های پنج سال گذشته در این راننده‌های نمونه نیست.

میزان انطباق با گروه‌ها

در حالت کلی، یک آزمون معنی‌دار کولموگروف - اسمیرنوف به دو معنی است، یا توزیع تئوری مناسب نیست، یا یک پارامتر نادرست برای ساخت آن توزیع به کار رفته است. به نتایج قبل نگاه کنید، برای تحلیل‌گر مشکل است که تصور کند توزیع پواسن برای مدل کردن تصادف‌های اتومبیل مناسب نیست. توزیع پواسن اغلب برای مدل کردن رویدادهای نادر به کار می‌رود و خوشبختانه، تصادف‌های اتومبیل نسبتاً نادر می‌باشد.

تحلیل‌گر نگران آن است که ممکن است جنسیت نتیجه آزمون را عوض نماید. میانگین جمع کل نمونه فرض کرده است که مردان و زنان تعداد تصادف مساوی داشته‌اند، اما این ممکن است درست نباشد. او نمونه را با جنسیت تقسیم خواهد کرد، از میانگین هر جنس به عنوان پارامتر پواسن در آزمون‌های جدا استفاده خواهد کرد.

تقسیم فایل

۱. برای شکستن فایل، مسیر **Data > Split File** را از منوی اصلی **Data Editor** برگزینید.
۲. گزینه **Compare groups** را انتخاب کنید.
۳. **Sex of insured** را به عنوان متغیر آزمون برگزینید.
۴. **Ok** را کلیک کنید.

اجرای تحلیل

برای اجرای دوباره آزمون تک نمونه‌ای **K-S**، کادر محاوره **One - Sample Kolpnogorov – Smirnov Test** را فراخوانی نمایید.

کادر محاوره به تنظیم مجدد نیاز ندارد، تحلیل‌گر می‌خواهد آماره‌های توصیفی را در متغیر آزمون با هر گروه به دست آورد.

۱. **Options** را کلیک کنید.

۲. گزینه **Descriptive** را فعال نمایید.

۳. دکمه **Continue** را کلیک کنید.

۴. **Ok** را در کادر محاوره **One - Sample Kolpnogorov – Smirnov Test** کلیک کنید.

آماره‌های توصیفی با گروه

جدول آماره‌ها نشان می‌دهد که یک پارامتر پواسن واحد برای هر دو جنس ممکن است صحیح نباشد. مردان در این نمونه‌گیری به طور میانگین حدود دو تصادف در پنج سال گذشته داشته‌اند، در حالی که زنان تعداد تصادف کمتری داشته‌اند (a).

Sex of insured		N	a. Mean	Std. Deviation	Minimum	Maximum
Male	Number of accidents past 5 years	250	1.98	1.608	0	7
Female	Number of accidents past 5 years	250	1.47	1.412	0	6

جداول آزمون با گروه

بعد از تعیین میزان انطباق، به خاطر آورید که یک آماره **Z** معنی‌دار حکایت از آن دارد که توزیع انتخاب شده به خوبی با داده‌ها منطبق نیست. برخلاف آزمون قبل، با تقسیم فایل به نتایج بهتری دسترسی می‌یابیم.

افزایش پارامتر پواسن از **1.72** به **1.98** به وضوح انطباق بهتری برای داده‌های تصادفات مردان ایجاد می‌نماید (a).

به طور مشابهی، کاهش پارامتر پواسن از **1.72** به **1.42** انطباق بهتری را برای داده‌های تصادفات زنان نشان می‌دهد (b).

Sex of insured				Number of accidents past 5 years
Male	N			250
	Poisson Parameter λ	Mean		1.46
	Most Extreme	Absolute		1.47
	Differences	Positive		1.47
		Negative		1.45
	Heterogeneity Statistic χ^2			a 15.01
	Asymp. Sig. (2-tailed)			b .002
Female	N			250
	Poisson Parameter λ	Mean		1.47
	Most Extreme	Absolute		1.74
	Differences	Positive		1.74
		Negative		1.42
	Heterogeneity Statistic χ^2			b 11.14
	Asymp. Sig. (2-tailed)			.002

a. Test distribution is Poisson.
b. Calculated from data.



ادامه فصل ۲۶ را در مرجع کاربردی SPSS 20 (48) دنبال نمایید.



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