فصل ششم

توزیع مشترک دو صفت متغیر

هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند :

۱ـ جامعه را برحسب دو صفت متغیر کمی گروه‌بندی نماید.

۲ـ جدول توزیع فراوانی‌های دو صفت متغیر را تشکیل دهد.

۳ـ توزیعهای حاشیه‌ای x و y را به‌دست آورد.

۴ـ توزیعهای شرطی y برحسب x و x برحسب y را به‌دست آورد.

۵ ـ مشخصه‌های عددی برای دو صفت متغیر را به تفکیک محاسبه نماید.

۶ ـ مفهوم پراکندگی توأم دو صفت متغیر را بیان نماید.

در فصلهای گذشته، راجع به مطالعه جامعه‌ها با یک صفت متغیر بحث کردیم. ولی اغلب برای یک محقق نیاز می‌شود که اعضاء جامعه را توسط چند صفت متغیر با هم مورد مطالعه قرار دهد. در این فصل فقط به مطالعه دو صفت متغیر از اعضاء جامعه می‌پردازیم.

توزیع مشترک دو صفت متغیر x و y

اگر برای اعضاء جامعه دو صفت متغیر x و y با هم اندازه‌گیری شوند، آنگاه برای هریک از اعضاء یک زوج اندازه یا مقدار به‌دست خواهد آمد. مجموع زوجهای به‌دست آمده از نتایج مشاهدات مقادیر دو صفت را می‌توان به‌صورت زیر نشان داد.

شماره اعضاء	۱	۲	۳	...	n
اندازه صفت x	x_1	x_2	x_3	...	x_n
اندازه صفت y	y_1	y_2	y_3	...	y_n

چنانچه حجم جامعه بزرگ باشد، به‌طوری که هر زوج از اندازه‌های x و y بیش از یک بار

تکرار شوند، لزوماً نتایج مشاهدات را در جدول توزیع فراوانی دو بعدی وارد می‌کنند که در آن فراوانیها (F) با دو اندیس i برای صفت x و j برای صفت y مشخص می‌گردند، یعنی به‌صورت F_{ij} نشان داده می‌شوند. مجموعه دو صفت متغیر را دستگاه دو متغیر نیز می‌نامند. جدول ۱ توزیع فراوانیها را برای دو صفت متغیر x و y نشان می‌دهد :

جدول ۱

x \ y	y_1	y_2	......	y_j	......	y_t	$F_{i\circ}$
x_1	F_{11}	F_{12}	......	F_{1j}	......	F_{1t}	$F_{1\circ}$
x_2	F_{21}	F_{22}	......	F_{2j}	......	F_{2t}	$F_{2\circ}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	F_{i1}	F_{i2}	......	F_{ij}	......	F_{it}	$F_{i\circ}$
$\vdots$	$\vdots$						$\vdots$
x_s	F_{s1}	F_{s2}	......	F_{sj}	......	F_{st}	$F_{s\circ}$
$F_{\circ j}$	$F_{\circ 1}$	$F_{\circ 2}$	......	$F_{\circ j}$	......	$F_{\circ t}$	n

در این جدول اگر فراوانیهای درون جدول (F_{ij}) را روی سطر جمع کنیم خواهیم داشت :

$$\sum_{j=1}^{t} F_{ij} = F_{i1} + F_{i2} + \cdots + F_{it} = F_{i\circ}$$

چون فراوانیها روی اندیس j جمع شده‌اند بنابراین جواب حاصل به‌صورت $F_{i\circ}$ نوشته خواهد شد. هم‌چنین اگر فراوانیها را روی اندیس i جمع ببندیم جواب حاصل به‌صورت $F_{\circ j}$ نوشته خواهد شد.

$$\sum_{i=1}^{s} F_{ij} = F_{1j} + F_{2j} + \cdots + F_{sj} = F_{\circ j}$$

بنابراین جمع سطرهای جدول به‌صورت $F_{1\circ}$ ، $F_{2\circ}$ ، ...، $F_{s\circ}$ درخواهد آمد و اگر آنها را روی اندیس i جمع کنیم حاصل حجم جامعه یعنی n را خواهد داد. و اگر $F_{\circ j}$ ها را روی j جمع کنیم نتیجه برابر حجم جامعه یعنی n خواهد شد.

$$\sum_{i=1}^{s} F_{i\circ} = F_{1\circ} + F_{2\circ} + \cdots + F_{s\circ} = n$$

$$\sum_{j=1}^{t} F_{\circ j} = F_{\circ 1} + F_{\circ 2} + \cdots + F_{\circ t} = n$$

مثال ۱: فرض کنید از جامعه دانش‌آموزان یک دبیرستان دو صفت متغیر x و y، که در آن x سن دانش‌آموزان و y تعداد افراد خانوار دانش‌آموزان، پرسش گردیده نتایج به‌صورت جدول توزیع فراوانیها (جدول۲) به‌دست آمده باشد.

جدول ۲

y تعداد افراد خانوار / x سن	۲	۳	۴	۵	۶	جمع سطر
۱۴	۳	۴	۶	۱۲	۵	۳۰
۱۵	۲	۷	۱۰	۱۱	۵	۳۵
۱۶	۴	۵	۹	۱۰	۴	۳۲
۱۷	۶	۴	۱۲	۱۰	۵	۳۷
۱۸	۳	۳	۷	۸	۷	۲۸
۱۹	۲	۶	۵	۴	۱	۱۸
جمع ستون	۲۰	۲۹	۴۹	۵۵	۲۷	۱۸۰

اعداد داخل جدول نشانه فراوانیهای هر زوج از دو صفت متغیر می‌باشد. برای مثال ۴ دانش‌آموز ۱۴ ساله تعداد نفرات خانوارشان ۳ نفره است و هم‌چنین ۱۰ دانش‌آموز ۱۵ ساله تعداد نفرات خانوارشان ۴ نفره می‌باشد.

توزیعهای حاشیه‌ای

از این جدول دوبعدی می‌توان توزیع هریک از صفات متغیر را جداگانه به‌دست آورد. مثلاً بدون درنظر گرفتن صفت y، می‌توان توزیع صفت متغیر x را به‌دست آورد.

جدول ۳

جمع	x_s	……	x_i	……	x_2	x_1	x
n	$F_{s\circ}$	……	$F_{i\circ}$	……	$F_{2\circ}$	$F_{1\circ}$	فراوانی $F_{i\circ}$

برای مثال ۱ توزیع سن دانش‌آموزان را بدون درنظر گرفتن تعداد افراد خانوار تشکیل می‌دهیم :

جدول ۴

جمع	۱۹	۱۸	۱۷	۱۶	۱۵	۱۴	سن x
۱۸۰	۱۸	۲۸	۳۷	۳۲	۳۵	۳۰	فراوانی $F_{i\circ}$

چون فراوانیها برای مقادیر صفت x (سن) از ستون کناری جدول قرار داده شده، به چنین توزیعی، توزیع کناری یا توزیع حاشیه‌ای x گویند. این توزیع نشان می‌دهد که توزیع سن دانش‌آموزان این دبیرستان چگونه است. همین‌طور می‌توان توزیع صفت y را به‌طور جداگانه از روی جدول دوبعدی ۱ به‌دست آورد :

جدول ۵

جمع	y_t	……	y_j	……	y_2	y_1	y
n	$F_{\circ t}$	……	$F_{\circ j}$	……	$F_{\circ 2}$	$F_{\circ 1}$	فراوانی $F_{\circ j}$

چنین توزیعی را توزیع حاشیه‌ای y نامند. چون در برابر مقادیر صفت y فراوانیها را از سطر پایین جدول ۱ قرار داده‌ایم.

توزیع حاشیه‌ای y را برای مثال ۱ به‌صورت جدول (۶) زیر می‌نویسیم :

جدول ۶

جمع	۶	۵	۴	۳	۲	y (تعداد افراد خانوار)
۱۸۰	۲۷	۵۵	۴۹	۲۹	۲۰	فراوانی $F_{\circ j}$

این توزیع به ما نشان می‌دهد که دانش‌آموزان این دبیرستان از نظر تعداد افراد خانوارشان چگونه توزیع شده‌اند.

توزیعهای شرطی

هرگاه بخواهیم برای اندازه معینی از صفت y، توزیع صفت x را بدانیم و یا برعکس برای اندازه یک x معین، توزیع صفت y را به‌دست آوریم. می‌توان با قرار دادن مقادیر صفت x در جدول و فراوانیهای متناظر با y معین را در برابر آنها، توزیع صفت x برحسب y معین را به‌دست آورد و آن‌را به‌صورت $X|Y=y_j$ نوشت و برای توزیع y برحسب x معین آن‌را به‌صورت $Y|X=x_i$ می‌نویسیم. به چنین توزیعهایی، توزیع شرطی x برحسب y و توزیع شرطی y برحسب x گویند. مثلاً اگر بخواهیم توزیع سن دانش‌آموزان را برای خانوارهای ۴ نفره به‌دست آوریم (با توجه به جدول ۲) خواهیم داشت:

جدول ۷

$X\|y=۴$	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	جمع
فراوانی $F_{i۴}$	۶	۱۰	۹	۱۲	۷	۵	۴۹

ملاحظه می‌شود که فراوانیها از ستون $y=۴$ در برابر مقادیر صفت x نوشته شده است.

هم‌چنین اگر توزیع نفرات خانوار برای دانش‌آموزان ۱۷ ساله موردنظر باشد، می‌توان آن‌را به‌صورت زیر به‌دست آورد:

جدول ۸

$Y\|x=۱۷$	۲	۳	۴	۵	۶	جمع
فراوانی $F_{۱۷j}$	۶	۴	۱۲	۱۰	۵	۲۷

در اینجا نیز فراوانیها از سطر $x=۱۷$ در برابر مقادیر y نوشته شده است.

از توزیعهای شرطی برای پی بردن به وجود بستگی بین دو صفت متغیر x و y استفاده می‌شود.

مشخصه‌های عددی برای توزیع مشترک دو صفت متغیر x و y

برای محاسبه میانگین صفت متغیر x و میانگین صفت متغیر y از توزیعهای حاشیه‌ای x و y استفاده نموده و مطابق طریقه محاسبه میانگین حسابی، می‌توان آنها را به‌دست آورد :

فرمولهای محاسباتی میانگینها به قرار زیر است :

$$\overline{X}=\frac{\sum_{i=۱}^{s}F_{i\circ}X_i}{n} \tag{۱}$$

$$\overline{Y}=\frac{\sum_{j=۱}^{t}F_{\circ j}Y_j}{n} \tag{۲}$$

برای مثال ۱، میانگین صفت متغیر x (سن دانش‌آموزان) و میانگین صفت متغیر y (تعداد افراد خانوار) را محاسبه می‌کنیم :

$$\overline{X}=\frac{\sum_{i=۱}^{s}F_{i\circ}X_i}{n}=\frac{۳۰\times۱۴+۳۵\times۱۵+۳۲\times۱۶+۳۷\times۱۷+۲۸\times۱۸+۱۸\times۱۹}{۱۸۰}$$

$$=\frac{۲۹۳۲}{۱۸۰}=۱۶/۲۹$$

به عبارت بهتر میانگین سن دانش‌آموزان ۱۶/۲۹ سال می‌باشد.

$$\overline{Y}=\frac{\sum_{j=۱}^{t}F_{\circ j}Y_j}{n}=\frac{۲۰\times۲+۲۹\times۳+۴۹\times۴+۵۵\times۵+۲۷\times۶}{۱۸۰}$$

$$=\frac{۷۶۰}{۱۸۰}=۴/۲۲$$

یعنی متوسط تعداد افراد خانوار دانش‌آموزان ۴/۲۲ نفر می‌باشد.

هم‌چنین می‌توان پراکندگی (واریانس) هریک از صفات متغیر x و y را توسط روابط زیر محاسبه نمود :

$$V(X)=\frac{\sum_{i=۱}^{s}F_{i\circ}X_i^۲}{n}-\overline{X}^۲ \tag{۳}$$

$$V(Y)=\frac{\sum_{j=۱}^{t}F_{\circ j}Y_j^۲}{n}-\overline{Y}^۲ \tag{۴}$$

برای مثال ۱ واریانس صفت متغیر X و واریانس صفت متغیر Y را محاسبه می‌کنیم :

$$V(X)=\frac{\sum_{i=1}^{s}F_{i\circ}X_i^2}{n}-\overline{X}^2$$

$$=\frac{۳۰\times۱۴^۲+۳۵\times۱۵^۲+۳۲\times۱۶^۲+۳۷\times۱۷^۲+۲۸\times۱۸^۲+۱۸\times۱۹^۲}{۱۸۰}-(۱۶/۲۹)^۲$$

$$=\frac{۴۸۲۱۰}{۱۸۰}-۲۶۵/۳۶=۲۶۷/۸۳-۲۶۵/۳۶=۲/۴۷$$

$$V(Y)=\frac{\sum_{j=1}^{t}F_{\circ j}Y_j^2}{n}-\overline{Y}^2$$

$$=\frac{۲۰\times۲^۲+۲۹\times۳^۲+۴۹\times۴^۲+۵۵\times۵^۲+۲۷\times۶^۲}{۱۸۰}-(۴/۲۲)^۲$$

$$=\frac{۳۴۷۲}{۱۸۰}-۱۷/۸۱=۱۹/۲۹-۱۷/۸۱=۱/۴۸$$

علاوه بر پراکندگی صفات متغیر x و y، پراکندگی توأم دو صفت متغیر x و y را نیز محاسبه می‌کنند که آن‌را کواریانس x و y نامند و به‌صورت Cov(x,y) نوشته می‌شود. فرمول محاسبه کواریانس x و y به قرار زیر است :

$$Cov(x,y)=\frac{\sum_i\sum_j F_{ij}X_iY_j}{n}-\overline{X}.\overline{Y} \qquad (۵)$$

از کواریانس x و y برای آگاهی از همبستگی بین دو صفت x و y استفاده می‌کنند.

هرگاه مقدار کواریانس x و y برابر صفر شود یعنی بین دو صفت x و y همبستگی خطی وجود ندارد و هرگاه مقدار کواریانس x و y بزرگتر از صفر (مثبت) باشد یعنی بین دو صفت x و y همبستگی به‌طور مستقیم وجود دارد. و چنانچه کوچکتر از صفر (منفی) باشد یعنی بین دو صفت x و y همبستگی به‌طور معکوس وجود دارد.

برای مثال ۱ کواریانس x و y را محاسبه می‌کنیم :

$$Cov(x,y)=\frac{\sum_i\sum_j F_{ij}X_iY_j}{n}-\overline{X}\overline{Y}$$

$$=\frac{۳\times۱۴\times۲+۴\times۱۴\times۳+۶\times۱۴\times۴+۱۲\times۱۴\times۵+۵\times۱۴\times۶+۲\times۱۵\times۲+}{}$$

$$\frac{\begin{array}{c}۷\times۱۵\times۳+۱۰\times۱۵\times۴+۱۱\times۱۵\times۵+۵\times۱۱\times۶+۴\times۱۶\times۲+۵\times۱۶\times۳+\\ ۹\times۱۶\times۴+۱۰\times۱۶\times۵+۴\times۱۶\times۶+۶\times۱۷\times۲+۴\times۱۷\times۳+۱۲\times۱۷\times۴+\\ ۱۰\times۱۷\times۵+۵\times۱۷\times۶+۳\times۱۸\times۲+۳\times۱۸\times۳+۷\times۱۸\times۴+۸\times۱۸\times۵\\ +۷\times۱۸\times۶+۲\times۱۹\times۲+۶\times۱۹\times۳+۵\times۱۹\times۴+۴\times۱۹\times۵+۱\times۱۹\times۶\end{array}}{۱۸۰}$$

$$=\frac{۱۲۳۵۲}{۱۸۰}-۱۶/۲۹\times۴/۲۲=۶۸/۶۲-۶۸/۷۴=-۰/۱۲$$

برای سهولت محاسبات می‌توان کلیه محاسبات را توسط جدول زیر انجام داد : (جدول ۹)

جدول ۹

x \ y	۲	۳	۴	۵	۶	$F_{i\circ}$	$F_{i\circ}X_i$	$F_{i\circ}X_i^۲$	$\sum_j F_{ij}Y_j$	$X_i\sum F_{ij}Y_j$
۱۴	۳	۴	۶	۱۲	۵	۳۰	۴۲۰	۵۸۸۰	۱۳۲	۱۸۴۸
۱۵	۲	۷	۱۰	۱۱	۵	۳۵	۵۲۵	۷۸۷۵	۱۵۰	۲۲۵۰
۱۶	۴	۵	۹	۱۰	۴	۳۲	۵۱۲	۸۱۹۲	۱۳۳	۲۱۲۸
۱۷	۶	۴	۱۲	۱۰	۵	۳۷	۶۲۹	۱۰۶۹۳	۱۵۲	۲۵۸۴
۱۸	۳	۳	۷	۸	۷	۲۸	۵۰۴	۹۰۷۲	۱۲۵	۲۲۵۰
۱۹	۲	۶	۵	۴	۱	۱۸	۳۴۲	۶۴۹۸	۶۸	۱۲۹۲
$F_{\circ j}$	۲۰	۲۹	۴۹	۵۵	۲۷	۱۸۰	۲۹۳۲	۴۸۲۱۰	۷۶۰	۱۲۳۵۲
$F_{\circ j}y_j$	۴۰	۸۷	۱۹۶	۲۷۵	۱۶۲	۷۶۰	←		↑	
$F_{\circ j}y_j^۲$	۸۰	۲۶۱	۷۸۴	۱۳۷۵	۹۷۲	۳۴۷۲				

محاسبه اندازه همبستگی بین دو صفت متغیر x و y و تفسیر آن‌را در جلد دوم کتاب روشهای آماری از نظر خواهید گذراند.

سؤالها و تمرینها

۱ــ توزیع مشترک دو صفت متغیر چیست؟
۲ــ هدف از مطالعه جامعه‌ها با بیش از یک صفت متغیر چیست؟
۳ــ فراوانی گروه ijام چیست؟
۴ــ توزیع حاشیه‌ای صفت متغیر x چگونه به‌دست می‌آید؟
۵ــ توزیع حاشیه‌ای صفت متغیر y چگونه به‌دست می‌آید؟
۶ــ توزیع شرطی صفت متغیر x برحسب y چه مفهومی دارد؟
۷ــ توزیع شرطی صفت متغیر y برحسب x چه مفهومی دارد؟
۸ــ از توزیع‌های شرطی برای چه منظوری استفاده می‌شود؟
۹ــ در توزیع مشترک دو صفت متغیر، میانگین حسابی دو صفت چگونه محاسبه می‌شود؟
۱۰ــ در توزیع مشترک دو صفت متغیر، پراکندگی صفات چگونه محاسبه می‌شود؟
۱۱ــ کواریانس دو صفت متغیر چیست؟ فرمول آن‌را بنویسید.
۱۲ــ از کواریانس برای آگاهی از چه خصوصیاتی از دو صفت متغیر استفاده می‌کنند؟
۱۳ــ هرگاه کواریانس دو صفت متغیر برابر صفر گردد چه مفهومی دارد؟
۱۴ــ نتایج مشاهدات دو صفت x و y در جدول زیر آمده است :

x	۸	۱۰	۱۱	۱۲	۱۴
y	۱۰	۶	۷	۳	۵

میانگین x و میانگین y را محاسبه کنید.
واریانس x و واریانس y را محاسبه کنید.
کواریانس x و y را محاسبه کنید.

۱۵ــ توزیع مشترک دو صفت متغیر x و y به‌شرح جدول زیر است :

x \ y	۳	۴	۵	۶	$F_{i\circ}$
۵	۲	۳	۳	۲	۱۰
۱۰	۳	۱۰	۱۰	۲	۲۵
۱۵	۱	۲	۷	۵	۱۵
$F_{\circ j}$	۶	۱۵	۲۰	۹	۵۰

مطلوبست :

الف ــ توزیع حاشیه‌ای x و توزیع حاشیه‌ای y

ب ــ توزیع شرطی y برحسب x = ۱۰

ج ــ توزیع شرطی x برحسب y = ۵

د ــ میانگین x و میانگین y

هـ ــ واریانس x و واریانس y

و ــ کواریانس x و y و تفسیر آن

منابع و مآخذ

۱ــ مفاهیم اساسی آمار تألیف علی مدنی ــ انتشارات فروردین.

۲ــ آنالیز آماری (جلد اول) تألیف علی مدنی.

۳ــ آمار توصیفی تألیف ژرار کالت ــ ترجمه حسن صادقی از انتشارات دانشگاه فردوسی مشهد.

۴ــ آمار و کاربرد آن در مدیریت تألیف مهدی صفاری ــ نشر دانا.

۵ــ آمار مقدماتی (جلد اول) تألیف وونا کات ــ ترجمه دکتر مشکانی.

۶ــ Statistical Methods Snedecor & Cochran.

۷ــ Statistical Analysis for Managerial Decisions Boot and Cox.

۸ــ Introduction to Mathematical Statistics Paul . G . Hoel.