

کاربرد انتسابهای چند گانه برای سوالات بدون پاسخ در بررسیهای آماری

مقدمه :

آمارشناسان به تجربه دریافته اند که در اغلب بررسیهای آماری، مواجه شدن با مشکلی به نام " بدون پاسخ " (۱) گریزناپذیر است. از طرفی تعداد بدون پاسخ های موجود در اطلاعات جمع آوری شده یکی از عوامل مهم تشخیص و تعیین کیفیت داده هاست، زیرا با وارد شدن آریبب احتمالی و کاهش تعداد نمونه های مورد بررسی به علت حذف مقادیر، بر آورد صفت مورد نظر و واریانس نمونه تحت تاثیر قرار می گیرد. به همین جهت آمارشناسان از دیرباز در جستجوی راه های مناسب جهت حل این مشکل بوده اند. اما آنچه که در این زمینه مهم می نمایاند شناخت نوع بدون پاسخ و تمایز بین آنها می باشد و به همین جهت تعیین و تفکیک در حالت زیر ضروریست :

۱ - بدون پاسخ بودن کل واحد نمونه، یعنی حالتی که در مورد هیچ یک از اقلام واحد نمونه اطلاعاتی در اختیار نیست.

۲ - بدون پاسخ بودن قلم یا اقلامی از يك واحد نمونه، یعنی حالتی که تنها در مورد يك یا چند قلم از اقلام يك واحد نمونه اطلاعی در دست نیست .

حالت اول ممکن است به علت عدم همکاری، قادر نبودن پاسخگو به همکاری، در دسترس نبودن پاسخگو و غیره رخ دهد . حالت دوم ممکن است ناشی از مواردی از قبیل عدم اطلاع پاسخگو یا حساسیت وی به ارایه پاسخ باشد . در حالت اول تنها اطلاعاتی که در دسترس است، مربوط به چارچوب نمونه گیری است . ولی در حالت دوم علاوه بر اطلاعات موجود در چارچوب نمونه گیری، پاسخ به سایر اقلام واحد نمونه نیز در اختیار می باشد .

جهت فائق آمدن بر مشکل بدون پاسخ و جبران حذف مقادیر ناشی از آن با توجه به نوع بدون پاسخ، عمدتاً در راه حل ارایه شده است : تعدیل وزنی برای کل واحد نمونه بدون پاسخ و انتساب مقادیر برای قلم یا اقلامی از واحد نمونه که بدون پاسخ مانده اند .

در این نوشتار سعی شده است که روشهای انتساب منفرد و چند گانه برای قلم یا اقلام بدون پاسخ با هم مقایسه و با ارایه يك مثال در مورد میانگین اجاره بهای واحد های مسکونی اجاری يك شهر فرضی که در بررسی نمونه ای آن تعداد ۲ واحد دارای اقلام بدون پاسخ می باشند با استفاده از روش انتساب چند گانه تحقیق به عمل می آید .

انتساب منفرد :

یکی از روشهای دستیابی به داده های کامل جهت تجزیه و تحلیل و نتیجه گیری در بررسی هایی که يك یا چند قلم از مشخصات واحد نمونه بدون پاسخ و در نتیجه بدون مقدار می باشد، روش انتساب منفرد است. بدین معنا که برای هر بدون پاسخی يك مقدار واقعی جایگزین می گردد . به عنوان مثال می توان میانگین مقدار بر مبنای به پاسخ دهندگان را جایگزین مقدار بدون پاسخ نمود و یا با

استفاده از مدل‌های رگرسیونی، مقدار حذفی ناشی از حذف و پاسخ را برآورد کرد.

بنابراین در این روش امکان تجزیه و تحلیل با استفاده از مجموعه داده‌های کامل فراهم می‌گردد و علاوه بر آن با شناخت دلایل عدم پاسخگویی، گرد آورنده داده‌ها می‌تواند برآورد نقطه‌ای مقدار منتسب شده را نیز بدست آورد. از آنجا که در این روش برای هر قلم بدو پاسخ تنها یک مقدار منتسب می‌گردد، میزان تغییرپذیری مقدار منتسب شده مشخص نیست.

انتساب چند گانه :

در انتساب چند گانه بجای هر مقدار حذفی، برداری متشکل از $M \geq 2$ مقدار ممکن جایگزین می‌گردد. بدین ترتیب هم امکان استفاده از مجموعه داده‌های کامل جهت تجزیه و تحلیل فراهم می‌آید و هم علاوه بر برآورد نقطه‌ای، میزان تغییرپذیری مقدار منتسب شده نیز قابل انعکاس است. تغییرپذیری مقدار منتسب شده می‌تواند ناشی از دو علت باشد: یکی نامشخص بودن دلایل عدم پاسخگویی و دیگری نمونه‌گیری (هر چند که دلایل عدم پاسخگویی مشخص باشد). به همین دلیل به منظور انعکاس تغییرات ناشی از نمونه‌گیری، برای هر مقدار حذفی بیش از یک مقدار منتسب می‌گردد و برای بازتاب تغییرات ناشی از عدم شناخت دلایل نیز دو یا چند مدل در نظر گرفته می‌شود. در یک مدل مفروض، هر انتساب، یک تکرار نامیده می‌شود که می‌توان برای دستیابی به برآوردی معتبرتر آنها را ادغام نمود و همچنین به منظور کشف حساسیت جوابها به دلایل بیان شده برای عدم پاسخگویی، نتایج حاصل از مدل‌های مختلف را با هم مقایسه کرد.

حال جهت تشریح بیشتر انتساب چند گانه که هدف این نوشتار است با ذکر مثالی، در مورد

میانگین اجاره بهای واحد‌های مسکونی مورد نظر بررسی به عمل می‌آید.

در یک شهر فرضی تعداد ۱۳ نمونه واحد مسکونی اجاری مورد بررسی قرار می‌گیرد ($n = 13$).

هدف از این بررسی برآورد میانگین اجاره بهای واحد های مسکونی اجاری شهر (\bar{Y}) با استفاده از برآورد نسبتی می باشد. یکی از متغیرهای مرتبط با مبلغ اجاره بها، سطح زیربنای واحد مسکونی (X) است که در پرسشنامه مربوطه در دسترس است. مقدار بیرجعه آوری شده (X و Y) مربوط به ۱۲ واحد نمونه اجاری که در جدول شماره ۱ منعکس گردیده است نشان می دهد که مقدار Y واحد های شماره ۷ و ۱۲ بدین پاسخ است. ضمناً اطلاعات موجود نشان می دهد که متوسط سطح زیربنای واحد های مسکونی اجاری در این شهر ۱۰۱ مترمربع می باشد.

برای اقلام بدین پاسخ واحد های ۷ و ۱۲ د و مدل در نظر گرفته می شود و در هرمد ل دومقدار منتسب می گردد (یعنی د و تکرار در هرمد ل).

درمد ل اول فرض می شود که اختلاف بین اجاره بهای واحد های ۷ و ۱۲ که سطح زیربنای آنها به ترتیب ۱۲۵ و ۸۴ مترمربع است با اجاره بهای واحد هایی که دارای همین سطح زیربنایی باشند تصادفی است. ولی درمد ل دوم فرض می شود که این اختلاف سیستماتیک است و اجاره بهای این واحد ها به علت واقع شدن در منطقه جغرافیایی خاص حدود ۳۰ درصد بالاتر از اجاره بهای واحدهای مشابه در دیگر مناطق است.

جهت انتساب مقدار برپه واحد هایی که اقلام آن بدین پاسخ ماند است، د و واحد که تطابق نزدیکی با هر واحد دارد انتخاب می گردد ($M = 2$). فاصله تطبیق بر مبنای مقدار سطح زیربنا تعریف می شود. بدین ترتیب برای واحد شماره ۷ که سطح زیربنای آن ۱۲۵ مترمربع است، واحد های ۱۱ و ۸ با سطح زیربنایی به ترتیب برابر با ۱۲۰ و ۱۳۰ مترمربع و به طریق مشابه برای واحد شماره ۱۲ واحد های ۳ و ۱ انتخاب می گردند. جدول شماره ۲ مقدار برانتسابی تحت د و مدل را نشان می دهد.

جدول شماره ۱

مشخصات واحد های مسکونی اجاری در یک شهر فرضی

i	y	x
شماره واحد	(اجاره بهایه تومان)	(سطح زیرینایه مترمربع)
۱	۱۴۰۰	۹۰
۲	۳۵۰۰	۱۱۰
۳	۲۰۰۰	۸۰
۴	۲۰۰۰	۸۰
۵	۱۹۰۰	۷۰
۶	۲۰۰۰	۱۵۰
۷	-	۱۲۵
۸	۲۷۰۰	۱۳۰
۹	۳۴۰۰	۱۰۰
۱۰	۳۵۰۰	۱۰۰
۱۱	۲۵۰۰	۱۲۰
۱۲	-	۸۴
۱۳	۲۵۰۰	۱۲۰

جدول شماره ۲

مقادیر انتسابی در مدل انتساب چندگانه

شماره واحد بدون پاسخ	مدل ۱		مدل ۲	
	تکرار		تکرار	
	۱	۲	۱	۲
۷	۲۷۰۰	۲۵۰۰	۲۵۱۰	۲۲۵۰
۱۳	۲۰۰۰	۱۴۰۰	۲۶۰۰	۱۸۲۰

در این جدول تحت مدل ۱، انتساب مقادیر ۲۷۰۰ و ۲۵۰۰ به ۷ واحد ۷ به ترتیب با عنوان تکرار ۱ و ۲ به صورت تصادفی انجام پذیرفته و تحت مدل ۲ به مقادیر فوق ۳۰ درصد اضافه گردیده است. در مورد واحد ۱۳ نیز به طریق مشابه عمل شده است. بدین ترتیب با فرض در مدل اول و انتساب هر مدل، چهار مجموعه داده‌های کامل حاصل می‌شود که به ترتیب درجد اول ۳ الی ۶ منعکس است.

با کارگیری چهار مجموعه داده‌های کامل جدول ۳ الی ۶ و برآورد نسبتی $\frac{\bar{y}}{\bar{x}}$ با واریانس مربوطه $SE_{\bar{y}}$ که در آن \bar{x} میانگین جامعه (متوسط سطح زیربنای واحد های مسکونی اجاری شهر = ۱۰۱)، \bar{y} و \bar{x} به ترتیب میانگین های Y و X در ۱۳ واحد نمونه تصادفی و

$$SE_{\bar{y}} = \frac{\sum_{i=1}^{13} (y_i - \bar{y} \cdot \frac{\bar{y}}{\bar{x}})^2}{13(13-1)}$$

می‌باشد، چهار برآورد و واریانس مربوطه به دست می‌آید که نتایج آن درجدول شماره ۷ منعکس

شده است.

جدول شماره ۴

مدل تکرار ۲

شماره واحد	y	x
۱	۱۴۰۰	۹۰
۲	۳۵۰۰	۱۱۰
۳	۲۰۰۰	
۴	۲۰۰۰	
۵	۱۹۰۰	
۶	۴۰۰۰	
۷	۲۵۰۰	
۸	۲۷۰۰	
۹	۳۴۰۰	
۱۰	۳۵۰۰	
۱۱	۲۵۰۰	
۱۲	۱۴۰۰	
۱۳	۲۵۰۰	۱۲۰
میانگین	۲۵۶۲	۱۰۴

جدول شماره ۳

مدل تکرار ۱

شماره واحد	y	x
۱	۱۴۰۰	۹۰
۲	۳۵۰۰	۱۱۰
۳	۲۰۰۰	۸۰
۴	۲۰۰۰	۸۰
۵	۱۹۰۰	۷۰
۶	۴۰۰۰	۱۵۰
۷	۲۷۰۰	۱۲۵
۸	۲۷۰۰	۱۳۰
۹	۳۴۰۰	۱۰۰
۱۰	۳۵۰۰	۱۰۰
۱۱	۲۵۰۰	۱۲۰
۱۲	۲۰۰۰	۸۴
۱۳	۲۵۰۰	۱۲۰
میانگین	۲۶۲۳	۱۰۴

جدول شماره ۶

مدل ۲ تکرار ۲

شماره واحد	y	x
۱	۱۴۰۰	۹۰
۲	۳۵۰۰	۱۱۰
۳	۲۰۰۰	
۴	۲۰۰۰	
۵	۱۹۰۰	
۶	۲۰۰۰	
۷	۳۲۵۰	
۸	۲۷۰۰	
۹	۳۴۰۰	
۱۰	۳۵۰۰	
۱۱	۲۵۰۰	
۱۲	۱۸۲۰	
۱۳	۲۵۰۰	۱۲۰
میانگین	۲۶۵۲	۱۰۴

جدول شماره ۵

مدل ۲ تکرار ۱

شماره واحد	y	x
۱	۱۴۰۰	۹۰
۲	۳۵۰۰	۱۱۰
۳	۲۰۰۰	
۴	۲۰۰۰	
۵	۱۹۰۰	
۶	۲۰۰۰	
۷	۳۵۱۰	
۸	۲۷۰۰	
۹	۳۴۰۰	
۱۰	۳۵۰۰	
۱۱	۲۵۰۰	
۱۲	۲۶۰۰	
۱۳	۲۵۰۰	۱۲۰
میانگین	۲۷۳۲	۱۰۴

جدول شماره ۷

برآورد و واریانس های مربوطه

برآورد و واریانس	مدل ۱		مدل ۲	
	تکرار		تکرار	
	۱	۲	۱	۲
برآورد $(\bar{X} \cdot \frac{\bar{Y}}{\bar{X}})$	۲۵۴۷	۲۴۸۸	۲۶۵۳	۲۵۷۵
واریانس (SE^2)	۴۷-۴۰	۵۳۸۵۳	۴۸۶۱۱	۵۱۱۰۶

اکنون برآورد های تحت هر مدل را می توان جهت رسیدن به يك نتیجه کلی برای میانگین اجاره بهای واحد های مسکونی اجاری (\bar{Y}) شهر مفروض با هم ادغام نمود. جدول شماره ۸ نتایج را نشان می دهد با این توضیح که در هر مدل برآورد حاصل، از متوسط برآورد ها به دست آمده است و کل واریانس از دو جزء متوسط واریانسهای داخل انتسابها و واریانس بین انتسابها تشکیل و با استفاده از فرمول ذیل محاسبه شده است:

$$(\text{واریانس بین انتسابها}) \times (1 + M^{-1}) + (\text{متوسط واریانسها}) = \text{کل واریانس}$$

بدیهی است که در اینجا $M = 2$ (تعداد تکرارها) می باشد.

جدول شماره ۸

برآورد و واریانس های ادغامی

برآورد و واریانسها	مدل ۱	مدل ۲
برآورد	۲۵۱۸	۲۶۱۴
متوسط واریانسهای داخل انتسابها	۵۰۴۴۶	۴۹۸۵۸
واریانس بین انتسابها	۱۷۴۱	۳۰۴۲
کل واریانس	۵۳۰۵۸	۵۴۴۲۱

در این جدول، به عنوان مثال، ارقام تحت مدل ۱ چنین محاسبه گردیده است:

$$\text{برآورد} = \frac{۲۵۲۷ + ۲۴۸۸}{۲} = ۲۵۱۸$$

$$\text{متوسط واریانسهای داخل انتسابها} = \frac{۴۷۰۴۰ + ۵۳۸۵۲}{۲} = ۵۰۴۴۶$$

$$\text{واریانس بین انتسابها} = \frac{(۲۵۲۷ - ۲۵۱۸)^2 + (۲۴۸۸ - ۲۵۱۸)^2}{۲ - ۱} = ۱۷۴۱$$

$$\text{کل واریانس} = ۵۰۴۴۶ + (۱ + ۲^{-1})(۱۷۴۱) = ۵۰۴۴۶ + \frac{۳}{۲}(۱۷۴۱) = ۵۳۰۵۸$$

فاصله اطمینان ۹۵ درصد برای \bar{Y} در مدل ۱ برابر است با (۲۹۶۹ و ۲۰۶۶) و در مدل ۲ برابر

است با (۳۰۷۱ و ۲۱۵۷).

بدین ترتیب در انتساب چندگانه با فرض برقراری شرایط مناسب با وجود ارقام بدون پاسخ

علاوه بر برآورد نسبتاً دقیق، فاصله اطمینان مناسبی نیز قابل ارایه می‌باشد.