تکنیک‌های آمــار اســتنباطی

آمار استنباطی شامل یک سری روش‌های آماری است که بر اساس اطلاعات نمونه، درباره ویژگی های جمعیت، پیشگویی‌هایی را فراهم می‌کند. تمرکز اولیه اکثر پژوهش‌ها بر روی پارامترهای جمعیت تحت مطالعه است. نمونه و آماره‌های توصیف‌کننده آن تنها تا جایی اهمیت دارند که اطلاعاتی را در مورد پارامترهای جمعیت فراهم کنند. بنابراین، یک جنبه مهم استنباط آماری، گزارش از صحّت احتمالی یا درجه اطمینان از آماره نمونه‌ای است که مقدار پارامتر جمعیت را پیشگویی می‌کند.

همان‌طور که ذکر شد از آمار استنباطی برای استنتاج الگوها و خصوصیات جمعیّت از خصوصیات نمونه استفاده می‌شود. **رایج‌ترین تکنیک‌های استنباطی در تحلیل یک متغیره، برآوردهای فاصله‌ای (خطای استاندارد) و در تحلیل دو یا چندمتغیره بهره گیری از سطح معنی‌داری (سطح خطا) است.**

۱) برآوردهای فاصله‌ای یا خطای استاندارد

اگر میانگین درآمد در نمونه‌ای ۱۸۰۰۰ دلار باشد میانگین درآمد در جمعیّت چقدر خواهد بود؟ چون بعید است که نمونه بازتاب کامل جمعیّت باشد(خطای نمونه‌گیری)، نمی‌توان فقط با استفاده از میانگین نمونه (که برآورد نمونه خوانده می‌شود) میانگین واقعی درآمد جمعیّت را (که پارامتر جمعیّت خوانده می‌شود) پیدا کرد.

لازم است راهی برای برآورد میزان دقت احتمالی برآورد نمونه پیدا کرد. چنانچه نمونه ما نمونه‌ای تصادفی باشد نظریه احتمالات چنین راه‌حلی در اختیار ما قرار می‌دهد. اگر تعداد زیادی نمونه تصادفی مختلف داشته باشیم، اکثر آن‌ها تقریبا نزدیک به پارامتر جمعیّت خواهند بود: میانگین اکثر آن‌ها نزدیک به میانگین واقعی جمعیّت خواهد بود و فقط معدودی از آن‌ها تفاوت زیادی با میانگین جمعیّت خواهند داشت. در واقع توزیع برآورد نمونه‌ها به توزیع نرمال نزدیک می‌شود.

مسأله این جاست که اگر فقط یک نمونه داشته باشیم (که دارای تعدادی عضو است) چگونه بدانیم میانگین نمونه ما چقدر به میانگین حقیقی جمعیّت نزدیک است. ممکن است نمونه ما یکی از نمونه‌های پرت باشد، اما بیشتر محتمل است که یکی از نمونه‌های نزدیک به میانگین حقیقی جمعیّت باشد(چون اکثر نمونه‌ها نزدیک به آن هستند). اما حتی در این صورت تا چه حد به آن نزدیک است؟ **برای برآورد میزان نزدیکی میانگین نمونه به میانگین جمعیّت از آماره‌ای به نام خطای استاندارد میانگین استفاده می‌کنیم** که فرمول آن بدین شرح است:

$$S_m = \frac{S}{\sqrt{N}}$$

که در آن S_m خطای استاندارد میانگین، S انحراف استاندارد و N تعداد کل نمونه است.

پس از محاسبه خطای استاندارد نظریه احتمالات به ما می گوید که میانگین جمعیّت احتمالا در چه دامنه‌ای از میانگین نمونه قرار دارد. طبق نظریه احتمالات در ۹۵ درصد نمونه‌ها، میانگین جمعیّت در محدوده ۲± واحد خطای استاندارد از میانگین نمونه قرار دارد. به عبارت دیگر به احتمال ۹۵ درصد میانگین جمعیّت در محدوده ۲± خطای استاندارد از میانگینِ نمونه ما قرار دارد. بنابراین می‌توانیم برآورد کنیم که میانگین جمعیّت در چه دامنه‌ای قرار دارد. به این دامنه **فاصله اطمینان** گفته می‌شود و میزان قطعیتِ قرار گرفتن میانگین جمعیّت در این فاصله، **سطح اطمینان** خوانده می-شود. رقمی که برای خطای استاندارد به دست می‌آوریم همواره برحسب واحد متغیر مورد بررسی است. مثلا اگر متغیر درآمد بر حسب دلار باشد و خطای استاندارد ۱۰۰۰ باشد، بدان معناست که خطای استاندارد ۱۰۰۰ دلار است. معنای همه این‌ها به زبان ساده چیست؟ گیریم میانگین درآمد در نمونه ما ۱۸۰۰۰ دلار است و خطای استاندارد آن ۱۰۰۰ دلار. در نتیجه به احتمال ۹۵ درصد میانگین درآمد جمعیّت در محدوده ۱۶۰۰۰ تا ۲۰۰۰۰ دلار است (یعنی ۱۸۰۰۰ به اضافه و منهای دو خطای استاندارد).

اندازه خطای استاندارد تابع حجم نمونه است. بنابراین برای این که بتوان میانگین جمعیّت را در محدوده کوچکتری برآورد کرد؛ یعنی برای این که بتوان فاصله اطمینان را کاهش داد باید خطای استاندارد را کم کرد. بدین منظور باید بر حجم نمونه افزود: با چهار برابر شدن حجم نمونه، خطای استاندارد نصف می‌شود.

۲) سطح معنی داری

بعد از این که مشخص شد دو متغیر با هم رابطه دارند باید دید رابطه‌ای که در نمونه مشاهده شده است به همان شدّت در جمعیّتی که نمونه از آن گرفته شده است نیز وجود دارد یا خیر: باید دید می‌توان نتایج حاصله از نمونه را با اطمینان تعمیم داد یا خیر. بدین منظور از آمار استنباطی (آزمون‌های معنی‌داری) سود می‌جوییم.

منطقی که در پس این آماره‌ها نهفته است چیست؟ معمولا ابتدا فرض می‌کنیم در جمعیّت بین دو متغیر رابطه‌ای وجود ندارد (یعنی r = ۰ = همبستگی). این فرضیه را فرضیه صفر می‌نامند. چنانچه مشاهده کردیم در نمونه بین دو متغیر رابطه‌ای وجود دارد (مثلا r = ./۴) تفاوت فرض عدم رابطه در جمعیّت با رابطه مشاهده شده در نمونه به دو صورت قابل تفسیر است:

۱- نمونه ما نمونه معرّف نیست. چه بسا باوجود استفاده از تکنیک‌های نمونه‌گیری تصادفی، بازهم نمونه ما نمونه‌ای غیر معرّف باشد. این را خطای نمونه‌گیری می‌نامند. بدین ترتیب ممکن است تفاوت بین نتیجه حاصل از نمونه (r= ./۰۴) و فرض عدم وجود رابطه در جمعیّت (r=۰) این باشد که نمونه، معرّف خوبی از جمعیّت نیست.
۲- فرض عدم رابطه در جمعیّت فرض غلطی است (فرض وجود رابطه در جمعیّت صحیح است).

اگر تفسیر اول را بپذیریم و حاصل نمونه (r= ./۰۴) را ناشی از خطای نمونه‌گیری بدانیم، آن‌گاه می گوییم بعید است که رابطه بین دو متغیر در جمعیّت از صفر تجاوز کند (یعنی فرض عدم رابطه را قبول می‌کنیم). از آن جا که معمولا هدف از مطالعه نمونه، استنتاج درباره جمعیّت است، فرض را بر آن می‌گیریم که رابطه مشاهده شده در نمونه (r= ./۰۴)، پیامد خطای نمونه‌گیری است و در نتیجه آن را معادل عدم رابطه در جمیعت (r=۰) می‌گیریم.

اما اگر تفسیر دوم را بپذیریم و فرض اولیه عدم رابطه در جمعیّت را غلط بدانیم، نتیجه حاصل از نمونه را بازتاب رابطه واقعا موجود در جمعیّت تفسیر می‌کنیم و چنان عمل می کنیم که گویی رابطه مشاهده شده در نمونه **واقعی** است و محصول نمونه غیر معرّف نیست. به بیان دیگر می‌توانیم وجود رابطه در جمعیّت را با توجه به وجود رابطه در نمونه بپذیریم.