آمار و احتمال

پایه ی یازدهم « رشته ی ریاضی فیزیک »

فصل ۴ : آمار استنباطی

تهیه کننده : جابر عامری

دبیر ریاضی شهرستان های اهواز و باوی

www.mathtower.ir

@amerimath

مهر ۱۴۰۰

درس ۱ : گردآوری اطلاعات آماری

بدیهی است که انجام هر کار آماری برای شناخت حداقل یکی از ویژگی های یک مجموعه از افراد یـا اشـیاء است. اوّلین اقدام در چنین فعالیت هایی، گردآوری اطلاعات(داده ها) پیرامون جامعه‌ی آماری (یا نمونه) مـی باشد. در فرآیند گردآوری اطلاعات به اصطلاحاتی از قبیل اصطلاحات زیـر روبـرو مـی شـویم. تعریـف ایـن اصطلاحات به صورت زیر است:

داده : داده ها واقعیت هایی درباره‌ی یک موضوع هستند که در محاسبه، استنباط یا برنامه ریزی به کار مـی روند.

مثال: وزن هر یک از نوزادانی که در ماه گذشته در شهر خاصی به دنیا آمده اند.

واحد آماری : به هر یک از افراد یا اشیاء گویند که داده های مربوط به آنها در یک بررسی آماری گردآوری می شوند.

مثال: هر کدام از نوزادان که در ماه گذشته در شهر خاصی به دنیا آمده اند.

مثال : وقتی گفته می شود که معدل علی محمدی ۱۴/۵ است. شخص علی محمـدی واحـد آمـاری و عـدد ۱۴/۵ یک داده است.

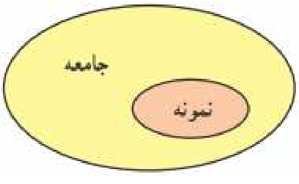
مثال : وقتی گفته می شود، تعداد دبیران ریاضی شهرستان رامهرمز ۳۴ نفر اسـت. شهرسـتان رامهرمـز واحـد آماری و عدد ۳۴ داده است.

مثال : وقتی گفته می شود که محسن محمدی از خـدمات رسـتوران گلهـا ناراضـی اسـت. شـخص محسـن محمدی واحد آماری و واژه‌ی ناراضی داده است.

جامعه‌ی آماری : مجموعه‌ی کل واحدهای آماری را جامعه‌ی آماری می نامند.

مثال : مجموعه‌ی کل نوزادانی که در ماه گذشته در شهر خاصی به دنیا آمده اند.

نمونه : هر زیر مجموعه از جامعه‌ی آماری را که با روش مشخصی انتخـاب شده باشند، یک نمونه‌ی آماری می نامند.



مثال : نوزادانی که روز های زوج ماه گذشته به دنیا آمده اند.

نمونه گیری : فرآیند انتخاب نمونه ای از یک جامعه‌ی آماری ، با منظور تعمیم اطلاعات آن به جامعه است.

مثال : فرآیند انتخاب نوزادانی که روز های زوج ماه گذشته به دنیا آمده اند.

از آنجا که جوامع آماری معمولاً از حجم و وسعت جغرافیایی زیادی برخوردارند و محققان نمی توانند به تمام آنها مراجعه کنند، بنابراین ناگزیرند به انتخاب جمعی از آنها به عنوان نمونه و تعمیم نتایج آن به جامعه‌ی مورد مطالعه اکتفا کنند. مثلاً محقق نمی تواند تمام روستاهای یک استان یا حتی یک شهرستان را مطالعه کند، یا برای مطالعه‌ی یک قشر اجتماعی مانند پزشکان نمی تواند به همه‌ی آنها مراجعه کند، یا دانش آموزان و دانشجویان یک کشور یا یک شهرستان را نمی تواند به طور کامل مطالعه کند. طبیعی است که انجام دادن تحقیقات در سطح وسیع، اقدام به شکل تمام شماری، نیازمند زمان و امکانات بسیار زیاد می باشد که عملاً مقدور نیست و کار تحقیق را غیر عملی می کند. تنها در سرشماری های نفوس و مسکن آن هم با یک بسیج عمومی از ناحیه‌ی دولت ها است که می توان به افراد جامعه مراجعه کرد.

بنابراین، محقق راه نمونه گیری را انتخاب می کند. محقق به دو شکل ممکن است نمونه را انتخاب کند. یک شکل آن ، این است که شانس انتخاب شدن را به تمامی افراد جامعه بدهد. یعنی روشی را برای انتخاب نمونه برگزیند که تمام افراد جامعه شانس مساوی برای انتخاب شدن داشته باشند. محقق برای انتخاب چنین نمونه ای با در نظر داشتن ویژگی های جامعه‌ی آماری می تواند از روش های گوناگون استفاده کند که به آن اشاره خواهد شد. این روش انتخاب نمونه را روش انتخاب **احتمالی** یا **اتفاقی** می گویند که از آن به روش تصادفی نیز تعبیر می شود. رعایت اصل احتمال و شانس انتخاب (مساوی یا غیر مساوی ولی غیر صفر) برای هر یک از افراد جامعه باعث می شود که نمونه‌ی منتخب معرف جامعه باشد و از ارزش علمی برخوردار بوده و صفات آن با صفات جامعه همخوانی و یکنواختی داشته باشد. به این ترتیب، می توان شاخص های محاسبه شده را بر اساس روشهای خاصی به کل جامعه تعمیم داد و پارامترهای آن را برآورد کرد.

روش دیگر برای انتخاب نمونه از بین واحدهای جامعه، **روش وضعی** و **غیر احتمالی** است. یعنی همه‌ی واحدها جامعه شانس مساوی برای انتخاب شدن به عنوان عضو نمونه را ندارند و در انتخاب واحدها در نمونه، محقق نظریات خود را دخالت می دهد و به واحدهای خاصی در جامعه چشم می دوزد. نتایج به دست آمده از چنین مطالعه ای قابلیت تعمیم ندارد. یعنی محقق نمی تواند مقادیر محاسبه شده برای صفات یا شاخص های

نمونه را به جامعه‌ی آماری تعمیم دهد و پارامترهای جامعه را از آن استنباط نماید. اینگونه نمونه ها را نمونه های **تورش دار** یا **اریب**[1] نیز می گویند.

بنابراین محقق وقتی قصد شناخت جامعه‌ی آماری را از طریق مطالعه‌ی نمونه ای از آن به بهترین شکل را داشته باشد، باید حتماً روش اوّل را برای گزینش نمونه انتخاب کند تا نمونه از این طریق با جامعه مشابهت داشته باشد و صفات آنها را بر یکدیگر مطابقت نماید و نمونه معرّف جامعه باشد.

با توجه به آنچه که اشاره شد. می توان تعاریف زیر را بیان کرد.

***

نمونه گیری احتمالی : نوعی نمونه گیری است که همه‌ی واحدهای آماری شانسی معلوم ولی غیر صفر برای انتخاب در نمونه داشته باشند. نمونه گیری احتمالی به روشهای مختلفی انجام می گیرد. مهمترین این روش ها به شرح زیر هستند.

الف : نمونه گیری تصادفی ساده

نوعی روش نمونه گیری، که در آن همه ی واحدهای آماری برای انتخاب شدن در نمونه و نیز برای گزینش هر یک از نمونه های ممکن، شانس یکسان دارند. این روش به شکل های مختلف انجام می شود. دو روش آن به صورت زیر است:

۱: استفاده از قرئه کشی

در این روش محقق به هریک از افراد جامعه یک کُد یا شماره‌ی مخصوص می دهد. سپس از مهره ها یا پلاک های شماره دار استفاده می کند در صورت نبود آن ، شماره‌ی هر یک از آنها روی کاغذی و یا مقوای کوچکی یادداشت می نماید. بنابراین، به تعداد افرد جامعه ، مهره یا پلاک یا کاغذ شماره دار در اختیار خواهد داشت. آنها را در داخل کیسه یا ظرفی می ریزد و به هم می زند. سپس مهره ها را یکی یکی خارج کرده ، شماره‌ی آنها را یادداشت می نماید و این کار را آنقدر ادامه می دهد تا به تعداد حجم نمونه برگزینند. آنگاه که تعداد افراد نمونه کامل شد، کار قرئه کشی به پایان رسیده، مطابق لیست، افراد خود را شناسایی می کند.

[1]. نمونه ای که انتخاب اعضای آن احتمالی نباشد یا اندازه‌ی کافی نداشته باشد، نمی تواند برآورد دقیقی از جامعه بدهد، چنین نمونه هایی را اریب یا تورش دار می نامند. اگر نمونه ای اریب نباشد، آن را نااریب می گویند.

۲: استفاده از جدول اعداد تصادفی

در این حالت از اعداد تصادفی برای انتخاب اعضای نمونه استفاده می شود. اعداد تصادفی را یا می توان از جداول اعداد تصادفی که در کتاب های مختلف آماری ، معمولاً درج شده اند، یا از ماشین حساب و یا رایانه آنها را تولید کرد.

مثال : یک نمونه‌ی تصادفی۴ تایی از یک جامعه‌ی ۱۰ نفره انتخاب کنید.

ابتدا اعضای جامعه را کد گذاری می کنیم.

نام	علی	حسن	مرتضی	زهرا	محمّد	فاطمه	مریم	مهدی	رضا	زینب
کد	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰

سپس نمونه گیری را انجام می دهیم.

مرحله	فشار دادن کلید های RAN و SHIFT	ضرب در عدد ۱۰	شماره‌ی عضو در نمونه (رُند کردن)
اوّل	۰/۳۵۲	۳/۵۲	۴
دوّم	۰/۴۵۲	۴/۵۲	۵
سوّم	۰/۸۲۵	۸/۲۵	۹
چهارم	۰/۴۸۸	۴/۸۸	–
پنجم	۰/۶۹۱	۶/۹۱	۷

به این ترتیب ۴ عضو نمونه مشخص شدند. این اعضا عبارتند از :

۱: زهرا　　۲: محمّد　　۳: رضا　　۴: مریم

ب : نمونه گیری طبقه ای

نوعی روش نمونه گیری که در آن با طبقه بندی، جامعه به زیر جامعه های مجزّا و همگن تقسیم می شود، سپس از هر طبقه یک نمونه‌ی تصادفی انتخاب می شود.

مثال: دانش آموزان پایه‌ی دهم در رشته تجربی یک روستا به دو طبقه دانش آموزان پسر و دانش آموزان دختر تقسیم می شوند. تعداد آنها طبق جدول زیر است. یک نمونه‌ی۱۰ نفره از این جامعه تعیین کنید.

تعداد کل	دانش آموزان پسر	دانش آموزان دختر
۱۰۰	۴۰	۶۰

ابتدا نسبت افراد نمونه از هر طبقه را تعیین می کنیم.

$$\frac{۱۰}{۱۰۰} = ۰/۱$$

تعداد کل	دانش آموزان پسر	دانش آموزان دختر
۱۰	$۴۰ \times ۰/۱ = ۴$	$۶۰ \times ۰/۱ = ۶$

حال از هر طبقه بطور مجزا به یکی از روش های تصادفی ساده یا منظم نمونه را انتخاب می‌کنیم.

توجه :

۱: نسبت افراد جامعه در هر طبقه آن باید در نمونه‌گیری مورد توجه قرار گیرد.

۲ : بدیهی است که اگر تعداد افراد طبقات جامعه مساوی باشند، نسبت افراد در نمونه برابرخواهد شد.

ج : نمونه گیری سامانمند یا سیستماتیک(منظم)

نوعی روش نمونه گیری طبقه ای است که در آن اندازه‌ی طبقات با هم برابر است و فقط از طبقه‌ی اوّل واحد های آماری به تصادف انتخاب می شود و با همان رویه از طبقات دیگر، این کار انجام می گیرد.

در این روش همانند روش های قبل فرض بر این است که افراد جامعه متجانس هستند و از این رو به هر یک از آنها از عدد ۱ تا N شماره (کُد) داده می شود. سپس افراد نمونه با نظمی خاص انتخاب می شوند. این روش نیز ساده است و محققان غالباً از آن استفاده می کنند. در روش منظم محقق سعی می کند، فاصله‌ی عددی دو نمونه را به طور ثابت مشخص کند. آنگاه برای تعیین کد یا شماره‌ی اوّلین فرد نمونه و مشخص کردن موقعیت آن در سلسله‌ی اعداد و نیز موقعیت سایر افراد نمونه می توان با افزودن یا کاستن عدد ثابت فاصله، اقدام کند. توجه داشته باشید که برای تعیین موقعیت اوّلین نمونه از روش انتخاب اتفاقی یا احتمالی ساده استفاده می شود. مثلاً بین اعداد ۱ تا ۹ را قرعه کشی کرد و یک عدد را انتخاب نمود. این عدد معرف اوّلین عضو نمونه خواهد بود. با اضافه کردن فاصله‌ی مورد نظر به کد اوّلین عضو کد عضو بعدی به دست می آید. این عمل تا انتخاب آخرین عضو ادامه پیدا می کند. به مثال زیر توجه کنید.

مثال : محققی می خواهد از بین افراد یک جامعه‌ی دانشجویی ۵۰۰ نفری نمونه ای به تعداد ۵۰ نفر را به روش منظم یا سیستماتیک انتخاب کند. برای این کار پس از کُدگذاری، ابتدا عدد ثابت k را محاسبه می کند.

$$k = \frac{N}{n} = \frac{۵۰۰}{۵۰} = ۱۰$$

سپس به روش قرئه کشی موقعیت اوّلین فرد نمونه را بین اعداد ۱ تا ۹ را مشخص می نماید. فرض کنید که این عدد ۶ انتخاب شود. عدد۶ کد اوّلین عضو نمونه است.کد اعضای دیگر به ترتیب زیر است.

۴۹۶ و .... و۵۶ و۴۶ و۳۶ و ۲۶ و ۱۶ و ۶

روش نمونه گیری منظم، باعث می شود تا افراد نمونه به طور یکنواخت در سراسر جامعه پراکنده شوند.

مثال : مدیر شهربازی اعلام می کند که در روز های گذشته حدود ۵۰۰ نفر از این شهر بازی دیدن کردند. برای تعیین میزان رضایت مشتریان، قرار است به میزان۵۰ فرم پرسش نامه توزیع شود. لذا با توجه به مثال فوق پرسش نامه به افراد دارای کُدهای تعیین شده (با توجه به قبض ورودی) انتخاب می شوند.

د : نمونه گیری خوشه ای

نوعی نمونه گیری که در آن، واحدهای نمونه گیری اوّلیه در جامعه، گروه ها یا خوشه ها می باشند که تفاوت داده ها درون خوشه ها بیشتر از بین خوشه ها است.

مثال: محققی می خواهد درباره‌ی معلمان یا دانش آموزان تحقیقی انجام دهد. با توجه به الگوی تقسیمات کشوری او باید در مرحله‌ی اوّل در سطح استان ها نمونه گیری کند و استان های مورد نظر را انتخاب نماید. در مرحله‌ی دوّم در بین استان های برگزیده شده، نمونه گیری را در سطح شهرستان انجام دهد. در مرحله‌ی سوّم باید بین شهرهای انتخاب شده ، تعدادی منطقه‌ی شهری را انتخاب نماید.

از آنجا که واحد تحلیل در تحقیق معلمان است. محقق باید چهارمین مرحله‌ی نمونه گیری را انجام دهد و در سطح شهر اقدام به انتخاب چند مدرسه نماید و در نهایت واحد تحلیل و مطالعه‌ی خود را برگزیند.

همانطور که گفته شد، در هر یک از مراحل نمونه گیری در هر یک از طبقات باید افراد نمونه را از روش احتمالی ساده یا منظم برگزیده شوند.

توجه : جدول زیر را مزایا و معایب روشهای نمونه گیری احتمالی را نشان می دهد.

روش نمونه گیری	مزایا	معایب
تصادفی ساده	همه‌ی اعضا شانس مساوی برای انتخاب دارند.	هزینه بر بودن و عدم دسترسی مناسب به واحدهای آماری در جامعه های بزرگ
خوشه ای	کم هزینه، سریع	شانس انتخاب واحد ها متفاوت است.
طبقه ای	اطمینان از انتخاب عضو از همه‌ی طبقات	زمان بر و هزینه بر همچنین شانس انتخاب واحدها متفاوت است.
سیستماتیک	شانس انتخاب یکسان اعضا و بهینه در صرف زمان و هزینه	بدون داشتن فهرستی از واحدی های آماری امکان انجام وجود ندارد.

***

نمونه گیری غیر احتمالی : نوعی نمونه گیری است که در آن برخی از واحد های آماری هیچ شانسی برای انتخاب شدن در نمونه را نداشته باشند. برخی از روش های نمونه گیری های غیر احتمالی عبارتند از:

۱: نمونه گیری در دسترس(اتفاقی): فقط از واحدهای آماری که در دسترس هستند،استفاده می کنیم.

مثال: این روش یکی از ساده ترین روش ها است.یعنی اینکه افرادی مورد مطالعه قرار می گیرند که در دسترس قرار دارند و مصاحبه گر در چهارچوب تعداد و حجم نمونه در مکانهای خاصی می ایستد و با هرکس از راه رسید مصاحبه می کند. مانند خبرنگاران رسانه ها، این مکان ها ممکن است، ایستگاه اتوبوس، بازار، فروشگاه بزرگ، سلف سرویس، رستوران، زیارتگاه، ورودی شهر و ... باشد.این نمونه ها را نمی توان تعمیم داد.

مثال: در فرودگاه مستقر می شویم و در مورد سفرهای نوروزی نظرخواهی می کنیم.

۲: نمونه گیری قضاوتی : فقط از گروه خاصی از جامعه، نمونه بگیریم.

مثال: گاهی اوقات محقق بر اساس تجربه شخصی یا تجارب تکراری و مشابه دیگران یک گروه اجتماعی را معرف جامعه ای که به آن تعلق دارند می یابد. مثلاً در بعد افکار و عقاید، آنها را منعکس کننده افکار جامعه می یابد و مورد مطالعه قرار می دهد. یا اینکه آموزگاری با شناخت خود از دانش آموزان کلاس درس خود، دانش آموزان ردیف آخر را انتخاب می کند.

مثال: فقط از قشر فرهنگی پیرامون رضایت شغلی (به نمایندگی کل کارمندان دولت) مصاحبه می کنیم.

۳: نمونه گیری سهمیه ای : برخی از اعضاء جامعه و یا بخشهای آن، سهم مشخصی در نمونه داشته باشند.

مثال: در این روش تعداد نمونه ها مشخص می شود و به همراه دستور العمل مصاحبه تحویل پرسشگر می گردد تا شخصاً به میدان بررسی رفته و خودش افراد نمونه را با توجه به تعدادی که به وی داده شده انتخاب کند و از طریق مصاحبه‌ی با آنها اطلاعات لازم را گردآوری نماید. این روش هر چند مورد حمایت عده ای قرار گرفته، چون در آن اصل شانس برابر برای کلیه‌ی افراد جامعه رعایت نمی شود، ارزش علمی مطلوب ندارد و نمی توان به تعمیم نتایج آن اعتماد کرد. مانند نمونه گیری برای پیش بینی اینکه کدام نامزد انتخابات رأی می آورد.

مثال: برای بررسی میزان رأی نامزدهای انتخابات قبل از رأی گیری، با اطلاع قبلی در نزدیکی ستاد یکی از نامزدهای انتخابات مستقر و به مدت ۲ ساعت پرسش می کنیم ولی در نزدیکی ستاد دیگر فقط ۱۵ دقیقه پرسش می کنیم.

***

تمرین۱ : فرض کنید جامعه‌ای از ۱۰۰ عضو تشکیل شده است و می خواهیم نمونه ای به اندازه‌ی ۲۰ نفر از آن انتخاب کنیم. در هر یک از حالت های زیر، احتمال انتخاب هر عضو جامعه به عنوان نمونه چقدر است؟ نام هر روش نمونه گیری را بنویسید.

الف: اگر جامعه به دو قسمت ۵۰ تایی تقسیم شود و بخواهیم از هر قسمت، نمونه‌ی تصادفی ۱۰ تایی انتخاب کنیم.

ب : اگر جامعه به تصادف به ۱۰ قسمت مساوی تقسیم شود و دو قسمت را به عنوان نمونه انتخاب کنیم.

پ : اگر جامعه به تصادف به ۲۰ قسمت مساوی تقسیم شود و از قسمت اوّل یک عضو به تصادف انتخاب شود. فرض کنید عضو انتخابی دوّمین عضو باشد و از قسمت های بعدی نیز دوّمین عضو انتخاب شود.

حل :

	نام روش	احتمال
الف	طبقه ای	$\frac{۱۰}{۵۰}=\frac{۱}{۵}$
ب	خوشه ای	$\frac{۱۰}{۱۰۰}=\frac{۱}{۱۰}$
پ	سیستماتیک	احتمال اوّلی $\frac{۱}{۵}$ و بقیه ۱ است.

تمرین۲ : برای بررسی وضع سلامت دهان و دندان افراد در یک روز ، صد نفر اوّل که به ترمینال غرب تهران (به عنوان نزدیکترین نقطه به محل زندگی محقق) می آیند را مورد کنترل و پرسش قرار می دهیم. به نظر شما کدام روش نمونه گیری در اینجا استفاده شده است.

حل : نمونه گیری غیر احتمالی و از روش در دسترس استفاده شده است. زیر ا همه‌ی اعضا شانس مساوی در انتخاب ندارند.

تمرین۳ : در هر مثال دلیل اُریبی روش نمونه گیری استفاده شده را ذکر کنید.

الف : نمونه گیری از اعضای تیم ملّی والیبال برای بررسی درآمد مالی والیبالیست ها در ایران

ب : نمونه گیری از اهالی نزدیک ترین محله در تهران برای بررسی سطح رفاه اجتماعی در این شهر

حل :

الف : انتخاب بخش و گروه خاصی از جامعه‌ی والیبال ایران و عدم وجود شانس انتخاب برای واحدهای آماری

ب : فقط از افراد در دسترس استفاده شده است و همه‌ی اعضا شانس انتخاب ندارند.

تمرین۴ : آیا در نمونه گیری خوشه ای، احتمال انتخاب واحدهای آماری برای عضویت در نمونه برابر است؟ چرا؟

جواب: در روش خوشه ای، ابتدا جامعه به گروه ها یا خوشه هایی تقسیم می شوند. سپس از بین خوشه ها تعدادی به تصادف انتخاب شده و سپس همه‌ی واحدهای آن مورد مطالعه قرار می گیرند. پس خوشه ها از شانس برابری برای انتخاب برخوردار هستند، امّا واحدهای آماری از شانس برابر برخوردار نیستند. البته اگر تعداد اعضای خوشه ها با هم برابر باشند، واحدهای آماری می توانند شانس برابر داشته باشند، امّا در حالت کلی این مورد برقرار نیست.

تمرین برای حل :

۵ : جای خالی را کامل کنید.

الف : هر زیر مجموعه از جامعه‌ی آماری که به روش مشخصی انتخاب شده باشد را یک .... می نامند.

ب : نوعی روش نمونه گیری که در آن، واحد های آماری برای انتخاب شدن در نمونه، احتمال یکسان دارند، نمونه گیری .... است.

پ : ۱۰۰ دانش آموز را به ۵ گروه مساوی تقسیم کرده و از میان این ۵ گروه ، ۲ گروه را که جمعاً ۴۰ نفر هستند به اردو می برند. در انتخاب این۴۰ نفر از نمونه گیری .... استفاده شده است.

ت: نمونه هایی که نمی توانند برآورد دقیقی از جامعه بدهند، را .... می نامند.

***

روش‌های گردآوری داده ها

داده ها واقعیت هایی درباره‌ی یک موضوع هستند که در محاسبه، استنباط یا برنامه ریزی به کار می روند.

عمل گردآوری داده ها به یکی از روشهای ممکن را **آمارگیری** و کسی کـه آمـارگیری انجـام مـی دهـد را **آمارگیر** می نامند. برای جمع آوری داده ها روش های مختلفی وجود دارد. معمول ترین این روش ها عبارتند از :

۱) مشاهده : گردآوری داده ها بدون نیاز به فرد پاسخ گو

مثال : شمارش تعداد وسایل نقلیه‌ی عبوری از یک تقاطع در هر ساعت

۲) پرسش از واحد های آماری

کتبی(پرسش نامه): مجموعه‌ی سئوال های از پیش تعیین شده که توسط تعدادی پاسخ دهنده تکمیل می شود.

مثال: پرسش نامه‌ی مربوط به رضایت معلمان پیرامون یک دوره‌ی ضمن خدمت که برگزار شده است.

شفاهی(مصاحبه): حالت خاصی از پرسش نامه که به صورت شفاهی صورت می گیرد.

مثال: مصاحبه پیرامون راه حل مناسب برای بر طرف نمودن مشکل ترافیک شهر

۳) دادگان : استفاده از داده های از پیش تهیه شده که در گذشته جمع آوری شده اند و می توان از آن هـا در تحقیقات بعدی بهره برد.

مثال : بررسی روند تغییر نمرات ریاضی یک دبیرستان از ده سال گذشته تاکنون

۴) آزمایش : جمع آوری اطلاعات از واحدهای آماری با انجام یکسری آزمایش روی آنها

این روش برای علوم تجربی بسیار مرسوم است.

مثال : بررسی اثر نور خورشید بر پوست انسان

تمرین۶ : بهترین روش جمع آوری داده ها را برای موارد زیر بیان کنید.

۱) میزان رضایت مشتریان بانک از نحوه‌ی برخورد و رسیدگی به درخواست های آنها

۲) سن همه‌ی دانش آموزان بر حسب ماه در پایه‌ی دهم

۳) تعداد سرنشینان خودروهای سواری در یکی از محورهای خروجی شهر

۴) بررسی رشد جمعیت در ۵۰ سال گذشته‌ی جهان

۵) بررسی سطح کیفی خدمات بانکی خاص در یک محله

۶) بررسی شرایط متقاضیان وام در یک بانک

۷) بررسی تعداد تصادف در خیابانی در مشهد

۸) بررسی میزان چربی خون افراد در سنین نوجوانی

حل :

۱) پرسش نامه ۲) دادگان ۳) مشاهده ۴) دادگان ۵) مصاحبه ۶) پرسش نامه ۷) مشاهده ۸) آزمایش

***

آمار استنباطی

هر ویژگی از اشخاص یا اشیاء که قرار است بررسی شود را **متغیر** می نامند. ویژگی اصلی متغیر این است که از هر عضو جامعه یا نمونه به عضو دیگر تغییر می کند. قد دانش آموزان و همچنین میوه‌ی مورد علاقه‌ی آنها نمونه هایی از متغیر ها هستند. متغیرها دارای دو نوع اصلی می باشد، کمی و کیفی

متغیر کمی : متغیرهایی هستند که مقادیر عددی می گیرند و برای آنها عملیات ریاضی از قبیل جمع ، تفریق و معدل گیری قابل انجام است.

مثال : سن ، نمره ، تعداد افراد خانواده ، طول مکالمه ی تلفنی و ....

متغیر کیفی : متغیر هایی هستند که صرفاً برای دسته بندی افراد یا اشیاء در گروه ها به کار می روند و لزوماً مقدار عددی نمی گیرند.

مثال : جنسیت (مرد و زن) ، کیفیت کالا (مرغوب، متوسط ، نامرغوب)

تمرین برای حل :

۷ : جای خالی را کامل کنید.

الف : به هر یک از ویژگی هایی که مورد پژوهش قرار می گیرند، .... می گویند.

ب : میزان بارندگی یک متغیر .... و گروه خونی یک متغیر .... است.

۸ : یک متغیر کیفی و یک متغیر کمی برای کتب های درسی بیان کنید.

شاخص های نمونه و جامعه

بعد از جمع آوری اطلاعات از جامعه یا نمونه می توان، شاخص هایی در مورد آنها را محاسبه کرد. بستگی به این که این شاخص ها از جامعه یا از نمونه به دست آمده باشند به دو دسته‌ی **پارامتر** و **آماره**، تقسیم می شوند.

هر مشخصه‌ی عددی که توصیف کننده‌ی جنبه‌ی خاصی از جامعه است را **پارامتر** می نامند. در محاسبه‌ی پارامتر ، داده های تمامی اعضای جامعه در نظر گرفته می شوند.

مانند:

الف : نسبت کل معلمان ورزشکار شهرستان به کل معلّمان این شهرستان

ب : میانگین سن کل معلمان ورزشکار شهرستان

هر مشخصه‌ی عددی که توصیف کننده‌ی جنبه ای خاص از نمونه باشد، را **آماره** می نامند. در محاسبه‌ی آماره فقط داده های اعضای نمونه شرکت داده می شوند.

مانند :

الف : نسبت معلمان ورزشکار در بین معلّمان دو دبیرستان شهرستان که به تصادف انتخاب می شوند.

ب : میانگین سن معلمان ورزشکار در بین معلّمان دو دبیرستان شهرستان که به تصادف انتخاب می شوند.

پارامترها و آماره های مهم که در آمار مورد استفاده قرار می گیرند، به شرح جدول زیر هستند.

شاخص	گروه	نماد کلی	نسبت	میانگین	واریانس
آماره	نمونه	$\hat{\theta}$	p	$\bar{x}$	$S^۲$
پارامتر	جامعه	θ	P	μ	$\sigma^۲$

توجه : چون در محاسبه‌ی پارامتر ها تمامی داده های جامعه در نظر گرفته می شوند، لذا پارامترها در جامعه ثابت هستند. ولی آماره ها در نمونه های مختلف از یک جامعه، متفاوت می باشند. علاوه بر این آماره های مختلف از یک جامعه ممکن است بیشتر و ممکن از کمتر از پارامتر باشند. بدیهی است که هر چه نمونه به صورت تصادفی انتخاب شود و هر چه حجم نمونه به حجم جامعه نزدیکتر شود، آماره به پارامتر نزدیک می شود.

هدف اصلی ازمحاسبه‌ی آماره در نمونه، تعمیم آن به پارامتر جامعه است. این امر مستلزم انجام دقیق و علمی مراحلی از کار است. فرآیند نتیجه گیری درباره‌ی پارامترهای جامعه بر اساس نمونه، **آمار استنباطی** نامیده می شود.

مثال ۱ : در یک مطالعه، ۱۲۶۱ مشتری غذاهای گیاهی یک شهرستان، سئوال شده است که برای کدام وعده‌ی غذایی (ناهار یا شام) سفارش داده اند. با توجه به این موضوع به سئوالات زیر پاسخ دهید.

الف: متغیر را مشخص کنید و نوع آن را بنویسید.

جواب : متغیر ، وعده‌ی غذایی و از نوع کیفی می باشد.

ب : کدام روش گردآوری داده ها برای این مطالعه مناسب است؟

جواب : پرسش نامه

ج : جامعه‌ی آماری در این جا چیست؟ در این مطالعه پارامتر و آماره چه چیزی می توانند باشند؟

جواب : جامعه‌ی آماری مشتریان غذای گیاهی شهرستان مورد نظر است. پارامتر و آماره نسبت سفارشات نهار (یا شام) به کل سفارشات می باشند.

مثال ۲ : در بررسی میانگین درآمد خانوارهای کشور، از هر استان، ۲۰ خانواده انتخاب کرده و میانگین آنها را به دست می آوریم. با توجه به این مطالعه سئوالات زیر را پاسخ دهید.

الف : متغیر مورد مطالعه و نوع آن را بنویسید؟

جواب : درآمد خانوارهای کشور ، کمی

ب : در این مطالعه ، نمونه و حجم آن را تعیین کنید.

جواب : خانوارهای تعیین شده ، حجم آن $۳۱ \times ۲۰ = ۶۲۰$ خانوار است. (۳۱ تعداد استان های کشور)

ج : پارامتر و آماره را بنویسید.

جواب : میانگین در آمد خانواره های کشور ، پارامتر و میانگین درآمد خانوارهای خانواده های انتخاب شده، آماره است.

مثال ۳ : برای تعیین میانگین درآمد خانوارهای یک روستا در استان کرمان ،۶ خانوار به تصادف با درآمدهای ۳/۵ و ۱/۵ و ۳ و ۵ و ۴ و ۱ میلیون تومان انتخاب شده اند. آماره‌ی میانگین نمونه را به دست آورید.

$$\overline{x}=\frac{\sum x_i}{n}=\frac{۱+۴+۵+۳+۱/۵+۳/۵}{۶}=۳$$

تمرین برای حل :

۹ : جای خالی را کامل کنید.

الف : در یک جامعه‌ی آماری .... ثابت است ولی آماره بین نمونه های مختلف جامعه .... می کند.

ب : هرچه نمونه، تصادفی انتخاب شده و به اندازه‌ی کافی بزرگ باشد، در این صورت ...... به پارامتر نزدیک می شود.

پ : فرآیند نتیجه گیری درباره‌ی پارامترهای جامعه بر اساس نمونه، .... نامیده می شود.

ت : تفاوت در نوع و کاربردهای آماره و پارامتر ، موجب تقسیم بندی آمار به آمـار توصـیفی و آمـار ..... شـده است. آمار توصیفی به خلاصه کردن داده ها می پردازد و هدف آن بررسی داده های گردآوری شده است ولی آمار استنباطی به تجزیه و تحلیل نمونه و ...... آن به جامعه می پردازد.

***

تهیه کننده : جابر عامری

عضو گروه ریاضی دوره ی دوم متوسطه

استان خوزستان

درس۲ : برآورد (تخمین)

علم آمار بر اساس میزان اطلاعات در دسترس از نمونه یا جامعه به دو بخش آمار توصیفی و آمار استنباطی تقسیم می شود. آمار توصیفی به خلاصه کردن داده های نمونه می پردازد. به کمک اطلاعات این نمونه ها، می توانیم به برآوردی درست از ویژگی های جامعه برسیم. در این حالت از آمار استنباطی به تجزیه و تحلیل نمونه و بررسی ویژگی های جامعه می پردازد.

فرض کنید می خواهیم میزان رضایت افراد جامعه از زندگیشان را بررسی کنیم. برای بررسی لازم هست از پارامترهای مرکزی و پراکندگی در جامعه استفاده کنیم.اگر اطلاعات همه‌ی افراد جامعه را داشته باشیم، می توان از پارامترها استفاده کرد. امّا در اغلب موارد اطلاعات کامل جامعه را نداریم. پس مجبوریم از نمونه گیری تصادفی استفاده کنیم. پس از نمونه گیری به کمک آماره‌ی حاصل از نمونه، می توان تخمینی از مقادیر پارامتر جامعه‌ی مربوط به آن داشت. بر این اساس می توان به طور خلاصه نوشت:

آمار توصیفی : خلاصه کردن داده ها

آمار استنباطی : تجزیه و تحلیل نمونه و تعمیم آن به جامعه

برای تحلیل یک پارامتر مجهول در جامعه، باید نمونه ای از جامعه را انتخاب کرده و بر اساس نتایج به دست آمده از روی نمونه با انتخاب یک آماره‌ی مناسب (جایگزین پارامتر در نمونه) به بحث و نتیجه گیری در مورد پارامتر مورد نظر پرداخت. یکی از روشهای انجام این عمل « برآورد» پارامتر مجهول جامعه است. برآورد به دو شکل برآورد نقطه‌ای یا برآورد فاصله‌ای پارامتر انجام می گیرد که در ادامه به توضیح این دو می پردازیم.

الف : برآورد نقطه ای

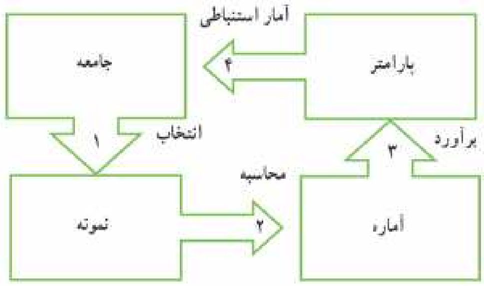
برآورد نقطه ای پارامتر جامعه برابر با مقدار عددی حاصل از جایگذاری اعداد نمونه‌ی تصادفی در آماره‌ی نظیر پارامتر است. یعنی پارمتر جامعه را برابر آماره‌ی نمونه می گیریم.

آماره = پارامتر

مثلاً : برای اگر میانگین نمونه ۱۵/۲ باشد، $\mu_x = \bar{x} = ۱۵/۲$، یعنی میانگین جامعه هم ۱۵/۲ می باشد.

واضح است که احتمال صحت برآورد نقطه ای بسیار ناچیز است و شاید بتوان گفت برابر صفر است. امّا چون برآورد نقطه‌ای از نمونه‌ی تصادفی تبعیت می کند، برای هر پارامتر مجهول ، برآوردهای مختلف به ازای

نمونه های مختلف به دست می آید. امّا توجه به این نکته مهم است که **با زیاد شدن اندازه‌ی نمونه، فاصله‌ی آماره تا پارامتر کمتر می شود**. لذا برآورد دقیق تری خواهیم داشت.



برآورد نقطه‌ی میانگین جامعه

با توجه به آنچه که در بالا گفته شد، یک روش مناسب برآورد پارامتر ، انتخاب چند نمونه‌ی مجزا از جامعه است. طوری که برای هر یک میانگین جداگانه‌ای حساب می شود. واضح است که این میانگین ها برابر نخواهند بود. به انحراف معیار میانگین های نمونه های جامعه، خطای معیار میانگین گفته می شود. خطای معیار میانگین، شاخص پراکندگی میانگین های نمونه ها از میانگین جامعه است.

هر چقدر مقدار خطای معیار میانگین، کمتر باشد، میانگین نمونه ها با خطای کمتری، میانگین جامعه را نشان می دهد و برآورد دقیق تری از میانگین جامعه خواهد داشت.

خطای معیار میانگین ، میزان برآورد از مقدار واقعی میانگین را نشان می دهد. به خطای معیار میانگین ، انحراف معیار برآورد میانگین هم گفته می شود.

آمار دانان انحراف معیار میانگین را به صورت زیر محاسبه می کنند.

$$\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{n}}$$

که در آن σ_x انحراف معیار جامعه ، n حجم نمونه و $\sigma_{\overline{x}}$ انحراف معیار میانگین می باشد.

توجه داشته باشد که انحراف معیار جامعه یعنی σ_x ثابت است و لذا با توجه به این رابطه واضح است که هر چقدر n افزایش یابد، با توجه به ثابت بودن σ_x در جامعه ، مقدار خطای میانگین ها کاهش می یابد.

مثال : میانگین یک آزمون در سطح دبیرستان های یک استان، ۵۰ و انحراف معیار آن ۱۴ می باشد.

فرض کنید از جامعه، چندین گروه تصادفی با حجم $n = ۴۹$ انتخاب و میانگین هر یک از گروه ها محاسبه شده است. خطای معیار میانگین را محاسبه کنید.

حل :

$$\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{۱۴}{\sqrt{۴۹}} = \frac{۱۴}{۷} = ۲$$

تمرین : از اعداد صفر تا N، به تعداد ۱۰ عدد به تصادف انتخاب شده است. اگر اعداد انتخابی برابر ۸ و ۱۳ و ۱۲ و ۵ و ۱۶ و ۶ و ۹ و ۱۴ و ۱۰ و ۷ باشند، مقدار N را برآورد کنید.

حل :

میانگین نمونه $\overline{x} = \dfrac{۸+۱۳+۱۲+۵+۱۶+۶+۹+۱۴+۱۰+۷}{۱۰} = \dfrac{۱۰۰}{۱۰} = ۱۰$

میانگین جامعه $\mu = \dfrac{۰+۱+۲+۳+...+N}{N+۱} = \dfrac{\frac{N(N+۱)}{۲}}{N+۱} = \dfrac{N(N+۱)}{۲(N+۱)} = \dfrac{N}{۲}$

اکنون این دو میانگین را برابر هم قرار می دهیم.

$$\frac{N}{۲} = ۱۰ \rightarrow N = ۲۰$$

تمرین : با انتخاب همه‌ی نمونه های ۱۰۰ تایی از جامعه ای با واریانس ۱/۹۶ ، انحراف معیار برآورد میانگین، چقدر است؟

حل :

$\delta^۲ = ۱/۹۶ \rightarrow \delta = \sqrt{۱/۹۶} = ۱/۴$ انحراف معیار جامعه

$\delta_{\overline{x}} = \dfrac{\delta}{\sqrt{n}} \rightarrow \delta_{\overline{x}} = \dfrac{۱/۴}{\sqrt{۱۰۰}} = \dfrac{۱/۴}{۱۰} = ۰/۱۴$ انحراف معیار برآورد میانگین

تمرین : اگر انحراف معیار بعد خانوار (میانگین تعداد اعضای خانوارها) در کشوری ۴ نفر باشد.

الف : انحراف معیار برآورد بعد خانوار را برای نمونه هایی با حجم ۱۰۰۰۰ نفری محاسبه کنید.

ب : اگر اندازه‌ی نمونه را ۴ برابر کنیم، انحراف معیار برآورد میانگین چه تغییری می کند؟

حل :

الف) $\delta_{\overline{x}} = \dfrac{\delta}{\sqrt{n}} \rightarrow \delta_{\overline{x}} = \dfrac{۴}{\sqrt{۱۰۰۰۰}} = \dfrac{۴}{۱۰۰} = ۰/۰۴$

ب) $\dfrac{\delta_{\overline{y}}}{\delta_{\overline{x}}} = \dfrac{\dfrac{\delta}{\sqrt{۴n}}}{\dfrac{\delta}{\sqrt{n}}} = \dfrac{\delta\sqrt{n}}{\delta\sqrt{۴n}} = \dfrac{۱}{۲}$

انحراف معیار برآورد میانگین، نصف می شود.

نتیجه : اگر اندازه‌ی نمونه را k برابر کنیم. انحراف معیار برآورد میانگین، $\dfrac{۱}{\sqrt{k}}$ برابر می شود.

تمرین : نمونه ای از یک جامعه به صورت ۷ و ۶ و ۶ و ۶ و ۵ و ۴ و ۳ و ۱ و ۱ و ۱ داده شده است. اگر انحراف معیار جامعه را با آماره‌ی متناظرش در نمونه را برآورد کنیم. انحراف معیار توزیع نمونه ای میانگین ها را برای نمونه‌ی ۴ تایی به دست آورید.

حل :

x_i	F_i	$F_i(x_i - \overline{x})^۲$
۱	۳	$۳(۱-۴)^۲ = ۲۷$
۳	۱	$۱(۳-۴)^۲ = ۱$
۴	۱	$۱(۴-۴)^۲ = ۰$
۵	۱	$۱(۵-۴)^۲ = ۱$
۶	۳	$۳(۶-۴)^۲ = ۱۲$
۷	۱	$۱(۷-۴)^۲ = ۹$
$\sum x_i = ۴۰$	$\sum F_i = ۱۰$	$\sum F_i(x_i - \overline{x})^۲ = ۵۰$

میانگین نمونه $\overline{x} = \frac{\sum x_i}{n} = \frac{۴۰}{۱۰} = ۴$

انحراف معیار جامعه $\delta^{۲} = \frac{\sum F_i (x_i - \overline{x})^{۲}}{n} = \frac{۵۰}{۱۰} = ۵ \to \delta = \sqrt{۵}$

$$\delta_{\overline{x}} = \frac{\delta}{\sqrt{n}} = \frac{\sqrt{۵}}{\sqrt{۴}} = \frac{\sqrt{۵}}{۲}$$

تمرین : در یک مطالعه ، انحراف معیار برآورد میانگین نمونه ها ، از ۰/۰۲ انحراف معیار جامعه کمتر است.حداقل تعداد اعضای نمونه ها چقدر است؟

حل :

$$\delta_{\overline{x}} < ۰/۰۲\delta \xrightarrow{\delta_{\overline{x}} = \frac{\delta}{\sqrt{n}}} \frac{\delta}{\sqrt{n}} < ۰/۰۲\delta \to \frac{۱}{\sqrt{n}} < \frac{۲}{۱۰۰} \to \sqrt{n} > ۵۰ \to n > ۲۵۰۰$$

پس باید حداقل تعداد اعضای نمونه برابر ۲۵۰۱ باشد.

تمرین برای حل :

۱ : سن ۶ نفر از اعضای یک تیم والیبال ۱۹ و ۲۲ و ۲۲ و ۲۳ و ۲۴ و ۲۵ است. برآورد نقطه ای میانگین سن همه‌ی اعضای تیم را به دست آورید.

۲ : اگر انحراف معیار جامعه ای ۲۰ و نمونه ای ۲۲۵ عضوی از آن انتخاب شود. انحراف معیار برآورد میانگین را تعیین کنید.

***

ب : برآورد فاصله ای

گفتیم برای بررسی پارامترهای جامعه، از آن نمونه‌ی تصادفی گرفته و به کمک آماره‌ی مناسب، پارامتر را برآورد نقطه ای می کنیم. چون برآورد بر اساس داده های موجود در نمونه صورت گرفته است، امکان خطا در آن وجود دارد. این خطا، به میزان تفاوت در میانگین نمونه های انتخاب شده بستگی دارد که با اصل افزایش تعداد اعضای نمونه، مقدار خطا کاهش می یابد.

میانگین نمونه (آماره)، بهترین برآورد نقطه ای میانگین جامعه می باشد ولی امکان خطا در آن زیاد است. باتوجه به این توضیحات لزوم ارائه‌ی روشی دیگر، غیر از روش برآورد نقطه ای وجود دارد. به این برآورد، برآورد بازه ای یا فاصله ای گفته می شود.

اگر دقت برآورد یا خطای نمونه گیری را با ε نمایش دهیم، برآورد فاصله ای به صورت زیر بیان می شود.

$$\text{آماره} \pm \varepsilon$$

برای مثال برای میانگین می توان نوشت:

$$\bar{x} \pm \varepsilon \rightarrow \bar{x} - \varepsilon < \mu < \bar{x} + \varepsilon \rightarrow \mu \in (\bar{x} - \varepsilon, \bar{x} + \varepsilon)$$

در برآورد فاصله ای، فاصله‌ی بدست آمده را فاصله‌ی اطمینان و حدود بدست آمده را حدود اطمینان می نامند.

در ادامه چند مورد از برآورد های فاصله ای پارامتر جامعه را معرفی می کنیم.

برآورد فاصله ای میانگین جامعه

برآورد بازه ای میانگین جامعه عبارت است از بازه های عددی برای پارامتر میانگین به همراه یک درصد اطمینان که به ضریب اطمینان شهرت دارد. اگر نمونه ای تصادفی به اندازه‌ی n در اختیار داشته باشیم، با اطمینان بیش از ۰/۹۵ می توانیم بگوییم:

$$\bar{x} - \frac{۲\sigma_X}{\sqrt{n}} < \mu < \bar{x} + \frac{۲\sigma_X}{\sqrt{n}} \rightarrow (\bar{x} - \frac{۲\sigma_X}{\sqrt{n}}, \bar{x} + \frac{۲\sigma_X}{\sqrt{n}})$$

که در آن μ میانگین جامعه، σ_X انحراف معیار جامعه، n حجم نمونه و $\bar{x}$ میانگین نمونه است.

مثال : در نمونه ای به اندازه‌ی ۱۰۰ و میانگین ۵۰ اگر انحراف معیار جامعه ۲۰ باشد، بازه‌ی اطمینان میانگین جامعه را با اطمینان ۰/۹۵ محاسبه کنید.

حل : $\overline{x}=۵۰$ ، $\sigma_x=۲۰$ و $n=۱۰۰$.

$$(\overline{x}-\frac{۲\sigma_x}{\sqrt{n}},\overline{x}+\frac{۲\sigma_x}{\sqrt{n}})=(۵۰-\frac{۲\times۲۰}{\sqrt{۱۰۰}},۵۰+\frac{۲\times۲۰}{\sqrt{۱۰۰}})=(۴۶,۵۴)$$

یعنی به اطمینان ۹۵ درصد ، میانگین جامعه در فاصله‌ی $(۴۶,۵۴)$ قرار می گیرد.

مثال : یک نمونه‌ی تصادفی از ۶۴ لامپ، نشان می دهد که عمر متوسط نمونه ۳۵۰ ساعت است. فاصله‌ی اطمینان ۹۵ درصد، برای متوسط طول عمر واقعی لامپ ها با فرض $\sigma_x=۱۰۰$ را به دست آورید.

حل :

$$(\overline{x}-\frac{۲\sigma_x}{\sqrt{n}},\overline{x}+\frac{۲\sigma_x}{\sqrt{n}})=(۳۵۰-\frac{۲\times۱۰۰}{\sqrt{۶۴}},۳۵۰+\frac{۲\times۱۰۰}{\sqrt{۶۴}})\rightarrow ۳۲۵<\mu<۳۷۵$$

تمرین : رئیس یک دانشگاه علاقمند است، میانگین سن دانشجویانی که در سال جاری ثبت نام کرده اند را بداند. برای این منظور او یک نمونه‌ی تصادفی از ۲۵ دانشجو را انتخاب کرده و میانگین سن آنها را ۲۲ سال بدست آورد. اگر در بررسی های گذشته انحراف معیار سن دانشجویان این دانشگاه برابر ۱/۹ سال باشد، بازه ی اطمینان ۹۵ درصدی برای میانگین سن دانشجویان محاسبه کنید.

حل :

$$(\overline{x}-\frac{۲\delta}{\sqrt{n}},\overline{x}+\frac{۲\delta}{\sqrt{n}})=(۲۲-\frac{۲(۱/۹)}{\sqrt{۲۵}},۲۲+\frac{۲(۱/۹)}{\sqrt{۲۵}})=(۲۱/۲۴,۲۲/۷۶)$$

تمرین : پنج نفر از کارمندان یک اداره را به عنوان نمونه انتخاب کرده ایم. اگر میزان درآمد سالیانه‌ی این افراد را ۱۲ و ۱۸ و ۲۲ و ۲۸ و ۲۰ میلیون تومان باشد و واریانس در آمد کارمندان اداره طبق آمار های قبلی برابر $\frac{۵}{۴}$ باشد. با اطمینان بیش از ۹۵ درصد ، بازه ای برای میانگین درآمد کارمندان اداره بنویسید.

حل :

میانگین نمونه $\overline{x}=\frac{۲۰+۲۸+۲۲+۱۸+۱۲}{۵}=\frac{۱۰۰}{۵}=۲۰$

واریانس جامعه $\delta^۲=\frac{۵}{۴}$

انحراف معیار جامعه $\delta = \sqrt{\frac{۵}{۴}} = \frac{\sqrt{۵}}{۲}$

پس فاصله‌ی اطمینان برآورد میانگین درآمد کارمندان این اداره می شود.

$$(\overline{x} - \frac{۲\delta}{\sqrt{n}}, \overline{x} + \frac{۲\delta}{\sqrt{n}}) = (۲۰ - \frac{۲ \times \frac{\sqrt{۵}}{۲}}{\sqrt{۵}}, ۲۰ + \frac{۲ \times \frac{\sqrt{۵}}{۲}}{\sqrt{۵}}) = (۲۰ - ۱, ۲۰ + ۱) = (۱۹, ۲۱)$$

تمرین : اگر یک نمونه به اندازه‌ی ۴ داشته باشیم و مشاهدات آن ۵ و ۲ و ۱ و ۰ باشند و انحراف معیار جامعه برابر ۱/۸۷ در نظر گرفته شود. یک فاصله‌ی اطمینان ۹۵ درصدی برای میانگین جامعه محاسبه کنید.

حل :

میانگین نمونه $\overline{x} = \frac{۰ + ۱ + ۲ + ۵}{۴} = ۲$

فاصله‌ی اطمینان $(\overline{x} - \frac{۲\delta}{\sqrt{n}}, \overline{x} + \frac{۲\delta}{\sqrt{n}}) = (۲ - \frac{۲(۱/۸۷)}{\sqrt{۴}}, ۲ + \frac{۲(۱/۸۷)}{\sqrt{۴}}) = (۰/۱۳, ۳/۸۷)$

نتیجه : با توجه به برآورد فاصله ای میانگین جامعه، می توان گفت که دقت برآورد میانگین جامعه به اندازه-ی نمونه و انحراف معیار جامعه بستگی دارد. اگر اندازه‌ی نمونه زیاد شود یا انحراف معیار جامعه کم شود، دقت برآورد بیشتر می شود.

مثال : در برآورد فاصله‌ای میانگین، طول بازه‌ی اطمینان را محاسبه کنید. سپس جای خالی را در هر یک از جملات زیر کامل کنید.

الف : اگر اندازه‌ی نمونه افزایش یابد، طول بازه‌ی اطمینان .... می یابد.

ب : اگر انحراف معیار جامعه افزایش یابد، طول بازه‌ی اطمینان .... می یابد.

حل :

فاصله‌ی اطمینان $\overline{x} - \frac{۲\sigma_X}{\sqrt{n}} < \mu < \overline{x} + \frac{۲\sigma_X}{\sqrt{n}}$

طول بازه‌ی اطمینان $\rightarrow (\overline{x} + \frac{۲\sigma_X}{\sqrt{n}}) - (\overline{x} - \frac{۲\sigma_X}{\sqrt{n}}) = \overline{x} + \frac{۲\sigma_X}{\sqrt{n}} - \overline{x} + \frac{۲\sigma_X}{\sqrt{n}} = \frac{۴\sigma_X}{\sqrt{n}}$

الف : کاهش ، زیرا بین طول بازه‌ی اطمینان با جذر اندازه‌ی نمونه رابطه‌ی معکوس وجود دارد.

ب : افزایش ، زیرا بین طول بازه‌ی اطمینان با انحراف معیار جامعه رابطه‌ی مستقیم وجود دارد.

تعریف : طول بازه ای اطمینان برابر تفاضل حد بالا و پایین، فاصله‌ی اطمینان می باشد. بنابراین :

فاصله‌ی اطمینان $(\bar{x}-\frac{۲\delta}{\sqrt{n}},\bar{x}+\frac{۲\delta}{\sqrt{n}})$

طول بازه‌ی اطمینان $(\bar{x}+\frac{۲\delta}{\sqrt{n}})-(\bar{x}-\frac{۲\delta}{\sqrt{n}})=\frac{۴\delta}{\sqrt{n}}$

توجه : اگر بازه‌ی برآورد میانگین جامعه به صورت (a,b) باشد. در این صورت :

میانگین نمونه
$$(\bar{x}-\frac{۲\delta}{\sqrt{n}},\bar{x}+\frac{۲\delta}{\sqrt{n}})\rightarrow\begin{cases}\bar{x}-\frac{۲\delta}{\sqrt{n}}=a\\ \bar{x}+\frac{۲\delta}{\sqrt{n}}=b\end{cases}\rightarrow ۲\bar{x}=a+b\rightarrow\bar{x}=\frac{a+b}{۲}$$

طول بازه‌ی اطمینان
$$(\bar{x}-\frac{۲\delta}{\sqrt{n}},\bar{x}+\frac{۲\delta}{\sqrt{n}})\rightarrow\begin{cases}\bar{x}-\frac{۲\delta}{\sqrt{n}}=a\\ \bar{x}+\frac{۲\delta}{\sqrt{n}}=b\end{cases}\rightarrow\frac{۴\delta}{\sqrt{n}}=b-a$$

توجه :

الف : اگر اندازه‌ی نمونه افزایش یابد، طول فاصله‌ی اطمینان کاهش می باشد.

ب : اگر انحراف معیار جامعه افزایش یابد، طول بازه‌ی اطمینان افزایش می یابد.

تمرین : از جامعه ای با انحراف معیار ۴/۵ ، نمونه ای تصادفی ۱۰۰ عضوی انتخاب می کنیم. طول فاصله-ی اطمینان برآورد میانگین جامعه را محاسبه کنید.

حل :

طول بازه‌ی اطمینان $\frac{۴\delta}{\sqrt{n}}=\frac{۴(۴/۵)}{\sqrt{۱۰۰}}=\frac{۱۸}{۱۰}=۱/۸$

تمرین : میانگین جامعه با اطمینان بیش از ۹۵ درصدی در بازه‌ی $(۱۲/۵ ,۱۸/۵)$ قرار دارد. اگر انحراف معیار جامعه ۱۵ باشد، تعداد اعضای نمونه‌ی انتخاب شده را تعیین کنید.

حل :

$$\frac{۴\delta}{\sqrt{n}} = b - a \rightarrow \frac{۴(۱۵)}{\sqrt{n}} = ۱۸/۵ - ۱۲/۵ \rightarrow \frac{۶۰}{\sqrt{n}} = ۶ \rightarrow \frac{۱۰}{\sqrt{n}} = ۱$$

$$\rightarrow \sqrt{n} = ۱۰ \rightarrow n = ۱۰۰$$

تمرین : اگر بازه‌ی برآورد میانگین جامعه با ضریب اطمینان ۹۵ درصد ، به صورت $(۱۴,۱۸)$ باشد و نمونه‌ی تصادفی انتخاب شده از این جامعه ، ۱۰۰ عضوی باشد. ضریب تغییرات جامعه را برآورد کنید.

حل :

$$\bar{x} = \frac{a+b}{۲} = \frac{۱۴+۱۸}{۲} = ۱۶ \quad \text{میانگین نمونه}$$

$$\frac{۴\delta}{\sqrt{n}} = b - a \rightarrow \frac{۴\delta}{\sqrt{۱۰۰}} = ۱۸ - ۱۴ = \frac{۴\delta}{۱۰} = ۴ \rightarrow \delta = ۱۰$$

$$\text{ضریب تغییرات } CV = \frac{\delta}{\bar{x}} = \frac{۱۰}{۱۶} = \frac{۵}{۸}$$

تمرین : اگر حد بالا و پایین فاصله‌ی اطمینان ۹۵ درصدی برای میانگین جامعه، برابر ۱۸ و ۱۰ باشند و اندازه‌ی نمونه‌ی تصادفی برابر ۸۱ باشد، میانگین نمونه و واریانس جامعه را محاسبه کنید.

حل :

$$\text{میانگین نمونه } \bar{x} = \frac{a+b}{۲} = \frac{۱۰+۱۸}{۲} = ۱۴$$

$$\frac{۴\delta}{\sqrt{n}} = b - a \rightarrow \frac{\delta}{\sqrt{۸۱}} = ۱۸ - ۱۰ \rightarrow \frac{۴\delta}{۹} = ۸ \rightarrow ۴\delta = ۷۲ \rightarrow \delta = ۱۸ \quad \text{انحراف معیار جامعه}$$

$$\rightarrow \delta^۲ = ۱۸^۲ = ۳۲۴ \quad \text{واریانس جامعه}$$

مثال : در یک جامعه، می دانیم میانگین برآورد شده توسط نمونه، ۴ واحد با میانگین جامعه اختلاف دارد.

الف : اگر نمونه ۴۰۰ تایی باشد، حداقل انحراف معیار جامعه را به دست آورید.

حل :

$$\bar{x}-\frac{۲\sigma_X}{\sqrt{n}}<\mu<\bar{x}+\frac{۲\sigma_X}{\sqrt{n}} \to |\mu-\bar{x}|<\frac{۲\sigma}{\sqrt{n}} \to ۴<\frac{۲\sigma}{\sqrt{۴۰۰}} \to ۲<\frac{\sigma}{۲۰} \to \sigma>۴۰$$

ب : اگر واریانس جامعه ۱۰۰ واحد باشد، حداکثر حجم نمونه را بیابید.

حل :

$$\bar{x}-\frac{۲\sigma_X}{\sqrt{n}}<\mu<\bar{x}+\frac{۲\sigma_X}{\sqrt{n}} \to |\mu-\bar{x}|<\frac{۲\sigma_X}{\sqrt{n}} \to ۴<\frac{۲\times۱۰}{\sqrt{n}} \to \sqrt{n}<۵ \to n\leq ۲۵$$

تمرین : اگر واریانس حقوق افراد جامعه ای۱۶ میلیون تومان باشد، تعداد افراد نمونه را چقدر بگیریم، تا طول بازه‌ی اطمینان ۹۵ درصدی برای بر آورد خط فقر در این جامعه ، از ۱۰ درصد بیشتر نباشد؟

حل : خط فقر برابر نصف میانگین درآمد افراد جامعه می باشد. لذا برآورد خط فقر با اطمینان ۹۵ درصدی به شکل زیر محاسبه خواهد شد.

فاصله‌ی اطمینان ۹۵ درصدی از خط فقر $(\bar{x}-\frac{۲\delta}{\sqrt{n}},\bar{x}+\frac{۲\delta}{\sqrt{n}}) \xrightarrow{\div ۲} (\frac{\bar{x}}{۲}-\frac{\delta}{\sqrt{n}},\frac{\bar{x}}{۲}+\frac{\delta}{\sqrt{n}})$

لذا طول بازه‌ی اطمینان ۹۵ درصدی از خط فقر می شود :

$$(\frac{\bar{x}}{۲}+\frac{\delta}{\sqrt{n}})-(\frac{\bar{x}}{۲}-\frac{\delta}{\sqrt{n}})=\frac{۲\delta}{\sqrt{n}}$$

$\delta^۲=۱۶ \to \delta=۴$ واریانس حقوق افراد جامعه

$$\frac{۲\delta}{\sqrt{n}}\leq ۰/۱۰ \to \frac{۲(۴)}{\sqrt{n}}\leq\frac{۱۰}{۱۰۰} \to \frac{۸}{\sqrt{n}}\leq\frac{۱}{۱۰} \to \sqrt{n}\geq ۸۰ \to n\geq ۶۴۰۰$$

***

تمرین برای حل :

۳ : جای خالی را کامل کنید.

الف : مقدار عددی حاصل از جایگذاری اعداد نمونه‌ی تصادفی در آماره‌ی نظیر پارامتر را برآورد .... پارامتر جامعه می نامند.

ب : در صورتی که اندازه‌ی نمونه را کمتر کنیم، امکان نزدیک شدن برآورد به پارامتر .... می شود.

پ : زمانی که اندازه‌ی نمونه با .... برابر باشد، آماره با پارمتر برابر می شود.

ت : انحراف معیار برآورد میانگین، از تقسیم .... بر .... به دست می آید.

۴ : در یک اداره که انحراف معیار سن همه‌ی کارمندان آن ۲۰ است، نمونه ای ۱۶ تایی به طور تصادفی انتخاب می کنیم که میانگین سن آنها ۴۰ است. یک فاصله‌ی اطمینان ۹۵ درصدی برای میانگین سن همه‌ی کارمندان این اداره به دست آورید.

۵ : در یک نمونه‌ی۱۰۰ تایی از جامعه، میانگین برابر ۱۰ و ضریب تغییرات ۲ است. اگر بدانیم انحراف معیار جامعه با انحراف معیار نمونه برابر است، یک فاصله‌ی اطمینان ۹۵ درصدی برای میانگین جامعه به دست آورید و طول این بازه را مشخص کنید.

۶ : اگر از یک کلاس۳۶ نفری، ۴ نفر را به عنوان نمونه انتخاب کنیم و نمرات در آمار و احتمال آنها برابر ۱۲ و ۱۶ و ۱۴ و ۱۰ باشد و اگر بدانیم واریانس جامعه برابر ۱۷ باشد. یک فاصله‌ی اطمینان ۹۵ درصدی برای میانگین نمرات درس آمار و احتمال این کلاس بیابید.

۷ : در نمونه گیری از یک جامعه‌ی بزرگ، اگر میانگین ۲۲ و واریانس ۴ برابر اندازه‌ی نمونه باشد. با ضریب اطمینان ۹۵ درصدی، حد بالا و پایین فاصله‌ی اطمینان برای میانگین را به دست آورید. (اندازه‌ی نمونه بیشتر از ۴۰ نفر است.)

***

برآورد فاصله نسبت جامعه

((برای مطالعه))

از ساده ترین مسائل در آمار استنباطی، برآورد کردن یک نسبت در جامعه است. مثلاً برآورد نسبتی از مردم که می خواهند در تعطیلات نوروز سال آینده به مسافرت بروند. یا درصدی از کشاورزان که در سال گذشته از آبیاری قطره ای در مزارع خود استفاده کرده اند و ...

با توجه به اهمیت این مسائل در برنامه ریزیها و برای برآورد کردن درست این نسبت ها، از برآورد فاصله ای نسبت استفاده می شود. این برآورد مانند برآورد میانگین جامعه با اطمینان ۹۵ درصد انجام می شود.

اگر از جامعه ای نسبتاً بزرگ، n نمونه‌ی تصادفی ساده انتخاب کنیم و m تا از آنها ویژگی مورد مطالعه‌ی ما را داشته باشد، آنگاه نسبت واقعی افرادی از جامعه که آن ویژگی را دارند، با اطمینان ۹۵ درصد بازه‌ی زیر است.

$$(p-۲\sqrt{\frac{p(۱-p)}{n}}\ ,p+۲\sqrt{\frac{p(۱-p)}{n}})$$

که در آن $p=\frac{m}{n}$ برآورد پارامتر نسبت ویژگی در جامعه است.

مثال : فرض کنید، از ۴۸ دانش آموز مدرسه پرسیده ایم که ، آیا برای آمدن به مدرسه، از وسایل نقلیه‌ی عمومی استفاده می کنید؟ از بین این دانش آموزان به تعداد ۳۶ نفر به سئوال ما جواب مثبت داده اند. در این صورت، تعیین کنید که چند درصد از دانش آموزان مدرسه جوابشان به این سئوال مثبت خواهد شد؟ (با اطمینان ۹۵ درصد)

حل : از صورت مسئله معلوم می شود که تعداد نمونه ۴۸ است و ۳۶ نفر این ویژگی را دارند. یعنی $n=۴۸$ و $m=۳۶$ ، پس $p=\frac{۳۶}{۴۸}=۰/۷۵$ لذا داریم:

$$(p-۲\sqrt{\frac{p(۱-p)}{n}}\ ,\ p+۲\sqrt{\frac{p(۱-p)}{n}})$$
$$=(۰/۷۵-۲\sqrt{\frac{۰/۷۵(۱-۰/۷۵)}{۴۸}}\ ,۰/۷۵+۲\sqrt{\frac{۰/۷۵(۱-۰/۷۵)}{۴۸}})$$
$$=(۰/۷۵-۰/۱۲۵\ ,۰/۷۵+۰/۱۲۵)=(۰/۶۲۵,۰/۸۷۵)\longrightarrow\frac{۶۳}{۱۰۰}<P<\frac{۸۸}{۱۰۰}$$

یعنی با اطمینان ۹۵ درصد، نسبت دانش آموزان مدرسه که از وسایل نقلیه‌ی عمومی استفاده می کنند در بازه‌ی $(۰/۶۳\ ,۰/۸۸)$ قرار دارد.

تمرین : برای اینکه بدانیم چند درصد دانش آموزان شهر تهران برای آمدن بـه مدرسـه از وسـایل نقلیـه‌ی عمومی استفاده می کنند. هزار نفر را به عنوان نمونه انتخاب کرده و متوجـه شـدیم کـه ۸۰۰ نفـر از آنهـا، از وسایل نقلیه‌ی عمومی برای رفتن به مدرسه استفاده می کنند. تعیین کنید کـه نـسبت واقعـی افـرادی کـه از وسایل نقلیه‌ی عمومی استفاده می کنند، با اطمینان ۹۵ درصد در چه بازه ای قرار دارد؟

$$p=\frac{m}{n}=\frac{۸۰۰}{۱۰۰۰}=\frac{۸}{۱۰}=۰/۸$$

$$(p-۲\sqrt{\frac{p(۱-p)}{n}}\ ,\ p+۲\sqrt{\frac{p(۱-p)}{n}})$$

$$=(۰/۸-۲\sqrt{\frac{۰/۸(۱-۰/۸)}{۱۰۰۰}}\ ,۰/۸+۲\sqrt{\frac{۰/۸(۱-۰/۸)}{۱۰۰۰}})$$

$$=(۰/۸-۰/۰۲۵۲\ ,۰/۸+۰/۰۲۵۲)=(۰/۷۷۹,۰/۸۲۱)$$

یعنی با اطمینان ۹۵ درصد، نسبت دانش آموزان مدرسه که از وسایل نقلیه‌ی عمومی استفاده می کنند در بازه‌ی $(۰/۷۷۹,۰/۸۲۱)$ قرار دارد. یا اینکه دانش آموزان شهر تهران از ۷۷/۹ تا ۸۲/۱ درصد ، از وسایل نقلیه‌ی عمومی استفاده می کنند.

محاسبه‌ی طول بازه‌ی اطمینان

$$\text{بازه‌ی اطمینان}\ (p-۲\sqrt{\frac{p(۱-p)}{n}},p+۲\sqrt{\frac{p(۱-p)}{n}})$$

$$\text{طول بازه‌ی اطمینان}\ (p+۲\sqrt{\frac{p(۱-p)}{n}})-(p-۲\sqrt{\frac{p(۱-p)}{n}})$$

$$=p+۲\sqrt{\frac{p(۱-p)}{n}}-p+۲\sqrt{\frac{p(۱-p)}{n}}=۴\sqrt{\frac{p(۱-p)}{n}}$$

تمرین : اگر در محاسبه‌ی طول بازه‌ی اطمینان نسبت، تعداد نمونه ها را k برابر کنیم.طول بازه‌ی اطمینان چه تغییری می کند؟

$$\text{طول بازه‌ی اطمینان قدیم}=۴\sqrt{\frac{p(۱-p)}{n}}$$

$$\text{طول بازه‌ی اطمینان جدید}=۴\sqrt{\frac{p(۱-p)}{kn}}$$

کافی است طول بازه‌ی اطمینان جدید را بر طول بازه‌ی اطمینان قدیم تقسیم کنیم.

$$\frac{۴\sqrt{\frac{p(۱-p)}{kn}}}{۴\sqrt{\frac{p(۱-p)}{n}}}=\frac{\sqrt{\frac{p(۱-p)}{kn}}}{\sqrt{\frac{p(۱-p)}{n}}}=\sqrt{\frac{\frac{p(۱-p)}{kn}}{\frac{p(۱-p)}{n}}}=\sqrt{\frac{p(۱-p)n}{p(۱-p)kn}}=\frac{۱}{\sqrt{k}}$$

پس طول بازه‌ی اطمینان باید $\frac{۱}{\sqrt{k}}$ برابر شود.

تمرین : اگر در محاسبه‌ی طول بازه‌ی اطمینان نسبت، تعداد نمونه ها را ۱۰۰ برابر کنیم. طول بازه‌ی اطمینان چه تغییری می کند؟

حل :

$$\frac{۱}{\sqrt{k}}=\frac{۱}{\sqrt{۱۰۰}}=\frac{۱}{۱۰}$$

یعنی طول بازه‌ی اطمینان نسبت، ۰/۱ کوچکتر می شود و این یعنی دقت محاسبه‌ی نسب، یک رقم اعشار بهترخواهد شد.

تمرین : یک موسسه‌ی نظر سنجی معتبر، یک روز قبل از برگزاری انتخابات ریاست جمهوری از یک نمونه-ی ۱۰۰۰ نفری از واجدان شرایط که به طور تصادفی از کل کشور انتخاب شده اند، پرسیده است که آیا در انتخابات شرکت خواهید کرد؟ اگر جواب ۷۰۰ نفر مثبت بوده است، یک بازه‌ی اطمینان ۹۵ درصدی برای درصد شرکت کنندگان در انتخابات به دست آورید؟

حل :

$$p=\frac{m}{n}=\frac{۷۰۰}{۱۰۰۰}=\frac{۷}{۱۰}=۰/۷$$

بازه‌ی اطمینان $\left(p-۲\sqrt{\frac{p(۱-p)}{n}}\ ,\ p+۲\sqrt{\frac{p(۱-p)}{n}}\right)$

$$=\left(۰/۷-۲\sqrt{\frac{۰/۷(۱-۰/۷)}{۱۰۰۰}}\ ,۰/۷+۲\sqrt{\frac{۰/۷(۱-۰/۷)}{۱۰۰۰}}\right)$$

$$=(۰/۷-۰/۰۲۹\ ,۰/۷+۰/۰۲۹)=(۰/۷۶۱,۰/۷۲۹)$$

یعنی اینکه با اطمینان ۹۵ درصد شرکت کنندگان در انتخابات بین ۷۶/۱ و ۷۲/۹ درصد خواهند بود.

توجه : در برآورد بالا، طول بازه‌ی اطمینان ۰/۹۵ می باشد. حال اگر تعداد نمونه ها را k برابر کنیم. طول بازه‌ی اطمینان بر $\sqrt{k}$ تقسیم می شود. پس اگر تعداد نمونه ها را ۴ برابر کنیم، طول بازه‌ی اطمینان نصف می شود و اگر تعداد نمونه ها را صد برابر کنیم، دقت محاسبه‌ی نسبت یک رقم اعشار بهتر خواهد شد.

در رابطه‌ی برآورد نسبت، p در واقع میانگین نسبت ویژگی در مورد مطالعه در نمونه ها و $\sqrt{\frac{p(1-p)}{n}}$ همان انحراف معیار نسبت های نمونه ای می باشند..

مثال : در یک نظر سنجی پیش از انتخابات، فرض کنید که از پیش بدانیم که آرای یکی از نامزدها نزدیک به۵۰ درصد است. اگر بخواهیم طول بازه‌ی اطمینان ۹۵ درصدی ، کمتر از یک درصد باشد، نمونه‌ی ما باید شامل چند نفر باشد؟

حل :

$$p-2\sqrt{\frac{p(1-p)}{n}}<P<p+2\sqrt{\frac{p(1-p)}{n}}$$

$$\to (p+2\sqrt{\frac{p(1-p)}{n}})-(p-2\sqrt{\frac{p(1-p)}{n}})<\frac{1}{100}$$

$$\to 2\sqrt{\frac{p(1-p)}{n}}+2\sqrt{\frac{p(1-p)}{n}}<\frac{1}{100}\to 4\sqrt{\frac{p(1-p)}{n}}<\frac{1}{100}\to \sqrt{\frac{p(1-p)}{n}}<\frac{1}{400}$$

طبق مسئله $p=0/5$ می باشد. پس :

$$\sqrt{\frac{0/5(1-0/5)}{n}}<\frac{1}{400}\to\frac{0/5}{\sqrt{n}}<\frac{1}{400}\to\frac{1}{2\sqrt{n}}<\frac{1}{400}\to\sqrt{n}>200\to n>40000$$

تذکر : طول فاصله‌ی اطمینان وقتی ماکزیمم است که برابر $4\sqrt{\frac{p(1-p)}{n}}$ است وقتی که n را ثابت فرض کنیم، وقتی ماکزیمم است که عبارت $p(1-p)$ ماکزیمم باشد. واضح است که

$$p(1-p)=p-p^2$$

و این عبارت درجه‌ی دوم، وقتی ماکزیمم می شود که $p=\frac{-b}{2a}=\frac{-1}{2(-1)}=\frac{1}{2}$ باشد.

حال با جایگذاری این مقدار در طول فاصله‌ی اطمینان بدست می آوریم :

$$۴\sqrt{\frac{p(۱-p)}{n}}=۴\sqrt{\frac{\frac{۱}{۲}(۱-\frac{۱}{۲})}{n}}=۴\sqrt{\frac{\frac{۱}{۴}}{n}}=۴\sqrt{\frac{۱}{۴n}}=\frac{۲}{\sqrt{n}}$$

که ماکزیمم طول فاصله‌ی اطمینان خواهد بود.

بنابراین در برآورد بازه ای نسبت، بازه‌ی $(p-\frac{۱}{\sqrt{n}}\ ,\ p+\frac{۱}{\sqrt{n}})$ شامل بازه‌ی اطمینان بیش از ۹۵ درصد است.

مثال : فرض کنید بدانیم که در انتخابات ریاست جمهوری، آرای یک از نامزدها نزدیک به ۹۵ درصد است. اگر بخواهیم طول بازه‌ی اطمینان ۹۵ درصدی ، حداکثر برابر یک درصد شود، حداقل تعداد نمونه چقدر باید باشد؟

حل:

$$p=\frac{۵۰}{۱۰۰}=\frac{۱}{۲}$$

پس طول بازه‌ی اطمینان ۹۵ درصدی نسبت، حداکثر برابر ۰/۰۱ خواهد شد و در نتیجه :

$$\frac{۲}{\sqrt{n}}\geq ۰/۰۱\rightarrow\frac{۲}{\sqrt{n}}\geq\frac{۱}{۱۰۰}\rightarrow\frac{۴}{n}\geq\frac{۱}{۱۰۰۰۰}\rightarrow\frac{n}{۴}\geq ۱۰۰۰۰\rightarrow n\geq ۴۰۰۰۰$$

بنابراین حداقل تعداد نمونه برابر ۴۰۰۰۰ نفر باشد.

تمرین : مدیر تولید یک روزنامه می خواهد، درصد روزنامه های معیوب را بررسی کند. برای این منظور ۱۰۰ روزنامه به تصادف انتخاب می شود، که معلوم شد ۱۶ تای آنها معیوب اند.

الف : یک فاصله‌ی اطمینان ۹۵ درصدی برای این روزنامه ها تعیین کنید.

ب : اگر بخواهیم طول فاصله‌ی اطمینان بیش از ۹۵ درصدی ، کمتر یا مساوی یک درصد باشد، باید n را چقدر در نظر بگیریم.

حل :

$$p=\frac{۱۶}{۱۰۰}=۰/۱۶$$

بازه‌ی اطمینان $\left(p-۲\sqrt{\frac{p(۱-p)}{n}}\ ,\ p+۲\sqrt{\frac{p(۱-p)}{n}}\right)$

$$=\left(۰/۱۶-۲\sqrt{\frac{۰/۱۶(۱-۰/۱۶)}{۱۰۰}}\ ,\ ۰/۱۶+۲\sqrt{\frac{۰/۱۶(۱-۰/۱۶)}{۱۰۰}}\right)$$

$$=\left(۰/۱۶-۲\sqrt{\frac{۰/۱۶\times ۰/۸۴}{۱۰۰}}\ ,\ ۰/۱۶+۲\sqrt{\frac{۰/۱۶\times ۰/۸۴}{۱۰۰}}\right)=(۰/۱۶-۰/۰۷۳\ ,\ ۰/۱۶+۰/۰۷۳)$$

$$=(۰/۰۸۷,\ ۰/۲۳۳)\xrightarrow{\times ۱۰۰}(۸/۷\ ,\ ۲۳/۳)$$

$$\frac{۲}{\sqrt{n}}\le ۰/۰۱\to\frac{۲}{\sqrt{n}}\le\frac{۱}{۱۰۰}\to\frac{۴}{n}\le\frac{۱}{۱۰۰۰۰}\to\frac{n}{۴}\ge ۱۰۰۰۰\to n\ge ۴۰۰۰۰$$

تمرین : اگر فاصله‌ی اطمینان ۹۵ درصدی نسبت ویژگی مورد نظر در جامعه $(۰/۶۲۵,۰/۸۷۵)$باشد، در این صورت نسبت و اندازه‌ی نمونه را بدست آورید.

حل :

بازه‌ی اطمینان $\left(p-۲\sqrt{\frac{p(۱-p)}{n}}\ ,\ p+۲\sqrt{\frac{p(۱-p)}{n}}\right)=(۰/۶۲۵\ ,۰/۸۷۵)$

$$\begin{cases}p-۲\sqrt{\frac{p(۱-p)}{n}}=۰/۶۲۵\\ p+۲\sqrt{\frac{p(۱-p)}{n}}=۰/۸۷۵\end{cases}\to ۲p=۱/۵\to p=۰/۷۵$$

طول فاصله‌ی اطمینان $=۰/۸۷۵-۰/۶۲۵=۰/۲۵$

$$\to ۴\sqrt{\frac{p(۱-p)}{n}}=۰/۲۵\xrightarrow{p=۰/۷۵}۴\sqrt{\frac{۰/۷۵(۱-۰/۷۵)}{n}}=\frac{۲۵}{۱۰۰}$$

$$\xrightarrow{\div ۴}\sqrt{\frac{۰/۷۵(۱-۰/۷۵)}{n}}=\frac{۱}{۱۶}\to\frac{۰/۷۵(۱-۰/۷۵)}{n}=(\frac{۱}{۱۶})^۲\to n=\frac{\frac{۳}{۱۶}}{\frac{۱}{۲۵۶}}=۴۸$$

تمرین : یک نمونه‌ی ۴۰۰ تای از کارمندان را برای یافتن نسبت کارمندانی که با موتور تردد می کنند، بررسی کرده ایم، طول بازه‌ی اطمینان ۹۵ درصدی برابر ۰/۳۲ شده است. برای کاهش طول بازه‌ی اطمینان به ۴ درصد تعداد نمونه را چقدر باید افزایش دهیم؟

حل : طول بازه‌ی اطمینان نسبت برابر $۴\sqrt{\frac{p(۱-p)}{n}}$ می باشد.اگر تعدا اعضای نمونه، k برابر شود ، طول بازه ی اطمینان $\frac{۱}{\sqrt{k}}$ برابر می شود.می خوهیم طول بازه ی اطمینان از ۰/۳۲ به ۰/۰۴ کاهش دهیم، پس طول بازه ی اطمینان بر ۸ تقسیم شده است. لذا :

$$\frac{۱}{\sqrt{k}}=\frac{۱}{۸}\to\sqrt{k}=۸\to k=۶۴$$

یعنی تعداد اعضای نمونه باید ۶۴ برابر شود. پس تعداد اعضای نمونه‌ی جدید برابر $۴۰۰\times ۶۴ = ۲۵۶۰۰$ خواهد شود. این یعنی باید $۲۵۶۰۰ - ۴۰۰ = ۲۵۲۰۰$ نفر به تعداد اولیه‌ی اعضای نمونه اضافه کنیم.

تمرین : اگر هیچ اطلاعی از نسبت دانش آموزانی که با علاقه کتاب می خوانند نداشته باشیم، برای برآورد این نسبت با اطمینان بیش از ۹۵ درصد برای اینکه طول بازه‌ی نسبت از ۴۰ درصد کمتر باشد، حداقل تعداد اعضای نمونه باید چقدر باشد؟

حل : طول بازه‌ی نسبت با اطمینان بیش از ۹۵ درصد برابر $\frac{۲}{\sqrt{n}}$ می باشد. پس :

$$\frac{۲}{\sqrt{n}}<۰/۴۰\to\frac{۲}{\sqrt{n}}<\frac{۴۰}{۱۰۰}\to\frac{۲}{\sqrt{n}}<\frac{۲}{۵}\to\sqrt{n}>۵\to n>۲۵$$

یعنی تعداد اعضای نمونه باید حداقل ۲۶ نفر باشد.

نتیجه‌ی ۱ : عبارت $p(۱-p)$ یک چند جمله ای درجه‌ی ۲ است و بیشترین مقدار خود را در حالت $p=۰/۵$ می گیرد. یعنی

$$p(۱-p)=۰/۵(۱-۰/۵)=(۰/۵)^۲$$

لذا برآورد فاصله این نسبت در این حالت به صورت $(p-\frac{۱}{\sqrt{n}}\ ,\ p+\frac{۱}{\sqrt{n}})$ خواهد شد.

لذا در برآورد فاصله ای نسبت، می توان گفت، با اطمینان بیشتر از ۹۵ درصد، نسبت مورد نظر در این بازه زیر است.

$$(p-\frac{1}{\sqrt{n}},p+\frac{1}{\sqrt{n}})$$

که در آن $p=\frac{m}{n}$ می باشد.

نتیجه‌ی۲ : در برآورد بازه ای برای نسبت ، دقت یا خطا عبارت است از

$$e=2\sqrt{\frac{p(1-p)}{n}}$$

می باشد. این مقدار میزان حداکثر خطای مجاز را نشان می دهد.

مثال : مدیر تولید یک روزنامه می خواهد درصد روزنامه های معیوب را بررسی کند. برای این منظور ۱۰۰ روزنامه به تصادف انتخاب می شود که ۱۶ تای آنها معیوب است. یک فاصله‌ی اطمینان ۰/۹۵ برای روزنامه های معیوب محاسبه کنید.

حل : طبق مسئله $p=\frac{16}{100}$ می باشد. حال داریم.

$$p-2\sqrt{\frac{p(1-p)}{n}}<P<p+2\sqrt{\frac{p(1-p)}{n}}$$

$$\rightarrow 0/16-2\sqrt{\frac{0/16\times 0/84}{100}}<P<0/16-2\sqrt{\frac{0/16\times 0/84}{100}}\rightarrow 0/152<P<0/167$$

مثال : با توجه به مثال قبل ، اگر بخواهیم، طول بازه‌ی اطمینان برابر یک درصد باشد، باید n را چقدر انتخاب کنیم.

حل : طبق مثال های قبل می دانیم که $4\sqrt{\frac{p(1-p)}{n}}$ طول بازه‌ی اطمینان می باشند. لذا برای این حالت داریم.

$$4\sqrt{\frac{p(1-p)}{n}}=\frac{1}{100}\rightarrow\sqrt{\frac{p(1-p)}{n}}=\frac{1}{400}\xrightarrow{p=0/16}\sqrt{\frac{0/16\times 0/84}{n}}=\frac{1}{400}$$

$$\rightarrow\sqrt{\frac{0/1344}{n}}=\frac{1}{400}\rightarrow\frac{0/1344}{n}=\frac{1}{160000}\rightarrow n=21504$$

مثال : به منظور ارزیابی کیفیت محصولات تولید شده، تعداد ۲۰۰ واحد محصول را به طور تصادفی انتخاب کرده ایم که بین آنها ۴۰ محصول نقص دار مشاهده شده است. دقت تخمین یا حداکثر خطای مجاز با اطمینان ۹۵ درصد را به دست آورید.

حل : با توجه به مسئله داریم: $m=۴۰$ و $n=۲۰۰$ لذا

$$p=\frac{۴۰}{۲۰۰}=\frac{۱}{۵}$$

پس حداکثر خطای مجاز به صورت زیر است.

$$e=۲\sqrt{\frac{p(۱-p)}{n}}=۲\sqrt{\frac{\frac{۱}{۵}\times\frac{۴}{۵}}{۲۰۰}}=\frac{۴}{۵۰\sqrt{۲}}=\frac{\sqrt{۲}}{۲۵}$$

مثال : حداقل حجم نمونه‌ی مناسب برای تخمین نسبت افرادی که در انتخابات آینده شرکت می کنند، با خطای ۲ درصد و ضریب اطمینان ۹۵ درصد تقریباً چقدر است؟

حل : زمانی که اطلاعات ما در مورد نسبت مورد نظر، محدود است، مقدار p را ۰/۵ در نظر می گیریم و برای تخمین n بر اساس آن محاسبات را انجام می دهیم.

$$e=۰/۰۲\rightarrow ۲\sqrt{\frac{p(۱-p)}{n}}=۰/۰۲\xrightarrow{p=۰/۵}۲\sqrt{\frac{\frac{۱}{۲}\times\frac{۱}{۲}}{n}}=\frac{۱}{۵۰}\rightarrow\frac{۱}{\sqrt{n}}=\frac{۱}{۵۰}$$

$$\rightarrow\sqrt{n}=۵۰\rightarrow n=۲۵۰۰$$

لذا حداقل حجم نمونه‌ی مناسب ۲۵۰۰ واحد آماری باید باشد.

***

تمرین برای حل :

۸ : در یکی از محله های مشهد از میان ۱۰۰ نفر که به تصادف انتخاب شده اند، ۲۰ نفر زائر غیر مشهدی هستند. یک فاصله‌ی اطمینان ۹۵ درصدی برای نسبت زائرین غیر مشهدی پیدا کنید.

۹ : از نظرسنجی از ۴۸ خانواده ، دریافتیم که ۱۶ خانواده مستأجر هستند. یک فاصله‌ی اطمینان ۹۵ درصد برای مستأجران کل کشور پیدا کنید.

***