شاخص های مرکزی و پراکندگی

شاخص‌های مرکزی:

میانگین:

مهم‌ترین و مفیدترین شاخص مرکزی

متوسط (معدل) مقادیر داده‌ها

جمع کل داده‌ها تقسیم بر تعداد

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} f_i x_i$$

ویژگی‌ها:

استفاده از همه داده‌ها در محاسبه

اگر همه داده‌ها با یک عدد خاص جمع، تفریق، ضرب یا تقسیم شوند میانگین به همان نسبت تغییر می‌کند.

مجموع انحراف داده‌ها از میانگین صفر است.

موارد استفاده:

برای داده‌های کمی

شاخص‌های مرکزی:

میانه:

مقداری که نصف داده‌ها از آن کوچکتر باشند (مقدار میانی یا وسطی)

اگر داده‌ها را از کوچک به بزرگ مرتب نماییم، عدد m را میانه این داده‌ها می‌نامیم اگر نصف داده‌ها در سمت چپ و نصف داده در سمت راست این عدد قرار گیرد.

در مجموعه داد مرتب رتبه میانه برابر است با $(n+1)/2$

موارد استفاده:

داده‌های کمی یا کیفی ترتیبی

برتری نسبت به میانگین:

حساس نبودن به مقادیر بسیار بزرگ یا کوچک و متفاوت با سایر داده‌ها

قابلیت استفاده برای متغیرهای ترتیبی

شاخص‌های مرکزی:

نما:

داده‌ای که فراوانی آن نسبت به دیگر داده‌ها بیشتر باشد، (M)

ویژگی:

سادگی محاسبه

موارد استفاده:

برای همه‌ی انواع داده‌ها

شاخص‌های مرکزی:

چارک‌ها

چارک‌های اول، دوم و سوم (Q_1، Q_2 و Q_3) به ترتیب داده‌هایی هستند که ۰/۲۵، ۰/۵۰ و ۰/۷۵ داده‌ها از آن کوچک‌تر هستند. چارک داده ها را به ۴ قسمت مساوی تقسیم میکنند.

دهک‌ها

دهک‌های اول تا نهم (D_1 ، D_2، ...، D_9) به ترتیب داده‌هایی هستند که ۰/۱، ۰/۲، ...، ۰/۹ داده‌ها از آن کوچک‌تر هستند. دهک داده ها را به ۱۰ قسمت مساوی تقسیم میکند.

صدک‌ها

صدک‌های اول تا نود و نهم (P_1، P_2 ، ...، P_{99}) به ترتیب داده‌هایی هستند که ۰/۰۱، ۰/۰۲، ...، ۰/۹۹ داده‌ها از آن کوچک‌تر هستند. صدک داده را به ۱۰۰ قسمت مساوی تقسیم میکند.

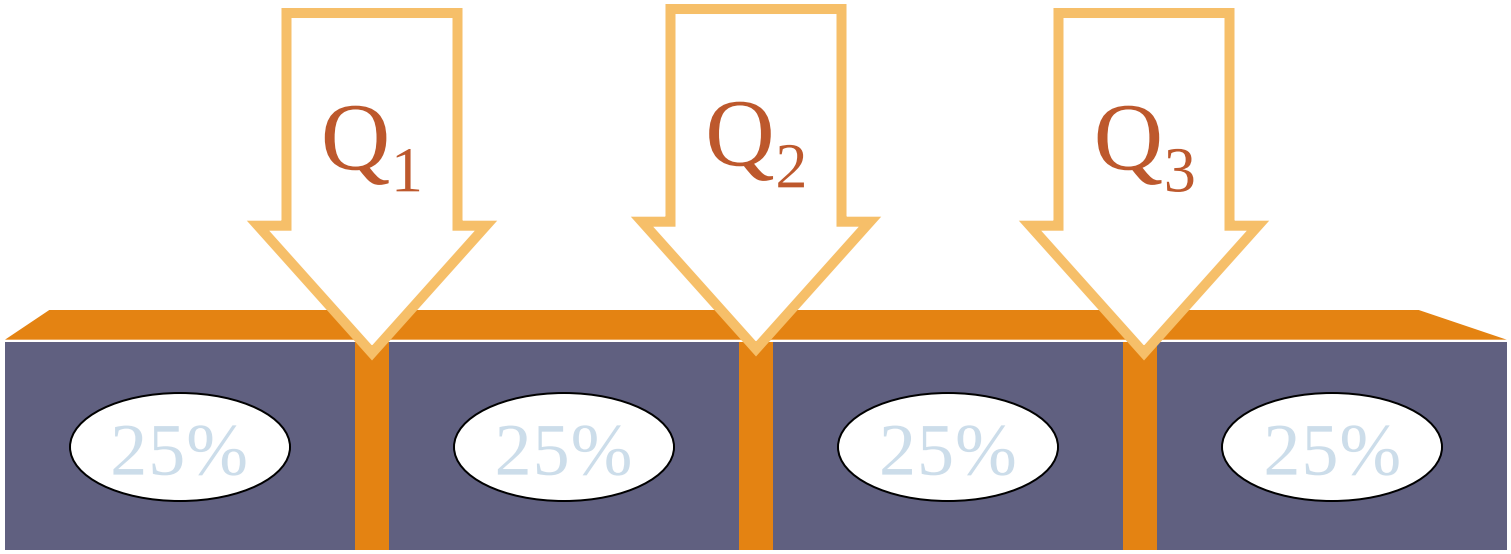
چندک ها:

برای محاسبه چندک ها از فرمول کلی **(n+1)*p** استفاده میشود.

مثال:

چارک اول	**(n+1)/4**
چارک دوم	**(n+1)/2**
چارک سوم	**3(n+1)/4**
دهک ۳	**3(n+1)/10**
دهک ۵	**5(n+1)/10**
صدک ۶۸	**68(n+1)/100**

چارک ها:



نمودار جعبه ای:

نمودار جعبه ای نموداری تصویری است که داده ها را بر اساس پنج مقدار نمایش می دهد. این مقادیر عبارتنداز:

کوچک ترین داده

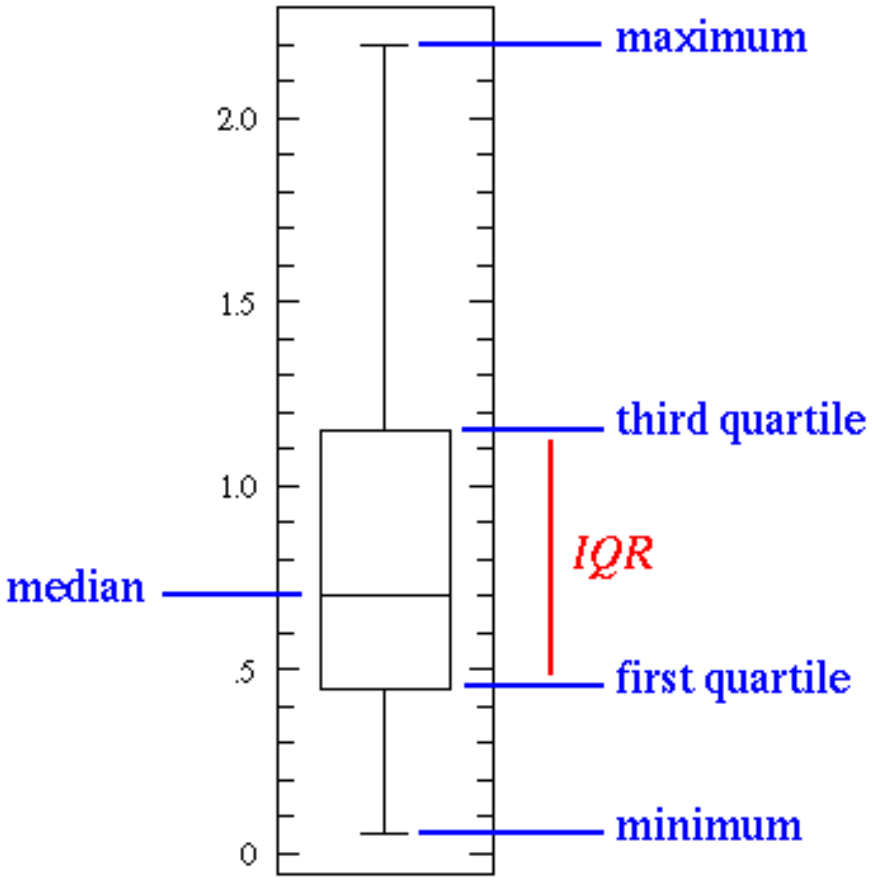
چارک اول

میانه

چارک سوم

بزرگ ترین داده

نمودار جعبه ای:



شاخص‌های پراکندگی:

اهمیت با ذکر مثال:

دو مجموعه داده:

۱۱، ۱۲، ۱۳، ۱۴، ۱۵

۶، ۷، ۱۳، ۱۹، ۲۰

میانگین و میانه برابر

اما تفاوت‌های ساختاری زیاد

لازم است علاوه بر اندازه‌گیری شاخص‌های مرکزی، در خصوص تغییرات، تفاوت‌ها و اختلاف‌های بین داده‌ها نیز کسب اطلاع نمود.

بدین منظور از شاخص‌های پراکندگی استفاده می‌شود.

شاخص‌های پراکندگی:

دامنه	کوچکترین $y_{(1)}$ و بزرگترین داده $y_{(n)}$ $R = y_{(n)} - y_{(1)}$
واریانس	$s^2 = \frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2$
انحراف استاندارد	$s = \sqrt{\frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}$
	$v = \frac{s}{\bar{x}}$

روشی ساده برای اندازه‌گیری پراکندگی داده‌ها

مطلوب نبودن در بسیاری از موارد تنها استفاده از دو داده

متوسط توان دوم اختلاف داده‌ها از میانگین. واحد اندازه‌گیری: توان دوم واحد اندازه‌گیری داده‌ها

جذر مثبت واریانس

واحد اندازه‌گیری مشابه داده‌ها

شاخص پراکندگی بدون واحد مقایسه پراکندگی متغیرهای با واحد اندازه‌گیری متفاوت

تعریف:

نمونه گیری فرایند انتخاب کردن تعداد کافی از میان اعضای جامعه آماری است، به طوری که با مطالعه گروه نمونه و فهمیدن خصوصیات یا ویژگیهای آزمودنیهای گروه نمونـه قـادر خواهیم بود این خصوصیات یا ویژگیها را به اعضای جامعه آماری تعمیم دهیم.

دلایل نمونه گیری:

در بررسی های پژوهشی که شامل چند صد و حتی چند هزار عضو جامعه آماری می شود عملا غیر ممکن است که اطلاعات را از هر عضو جمع آوری کنیم. حتی اگر امکان پذیر هم باشد به لحاظ زمان، هزینه و سایر مسایل منابع انسانی مقدور نیست.

علاوه بر اینها مطالعه یک گروه نمونه به جای کل جامعه آماری گاهی ممکن است منجر به نتایج معتبرتری شود به خاطر اینکه خستگی کمتری وجود خواهد داشت و از این رو خطاهای کمتری درجمع آوری اطلاعات پدید می آورد،مخصوصا موقعی که اعضای جامعه آماری بسیار وسیع باشد.

معرف بودن نمونه:

ما باید بتوانیم گروه نمونه را به صورتی انتخاب کنیم که معرف جامعه آماری باشد و ویژگیهای آن رادر برداشته باشد. به ندرت گروه نمونه جانشین کاملی برای جامعه آماری که از آن بیرون آمده است خواهد بود؛ ولی اگر گروه نمونه را به طور علمی برگزینیم می‌توانیم به طور منطقی مطمئن باشیم که آماره‌های گروه نمونه خیلی نزدیک به پارامترهای جامعه آماری خواهد بود.

نرمال بودن توزیع نمونه:

بسیاری از صفت ها یا ویژگیها در جامعه آماری عموما به طور طبیعی توزیع می شود یعنی صفت هایی مانند قد و وزن چنان هستند که بیشتر مردم اطراف میانگین قرار دارند. اگر قرار باشد ویژگیهای جامعه آماری را با دقت معقول از ویژگی های موجود در گروه نمونه براورد کنیم گروه نمونه باید چنان انتخاب شده باشد که توزیع ویژگیهای مورد نظر ما در آن همان نوع توزیع نرمال را که در جامعه وجود دارد دنبال کند. برای رسیدن به این هدف و داشتن یک گروه نمونه معرف باید اولا حجم نمونه به حد کافی بزرک بگیریم و ثانیا طرح نمونه برداری مناسبی را انتخاب کنیم.

انواع نمونه گیری:

دو نوع اصلی طرح نمونه گیری وجود دارد:

۱. نمونه برداری احتمالی

۲. نمونه برداری غیر احتمالی

نمونه گیری احتمالی:

در نمونه گیری احتمالی اعضای جامعه شانس یا احتمال شناخته شده ای دارند که به عنوان آزمودنی گروه نمونه انتخاب شوند.

طرح های نمونه گیری احتمالی موقعی به کار می رود که معرف بودن گروه نمونه برای اهداف تعمیم پذیری دارای اهمیت باشد.

- ✓ نمونه گیری احتمالی شامل دو دسته نامحدود ومحدود می باشد.

نمونه گیری احتمالی نا محدود (نمونه برداری تصادفی ساده):

در این نوع نمونه گیری اعضای جامعه آماری یک شانس معین وبرابر برای انتخاب شـدن بـه عنوان آزمودنی دارند. این نوع نمونه گیری کمترین سوگیری وبیشترین تعمیم پذیری را دارا می باشد. اما این روش پر زحمت وپر هزینه است وگاهی نمی توانیم فهرست کاملا جدیدی از جامعه آماری بدست آوریم.

نمونه گیری احتمالی محدود یا پیچیده:

مرسوم‌ترین طرحهای نمونه گیری احتمالی پیچیده که به عنوان جایگزین‌هایی برای طرح پرزحمت و پر هزینه احتمالی ساده به شمار می‌روند عبارتند از:

الف) نمونه گیری سیستماتیک

ب) نمونه گیری تصادفی طبقه‌ای

ج) نمونه گیری خوشه‌ای

د) نمونه گیری خوشه ای مرحله ای

الف) طرح نمونه گیری نظام دار(سیستماتیک):

طرح نمونه گیری منظم شامل بیرون آوردن هر n امین عضو جامعه آماری است, به گونه‌ای که این عضو به طور تصادفی بین ۱وn انتخاب می شود. اگر ما گروه نمونه ای از ۳۵ خـانوار از کـل جامعه آماری شامل ۲۶۰ خانه در یک محل خاص را بخواهیم, می‌توانیم با انتخاب شماره تصادفی مثلا ۷ هر هفتمین خانه را با شمارش ۱تا ۷ نمونه بـرداری کنـیم یعنـی خانـه هـای شـماره ۷و ۱۴و۲۱و۲۸...

ب) طرح نمونه گیری تصادفی طبقه ای:

نمونه گیری تصادفی طبقه‌ای مستلزم طبقه‌بندی جامعه آماری و سپس انتخاب تصادفی آزمودنی‌ها از هر طبقه است.در نمونه برداری تصادفی طبقه ای جامعه آماری ابتدا به گروههای ناسازگاری که دربافت پژوهش مرتبط، متناسب و معنا دار هستند تقسیم می‌شود. وسپس از هر طبقه نمونه برداری صورت می‌گیرد. پیدا کردن احتمالات در پارامترهای گروههای فرعی در داخل جامعه آماری بدون روش نمونه گیری تصادفی طبقه ای امکان پذیر نیست. وقتی جامعه آماری با روش معناداری طبقه‌بندی شد, نمونه گیری منظم انتخاب شود.

پ) نمونه گیری خوشه ای :

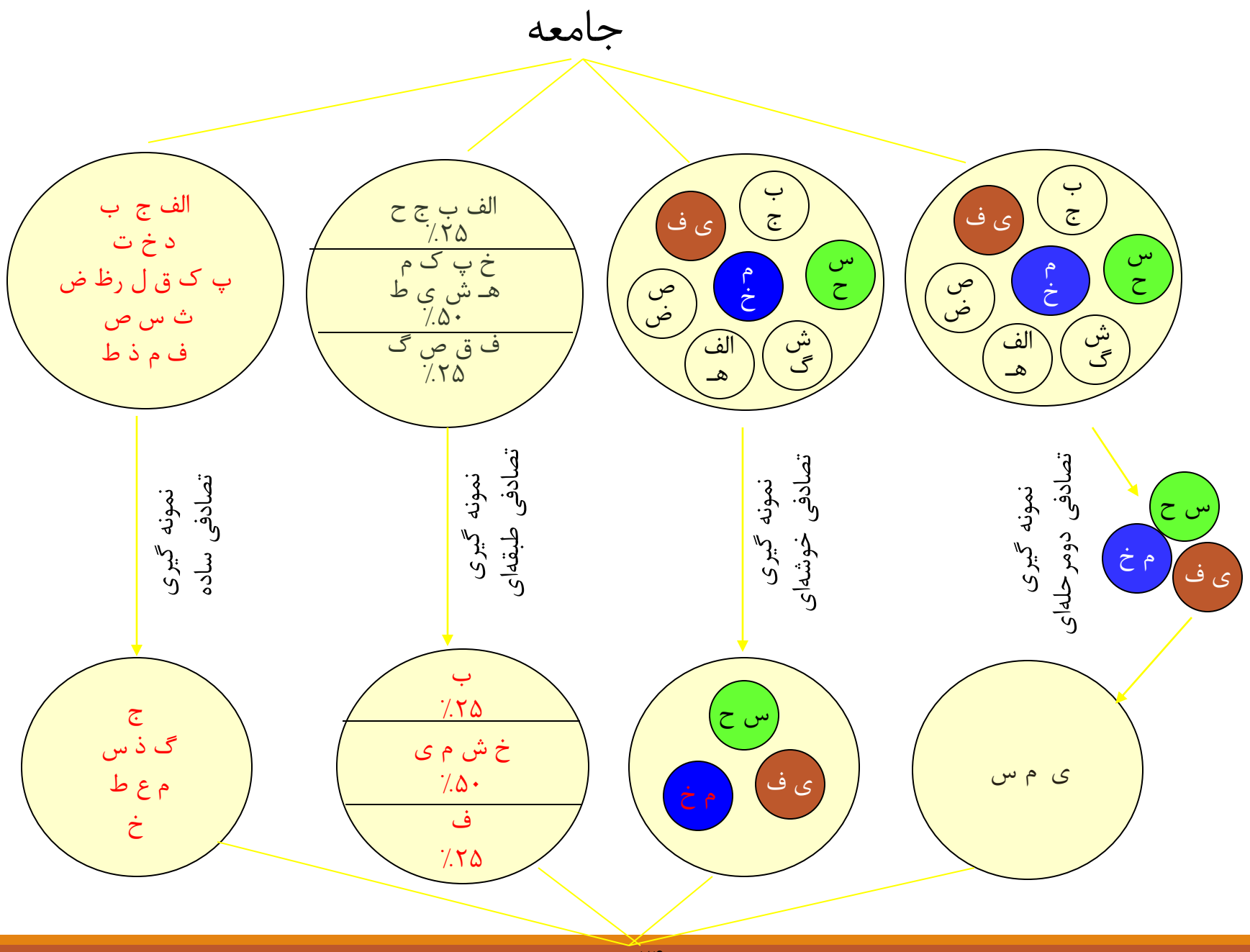
این طرح شامل تقسیم بندی جامعه آماری به خوشه‌های مناسب میشود، در حالی که بطور تصادفی تعداد لازم از خوشه‌ها را به عنوان آزمودنی‌های گروه نمونه انتخاب کرده و پس از آن هم اعضای جامعه آماری در هر یک از این خوشه‌ها مورد مطالعه قرار می‌گیرند.

گروه‌هایی از اعضای جامعه آماری به طوری که عدم تجانس در میان اعضای هر گروه وجود داشته باشد برای مطالعه انتخاب میشوند. نمونه گیری خوشه‌ای عدم تجانس بیشتری را در داخل گروه‌ها و تجانس بیشتری را در بین گروه‌ها ارائه می‌کند که عکس آن چیزی است که ما در نمونه گیری تصادفی طبقه‌ای می‌یابیم زیرا در آن تجانس در هر گروه و عدم تجانس در میان گروه‌ها وجود دارد.

نمونه گیری خوشه‌ای چند مرحله‌ای Multiple cluster sampling:

بواسطه گستردگی بیش از حد جامعه محقق ناگزیر می‌گردد نمونه را طی دو یا چند مرحله انتخاب کند .

- جامعه موردنظر را دقیقا تعریف کنید .
- واحدها یا خوشه‌های نمونه برداری را تعریف کنید .
- تعدادی از خوشه‌ها یا واحدها را به صورت تصادفی انتخاب کنید .
- از میان خوشه‌های انتخاب شده تعداد افراد مورد نظر را به روش تصادفی انتخاب کنید .


جامعه
الف ج ب
د خ ت
پ ک ق ل رظ ض
ث س ص
ف م ذ ط
الف ب ج ح
٪۲۵
خ پ ک م
هـ ش ی ط
٪۵۰
ف ق ص گ
٪۲۵
ی ف
ب ج
س ح
ص ض
خ م
الف هـ
گ ش
نمونه گیری تصادفی ساده
نمونه گیری تصادفی طبقه‌ای
نمونه گیری تصادفی خوشه‌ای
نمونه گیری تصادفی دومرحله‌ای
ج
گ ذ س
م ع ط
خ
ب
٪۲۵
خ ش م ی
٪۵۰
ف
٪۲۵
ی م س
نمونه

نمونه گیری غیر احتمالی:

اعضای جامعه آماری هیچ احتمالی برای انتخاب شدن در گروه نمونه به عنوان آزمودنی را ندارند. بدین معنی که یافته های مطالعه گروه نمونه را نمی توان به اطمینان به جامعه آماری تعمیم داد. و زمانی مورد استفاده قرار می گیرد که پژوهشگر به روش سریع و ارزان بیشتر علا قه من باشد تا تعمیم پذیری یافته ها.

- ✓ نمونه گیری غیر احتمالی دو گروه عمده نمونه برداری در دسترس ونمونه برداری هدف دار را شامل می شود .

الف) نمونه برداری در دسترس:

در این طرح, در دسترس ترین اعضا به عنوان آزمودنی انتخاب می‌شوند این طرح بـه هـیچ وجه قابل تعمیم نمی‌باشد و برای پژوهش علمی مناسب نیست و ممکـن اسـت در مـواقعی برای کسب اطلاعات سریع و استفاده از نتایج آن به کار رود.

ب) نمونه گیری هدف دار:

ممکن است گاهی ضروری باشد به جای کسب اطلاعات از کسانی که در دسترس هستند، اطلاعات را از افراد خاص بدست آوریم یعنی افرادی که قادر خواهند بود اطلاعات مطلوب را ارائه دهند به این دلیل که آنها تنها کسانی هستند که می‌توانند اطلاعات لازم را بدهند یا افرادی هستند که با معیار خاص که پژوهشگر در نظر دارد وفق دهد. چنین روش نمونه گیر ی را هدفمند می‌نامند.

انواع نمونه گیری هدفدار:

نمونه برداری قضاوتی

نمونه برداری سهمیه‌ای

برآورد حجم نمونه:

۱. حجم نمونه برای برآورد یک نسبت در جمعیت:

$$n = \frac{z^2_{1-\alpha/2} \times p(1-p)}{d^2}$$

$$d = 0.1p$$

مثال: محققی میخواهد برآورد نماید که احتمال یک خطر ۱۰ ساله بیماری قلبی عروقی در بین مردان ۴۰ تا ۷۹ ساله در شمال ایران چقدر است. درمطالعه ای در سال ۲۰۰۹ که نتایج آن در سال ۲۰۱۵ منتشر شد این شیوع برابر ۴۸/۹ درصد برآورد گردید. این محقق، به چند نمونه نیاز خواهد داشت، اگر خطای نوع اول را برابر ۵ درصد درنظر بگیرد؟

برآورد حجم نمونه:

۲. حجم نمونه برای برآورد یک میانگین در جمعیت:

$$n = \frac{z_{1-\alpha/2}^2 \times \sigma^2}{\delta^2}$$

مثال: محققی میخواهد میانگین فشار خون سیستولیک زنان ۱۸ سال و بالاتر را در شمال ایران تخمین بزند. براساس مطالعه ای که در گذشته بر روی همین جمعیت صورت گرفت برآورد فشار خون برابر ۱۷/۶۰ ± ۱۱۵/۴۲ بدست آمد. برای هدایت این مطالعه محقق به چند نمونه نیاز خواهد داشت، اگر دقت را در این مطالعه ۲ میلی متر جیوه در نظر گرفته باشد (البته با خطای یک درصد)؟.

برآورد حجم نمونه:

۳. برآورد حجم نمونه برای مقایسه میانگینهای دو جامعه با واریانسهای برابر:

$$n_A = \frac{(\varphi+1)(Z_{1-\alpha/2}+Z_{1-\beta})^2\sigma^2}{\varphi\delta^2}$$

$$n_B = \varphi n_A$$

۴. برآورد حجم نمونه برای مقایسه میانگینهای دو جامعه با واریانسهای نابرابر:

$$n_{A(Variances\ \ Unequal)} = \frac{(Z_{1-\alpha/2}+Z_{1-\beta})^2\left(\sigma_A^2+\frac{\sigma_B^2}{\varphi}\right)}{\delta^2}$$

$$n_{B(Variances\ \ Unequal)} = \varphi \times n_{A(Variance\ \ Unequal)}$$

مثال: فرض کنید محققی بخواهد میانگین فشار خون را در دو جمعیت مقایسه کند. او این فرضیه را در ذهن خود دارد که میانگین فشار خون در این دو جامعه با هم برابر نیست. اگر واریانسها باهم برابر باشند یا نباشند چه مقدار حجم نمونه مورد نیاز است در صورتی که انحراف معیارهای دو جامعه براساس برآوردهای قبلی برابر ۱۵ و ۲۰ میلیمتر جیوه باشد، خطای آلفا ۵ درصد، توان مطالعه ۹۰ درصد و حداقل تفاوت با اهمیت از نظر بالینی برابر ۱۰ میلیمتر جیوه اختیار شود. این محقق نسبت تخصیص را برابر ۱ میگیرد.

برآورد حجم نمونه:

۵. برآورد حجم نمونه برای مقایسه یک میانگین با یک عدد مرجع:

$$n = \frac{\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2 \sigma^2}{\varepsilon^2} \qquad \varepsilon = \left|\mu_1 - \mu_0\right|$$

مثال: فرض کنید محققی میخواهد تعیین کند آیا تفاوت معنی داری بین میانگین قند خون جامعه ای با یک عدد ثابت که آن را ۱۰۰ میلی گرم در دسی لیتر در نظر گرفت، وجود دارد یا خیر؟ فرض کنید او ۱۰ میلیگرم در دسی لیتر تفاوت در قند خون ناشتا را از لحاظ بالینی با اهمیت تلقی کرده و میخواهد بداند که آیا قند خون اندازه گیری شده بیش از ۱۰ میلی گرم با عدد ۱۰۰ تفاوت دارد یا خیر؟ محقق خطای آلفا را ۵ درصد و خطای بتا را ۲۰ درصد برای مطالعه در نظر میگیرد.

برآورد حجم نمونه:

۶. برآورد حجم نمونه برای مقایسه میانگینهای دو جامعه مستقل که نتایج در دو زمان متفاوت بدست میایند:

$$n = \frac{2\sigma_d^2 \times \left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\delta^2} \qquad \sigma_d^2 = \sigma_1^2 + \sigma_2^2 - 2\rho \times \sigma_1^2 \times \sigma_2^2 \qquad \delta = \left|\mu_1 - \mu_2\right|$$

مثال: محققی قصد دارد تغییرات میانگین فشارخونهای سیستولیک را در بین گروه درمانی و کنترل مقایسه نماید. براساس مطالعات قبلی انحراف معیارهای مقادیر پایه و پیگیری (با فاصله یکسال) به ترتیب برابر ۱۵ و ۱۲ میلی متر جیوه وضریب همبستگی بین دو اندازه گیری برابر ۰/۷ بدست آمد. همچنین کاهش میانگین برای گروه درمان در عرض ۱ سال ۱۰ میلیمتر جیوه و گروه کنترل ۳ میلیمتر جیوه بدست آمد. اگر این محقق بخواهد چنین مطالعه ای را با توان ۸۰ درصد و خطای نوع اول ۵ درصد هدایت نماید به چه حجم نمونه ای نیاز خواهد داشت؟