

حجم نمونه

Sample Size

اهمیت حجم نمونه

- یکی از مهمترین شاخص های تعیین کننده توان تصمیم گیری (power) یک مطالعه است.
- حجم نمونه هم باید طوری انتخاب شود که اختلاف یا تفاوت علمی مشاهده شده از نظر علمی معنا دار باشد و هم از نظر اقتصادی با محدودیتهای مواجه نشود.

عوامل موثر بر حجم نمونه

- میزان دقت مورد نیاز
- روش نمونه گیری
- نوع پژوهش
- حجم جامعه
- همگنی و یا عدم همگنی جامعه
- هزینه و در دسترس بودن اعضا

چرا نمونه گیری می کنیم؟

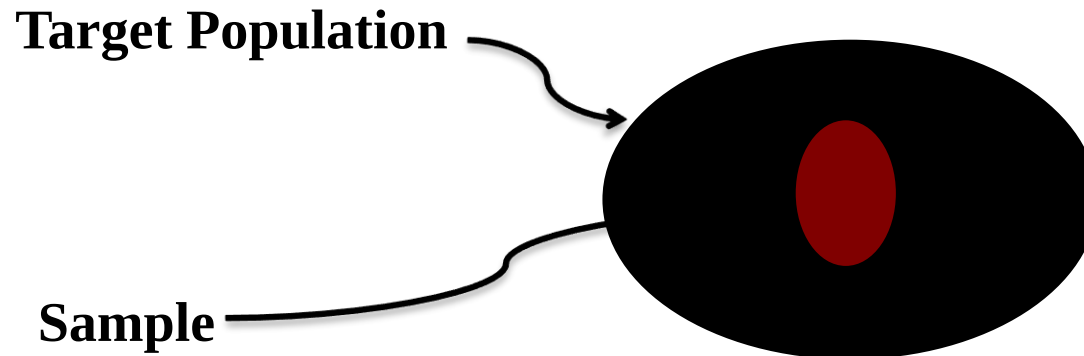
- کاهش هزینه ها
- افزایش سرعت
- افزایش دقت
- غیر عملی بودن سرشماری در بعضی شرایط

نمونه گیری

5

نمونه Sample: گروه یا مواردی که اطلاعات از آنها به دست می آید .

جامعه هدف Target Population: گروه بزرگتری که در صدد هستیم یافته های حاصل از بررسی و مطالعه را به آنها تعمیم دهیم .



آزمون فرضیه

6

- فشارخون سیستولیک طی سال اول و دوم دوره دانشکده افزایش می یابد.
- برای اثبات این موضوع باید فشار خون تمامی دانشجویان سال اول و دوم را اندازه گیری نموده و سپس با هم مقایسه نماییم.
- منتهی قادر به اندازه گیری فشارخون سیستولیک تمام دانشجویان سال اول و دوم نیستیم.
- لذا بطور تصادفی 5 دانشجو را از هر سال انتخاب کرده و فشار خون آنها را اندازه گیری می نماییم.
- دانشجویان سال اول: 120، 80، 90، 110، 95 (Mean=99)
- دانشجویان سال دوم: 105، 130، 145، 125، 115 (Mean=124)
- سؤال: آیا براساس این مشاهدات می توان نتیجه گرفت که واقعاً SBP در دانشجویان سال دوم بالاتر از دانشجویان سال اول می باشد؟

آزمون فرضیه

7

- آیا واقعاً فشارخون سیستمیک در دانشجویان سال دوم بالاتر از دانشجویان سال اول می باشد؟
- با مطرح نمودن فرضیه و انجام آزمون ، به سئوالات فوق پاسخ داده می شود.
- فرضیه صفر (H_0) و یا **Null hypothesis**: تفاوت مشاهده شده بعلت شانس می باشد و بین دو جمعیت تفاوتی وجود ندارد.
- فرضیه مقابل (H_1) و یا (alternative): تفاوت معنی داری وجود دارد و دو جمعیت با یکدیگر تفاوت دارند.

Type I and Type II Errors

8

هنگامی که تصمیم داریم فرضیه صفر را رد یا قبول نماییم، همیشه این **احتمال** وجود دارد که مرتکب خطا شویم.

خطای نوع اول (α): احتمال **رد نمودن** فرضیه صفر هنگامی که **صحیح** است. (بیان تفاوت هنگامی که تفاوتی وجود ندارد)

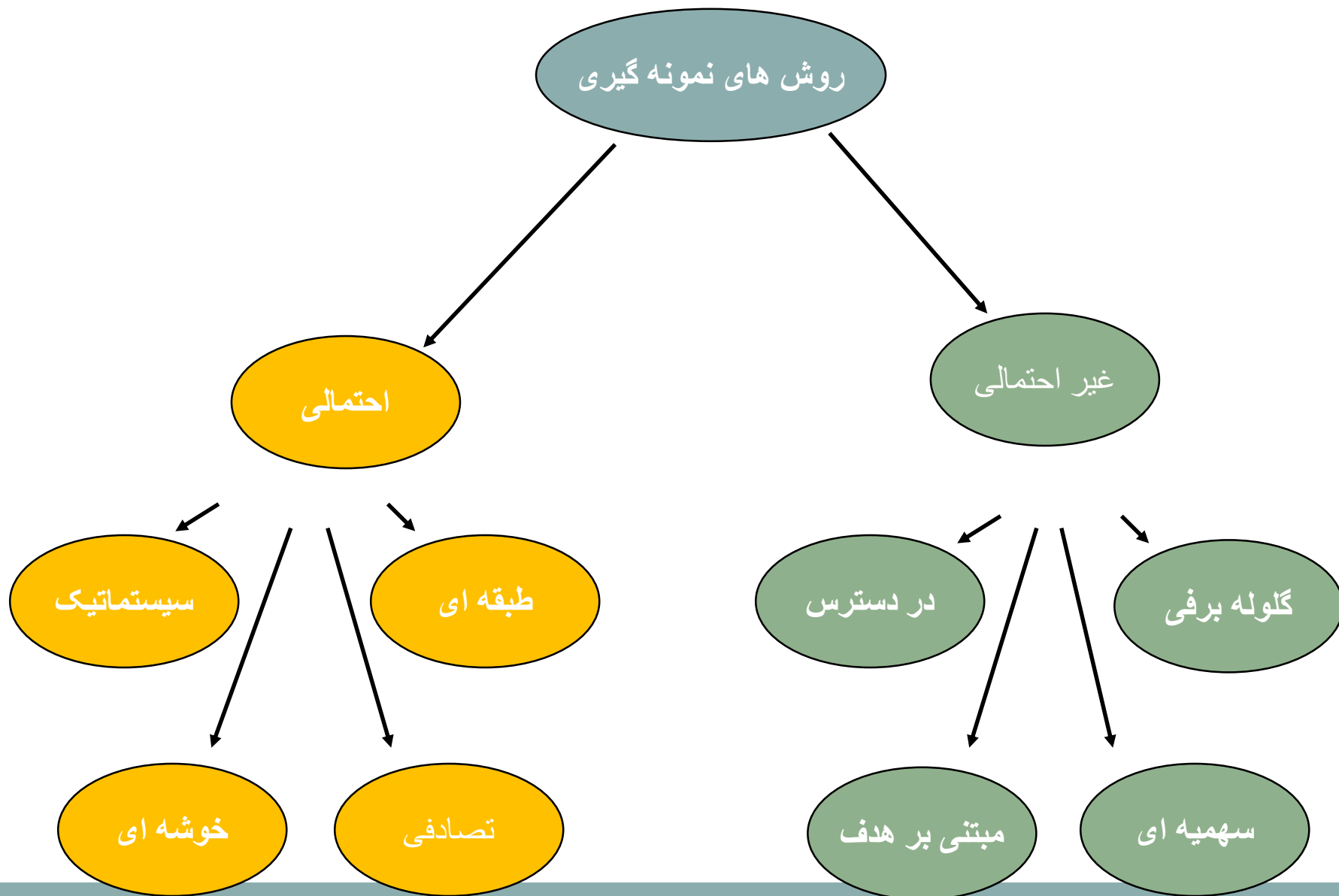
خطای نوع دوم (β): احتمال **رد کردن** فرضیه مقابل هنگامی که **صحیح** است. (ناتوانی در نشان دادن تفاوت هنگامی که تفاوت وجود دارد)

$1 - \alpha$ سطح اطمینان و α سطح معنی داری

$1 - \beta$ توان آزمون

روش هاي نمونه گيري

- غير احتمالي (Non-probability sampling)
- احتمالي (Probability sampling)



نمونه گیری غیر احتمالی

11

- نمونه گیری در دسترس (Convenience sampling): در این روش نمونه گیری از افرادی که بر راحتی در دسترس هستند برای شرکت در مطالعه دعوت بعمل می آید. برای مثال، محقق برای بررسی میزان افسردگی افراد دیابتیک از مراجعه کنندگان به کلینیک دیابت نمونه گیری انجام می دهد. (متداولترین روش)
- نمونه گیری گلوله برفی (Snowballing sampling): در این روش نمونه گیری، از افرادی که وارد مطالعه شده اند برای فراخوان افراد دیگری که شرایط ورود به مطالعه را داشته باشند، جهت شرکت در مطالعه دعوت بعمل می آید. در مواردی که افراد واجد شرایط شناخته شده نیستند و بجای خاصی مراجعه متقیم ندارند. برای مثال، محقق از معتادین تزریقی که برای تست HIV به درمانگاه مراجعه کرده اند می خواهد که به دیگری که می شناسد اطلاع دهد تا آنها نیز در مطالعه شرکت نمایند.

نمونه گیری غیر احتمالی

12

نمونه گیری سهمیه ای (Quota Sampling): نمونه گیری است که در آن محقق بر اساس اطلاعاتی که دارد تصمیم می گیرد که از هر زیر جامعه ای چه نسبتی برای نمونه گیری انتخاب شود. در واقع نمونه گیری سهمیه ای شکلی از نمونه گیری طبقه ای متناسب است که در آن، نسبت از قبل تعیین شده افراد از گروه های مختلف نمونه گیری می شوند اما بر مبنای نمونه گیری در دسترس برای مثال، محقق مشخص می کند که از هر رده شغلی 20 نفر انتخاب شود.

نمونه گیری مبتنی بر هدف (Purposive Sampling): در این نوع نمونه گیری، افرادی برای نمونه انتخاب می شوند که برای ارائه اطلاعات مورد نیاز در بهترین موقعیت قرار دارند. برای مثال، اگر محقق می خواهد بداند چگونه زنان به پست های مدیریت عالی دست می یابند، تنها افرادی که می توانند اطلاعات دست اول در این مورد به وی بدهند مدیران زنی هستند که در پست های ریاست، معاونت و سطوح عالی مدیریت اجرایی در سازمان ها، ادارات و دانشگاه ها قرار دارند.

معایب روش های نمونه گیری غیر احتمالی

- شانس انتخاب برای همه اعضاء جامعه مشخص نیست.
- نمونه حاصل معرف جامعه نمی باشد.
- ممکن است بعضی از واحدها زیاد و برخی کم انتخاب شوند.
- نتایج بدست آمده را نمی توان با ادعای قوی به جامعه هدف تعمیم داد.
- روش مناسب برای برآورد پارامترهای جامعه نمی باشد.

محاسن نمونه گيري احتمالي

- شانس هر فرد براي ورود به نمونه معين، و نامساوي صفر است.
- امکان سوگيري انتخاب را، کم مي کند.
- اجازه مي دهد که از تئوري هاي آماري در تحليل ها استفاده شود.

نمونه گيري تصادفي ساده

Simple random sampling

■ شانس انتخاب همه واحدهاي نمونه گيري مساوي است

■ روش اجرا:

- تهيه فهرست كل اعضاي جامعه (چارچوب نمونه گيري)
- انتخاب تصادفي هر نمونه

نمونه گيري تصادفي ساده

مزایا

آسان است

خطاي تصادفي (sampling error) به سادگي قابل اندازه گيري است.

معایب

احتیاج به چار جوب نمونه گيري دارد

ممکن است نمونه، بهترین معرف جامعه نباشد.

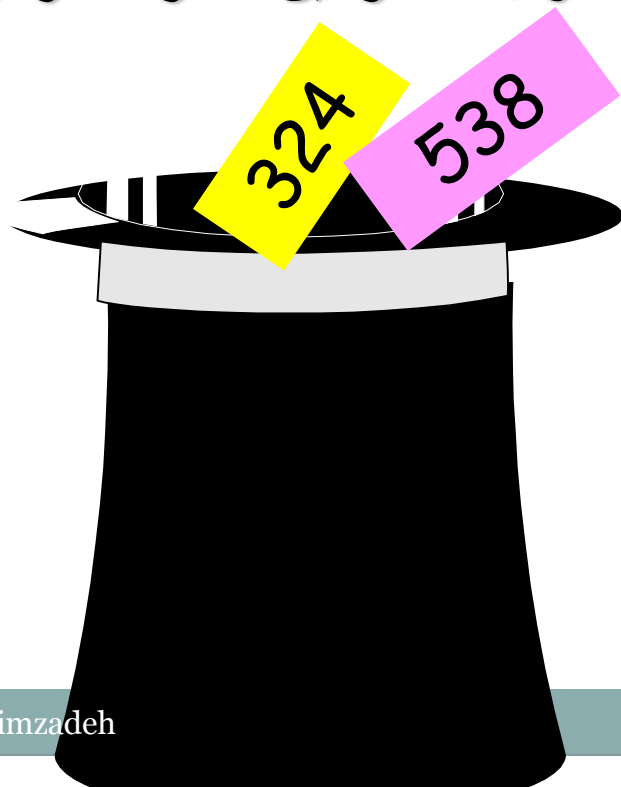
دسترسى سخت به واحدهاي نمونه انتخاب شده که پراکنده هستند.

براي انجام نمونه گيري تصادفي ساده بايد:

- فهرستي شماره دار از واحدهاي جامعه که قرار است نمونه از آنها انتخاب شود تهيه گردد.
- اندازه نمونه مشخص مي شود.
- براي هر واحد جامعه نمونه گيري یک عدد در نظر گرفته مي شود.
- واحدهاي مورد نیاز با قرعه کشي يا جدول اعداد تصادفي انتخاب شوند.

نمونه گیری تصادفی ساده

- برای مثال فرض کنید از بین 623 دانش آموز یک مدرسه، می خواهیم یک نمونه 40 نفری برای محاسبه میزان هموگلوبین خون داشته باشیم. برای این منظور مراحل زیر را دنبال می کنیم.
- 1- به هر یک از دانش آموزان عددی از 001 تا 623 نسبت می دهیم.
- 2- این شماره ها را در یک ظرف گذاشته و 40 تا از این شماره ها را بتصادف انتخاب می کنیم.



نمونه گیری تصادفی منظم یا سیستماتیک

Systematic sampling

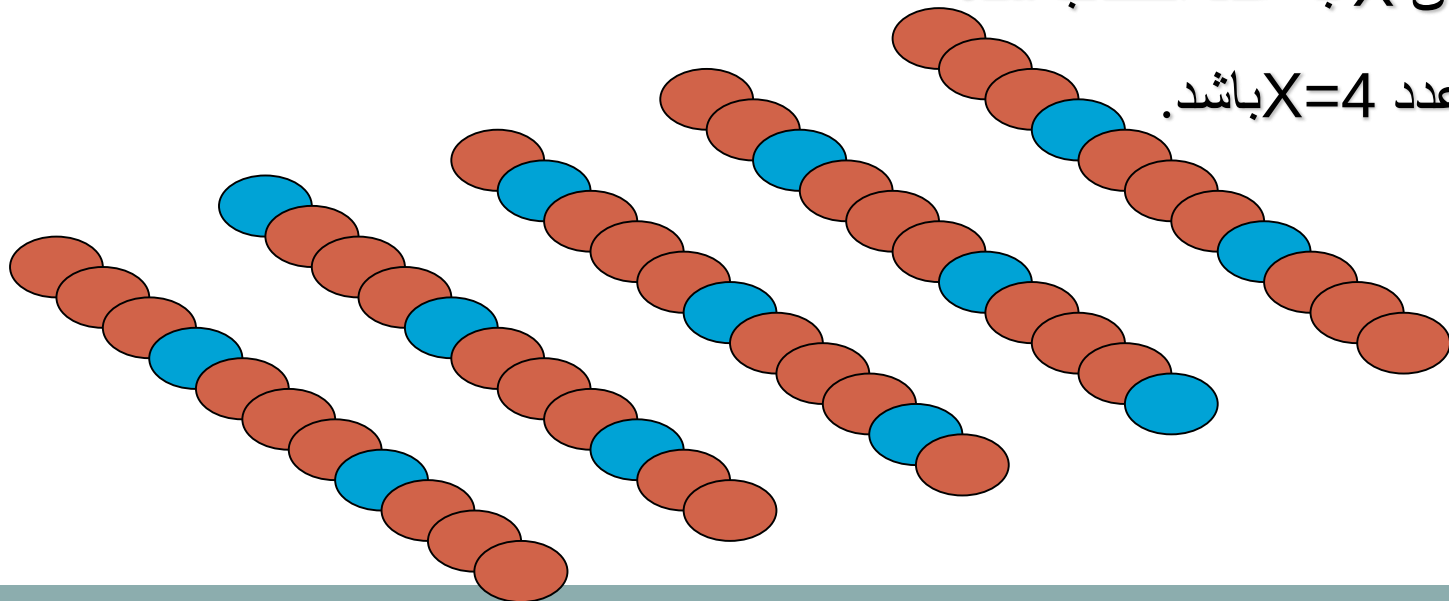
■ تهیه فهرست واحد های جامعه (چارچوب نمونه گیری)

■ تقسیم اندازه جامعه به تعداد نمونه مورد نیاز (X)

■ انتخاب یک عدد تصادفی بین 1 و X

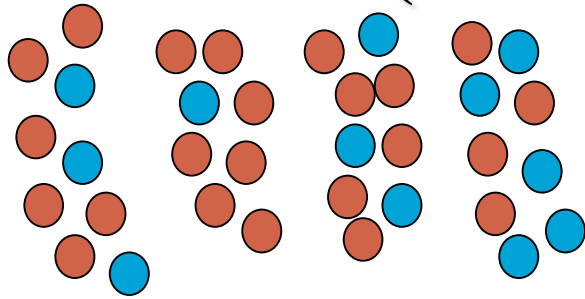
■ افزودن X به عدد انتخاب شده

■ اگر عدد $X=4$ باشد.



نمونه گيري طبقه ای

(Stratified sampling)



- جامعه را به تعدادي طبقه افراز مي كنيم
- بين طبقه ها نبايد هم پوشاني وجود داشته باشد
- هيچ فردي از جامعه نبايستي بيرون از طبقه بندي قرار بگيرد
- از هر طبقه تعدادي نمونه انتخاب مي كنيم
- يافته هاي حاصل از طبقات را روي هم مي ريزيم

هر چقدر داخل طبقات افراد بيشتر به هم شبیه باشند (homogeneity) ولي بين طبقه ها تفاوت زياد وجود داشته باشد، نمونه گيري طبقه اي بهتر است

نمونه گيري طبقه بندي شده

■ مزایا

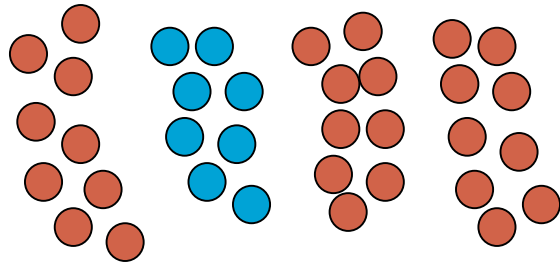
- اگر داخل طبقه ها افراد به هم شبیه باشند، کارایی (دقت) این روش از روش تصادفی ساده بیشتر است.
- به علت وجود داده های مجزا از هر يك از طبقات، امکان تحلیل جداگانه برای هر يك از طبقه ها وجود دارد.

■ معایب

- اگر صفت مورد نظر در داخل طبقه ها خیلی متفاوت باشد، این روش مناسب نیست
- محاسبه خطاي نمونه گيري (به دست آوردن SE Mean) مشکل است.

نمونه گيري خوشه اي (Cluster sampling)

- جامعه را به تعدادي گروه (خوشه) از واحدها تقسيم مي كنيم
- تعدادي از خوشه ها را به طور تصادفي انتخاب مي كنيم
- در هر خوشه همه واحدها (يا تعدادي از آنها) را انتخاب مي كنيم.



هر چقدر داخل خوشه ها افراد کمتر به هم شبیه (Heterogeneity) بیشتر باشند نمونه گيري خوشه اي موفقتر است.

نمونه گیری خوشه ای

Cluster sampling

■ مزایا

باعث صرفه جویی در وقت و هزینه گردد.

■ معایب

دقت این نمونه گیری کمتر از انواع قبلی است.

نمونه گیری چند مرحله‌ای

Multistage sampling

24

- نمونه گیری چند مرحله ای (multistage): یعنی ترکیب این روشها و تنظیم مدلی متناسب با نیازها و شرایط

بررسی میزان مصرف سیگار در دانش آموزان شهر کرج

- تقسیم شهر کرج به چند منطقه
- تقسیم مدارس هر منطقه به پسرانه و دخترانه
- انتخاب 3 مدرسه پسرانه و 3 مدرسه دخترانه در هر منطقه
- انتخاب یک کلاس از هر رده تحصیلی
- گزینش دانش آموزانی که فامیل آنها با “ک”، “د” و “ل” شروع می شود

محاسبه حداقل حجم نمونه برای برآورد میانگین (صفت کمی)

25

$$n = \left(\frac{Z_{1-\alpha/2} \times S}{d} \right)^2$$

α : احتمال خطاي نوع اول؛

اگر $\alpha = 0/05$ باشد $Z_{1-\alpha/2}$ برابر 1/96 است

s : انحراف معیار (SD) صفت مورد نظر

d: خطاي قابل قبول در برآورد میانگین

مثال: یک متخصص تغذیه در نظر دارد مطالعه ای برای برآورد مصرف پروتئین روزانه افراد نوجوان انجام دهد. با استفاده از تجربیات گذشته انحراف معیار مصرف پروتئین در جامعه مورد مطالعه ۲۰ گرم است اگر حداکثر خطای برآورد را ۵ گرم و میزان اطمینان را ۹۵٪ در نظر بگیریم اندازه نمونه چقدر است؟

$$s = 20$$

$$n = \frac{(z_{1-\alpha/2} \times s)^2}{d^2} = \frac{(1/96)^2 \times 20^2}{5^2} = 62$$

۹۵٪ سطح اطمینان

$$d = 5$$

$$d = 2.5 \text{ — } n = 246$$

خطا نصف شود

محاسبه حداقل حجم نمونه برای برآورد نسبت (صفت کیفی)

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

α : احتمال خطای نوع اول؛

اگر $\alpha = 0/05$ باشد $Z_{1-\alpha/2}$ برابر 1/96 است

p : تخمین نسبت (proportion) صفت مورد نظر

d : خطای قابل قبول در برآورد نسبت مورد نظر

مثال: یک سازمان منطقه ای بهداشت در نظر دارد میزان شیوع بیماری چاقی را در کودکان زیر ۵ سال یک منطقه برآورد کند چه تعداد نمونه نیاز دارد تا با اطمینان ۹۵٪ میزان شیوع را در فاصله ۰/۰۵ مقدار واقعی آن برآورد کند در صورتیکه میزان شیوع واقعی بیماری از ۲۰٪ تجاوز نمی کند.

۹۵٪ سطح اطمینان

$d=0.05$ دقت مطلق

$z = 1.96$ جدول توزیع نرمال

$P=0.2$

$$n = \frac{(z_{1-\alpha/2})^2 \cdot p(1-p)}{d^2} = \frac{(1.96)^2 \times 0.2 \times 0.8}{(0.05)^2} = 246$$

اصلاح حجم نمونه جامعه محدود

29

- فرمول هایی که در بخش های قبلی ارائه شد برای وقتی که حجم جامعه بزرگ باشد، مناسب است.
 - در صورتی که حجم جامعه کمتر از 3000 باشد با استفاده از فرمول زیر که به ضریب تصحیح جامعه محدود مشهور است حجم نمونه را اصلاح استفاده می کنیم.
- $$\hat{n} = \frac{Nn_0}{N + n_0 - 1}$$
- در مثال تغذیه حجم نمونه 62 نفر برآورد گردید، حال اگر حجم جامعه 500 نفر باشد، حجم نمونه با استفاده از فرمول اصلاح جامعه محدود 56 نفر به دست می آید.
 - هر چقدر جامعه کوچکتر نمونه به دست آمده با تصحیح جامعه محدود کوچکتر می شود. اگر حجم جامعه 100 باشد نمونه مورد نیاز به 39 نفر کاهش می یابد.

محاسبه حداقل حجم نمونه برای مقایسه میانگین ها در دو جامعه مستقل

$$n = \frac{(s_1^2 + s_2^2) (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{(\bar{x}_1 - \bar{x}_2)^2}$$

اطلاعات قبلی در مورد گروه اول $\left\{ \begin{array}{l} \bar{x}_1 \\ s_1 \end{array} \right.$

اطلاعات قبلی در مورد گروه دوم $\left\{ \begin{array}{l} \bar{x}_2 \\ s_2 \end{array} \right.$

مثال: می خواهیم تاثیر دو نوع رژیم غذایی را در کاهش فشار خون بررسی کنیم با توجه به مطالعه مقدماتی اطلاعات بدست آمده در دو گروه به شرح زیر است:

$$\begin{array}{l} \text{گروه A} \\ \text{گروه B} \end{array} \left\{ \begin{array}{l} \bar{X}_1 = 132.8 \\ S_1 = 15.3 \\ \bar{X}_2 = 127.4 \\ S_2 = 18.2 \end{array} \right.$$

$$\alpha = 0.05 \qquad 1 - \beta = \%80$$

$$n = \frac{(1.96 + 0.84)^2 (15.3^2 + 18.2^2)}{(132.8 - 127.4)^2} = 152$$

در هر گروه به 152 نفر برای مطالعه نیاز داریم.

محاسبه حداقل حجم نمونه برای مقایسه نسبت ها در دو جامعه مستقل

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \times [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}$$

α : احتمال خطای نوع اول؛ اگر $\alpha = 0/05$ باشد $Z_{1-\alpha/2}$ برابر 1/96 است

β : احتمال خطای نوع دوم؛ اگر $\beta = 0/2$ باشد $Z_{1-\beta}$ برابر 0/84 است

نسبت در گروه اول: P_1

نسبت در گروه دوم: p_2

مثال: در یک مطالعه می خواهیم درصد شیوع بیماری دیابت را در بین افراد بالای 50 سال که ورزش می کنند و نمی کنند با اطمینان 95 درصد و توان 80 درصد را با هم مقایسه کنیم.

با توجه به اطلاعات قبلی می دانیم که شیوع دیابت در افرادی که ورزش می کنند 5% و در کسانی که ورزش نمی کنند 10% می باشد.

$$P_1 = 0.05 \quad P_2 = 0.1$$

$$\alpha = 0.05 \quad 1 - \beta = 80\%$$

$$n = \frac{(1.96 + 0.84)^2 (0.05 \times 0.95 + 0.1 \times 0.9)}{(0.1 - 0.05)^2} = 432$$

در هر گروه نیاز به 432 نفر برای مطالعه داریم.

Calculations

34

• در آدرس زیر می توانید برنامه ای را پیدا کنید که حجم نمونه را برای فرمول های ارائه شده در این کارگاه برای شما محاسبه می نماید.

• <http://www.stat.ubc.ca/~rollin/stats/ssize/>

This page contains links to JavaScript based forms for simple power available at [UCSF Biostatistics: Power and Sample Size Programs](#).

- [Comparing a Mean to a Known Value](#)
- [Comparing Means for Two Independent Samples](#)
- [Comparing a Proportion to a Known Value](#)
- [Comparing Proportions for Two Independent Samples](#)
- [Unmatched Case Control Studies](#)

Comparing a Mean to a Known Value

35

- در این صفحه مقدار آلفا بطور پیش فرض 0.05 در نظر گرفته شده است و مقدار بتا را نیز 0.5 قرار دهید. مقداری که برای μ_0 و μ_1 قرار می دهید باید بگونه ای باشد که اختلاف آنها برابر با مقدار d باشد و مقدار σ نیز انحراف معیار می باشد.
- سپس دگمه calculate را فشار دهید تا حجم نمونه را برای شما محاسبه نماید.

$$n = \frac{(z_{1-\alpha/2} \times s)^2}{d^2} = \frac{(1/96)^2 \times 20^2}{5^2} = 62$$

- ☒ Calculate Sample Size (for specified Power)

- ☐ Calculate Power (for specified Sample Size)

Enter a value for μ_0 :

10

Enter a value for μ_1 :

5

Enter a value for σ :

20

- ☐ 1 Sided Test

- ☒ 2 Sided Test

Enter a value for α (default is .05):

.05

Enter a value for desired power (default is .80):

.5

The sample size is:

62

Calculate

Comparing Means for Two Independent Samples

- در این صفحه مقدار آلفا 0.05 و مقدار بتا 0.8 می باشد. مقدار واریانس در دو گروه برابر با هم و مساوی sigma (انحراف معیار) و مقدار آن 16 در نظر گرفته می شود.

- مقدار μ_0 میانگین در گروه اول و μ_1 میانگین در گروه دوم است.

$$n = \frac{(s_1^2 + s_2^2)(z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{(\bar{x}_1 - \bar{x}_2)^2} = 138$$

- ☒ Calculate Sample Size (for specified Power)

- ☐ Calculate Power (for specified Sample Size)

Enter a value for μ_1 :

Enter a value for μ_2 :

Enter a value for sigma:

- ☐ 1 Sided Test

- ☒ 2 Sided Test

Enter a value for α (default is .05):

Enter a value for desired power (default is .80):

The sample size (for each sample separately) is:

Comparing a Proportion to a Known Value

$$n = \frac{(z_{1-\alpha/2})^2 \cdot p(1-p)}{d^2} = \frac{(1/96)^2 \times 0/2 \times 0/8}{(0/05)^2} = 246$$

- در این صفحه مقدار آلفا بطور پیش فرض 0.05 در نظر گرفته شده است و مقدار بتا را نیز 0.5 قرار دهید. مقداری که برای p_0 قرار می دهید شیوع ویژگی مورد نظر و مقداری که برای p_1 قرار می دهید مجموع d با p_0 است
- سپس دگمه calculate را فشار دهید تا حجم نمونه را برای شما محاسبه نماید.

- ☒ Calculate Sample Size (for specified Power)
- ☐ Calculate Power (for specified Sample Size)

Enter a value for p_0 :

Enter a value for p_1 :

- ☐ 1 Sided Test
- ☒ 2 Sided Test

Enter a value for α (default is .05):

Enter a value for desired power (default is .80):

The sample size is:

Comparing Proportions for Two Independent Samples

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 * [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}$$

$$n = \frac{(1.96 + 0.84)^2 (0.05 * 0.95 + 0.1 * 0.9)}{(0.1 - 0.05)^2} = 432$$

مقدار p_1 و p_0 شیوع در گروه اول و دوم می باشد.
مقدار آلفا و بتا نیز بطور پیش فرض بر روی 0.05 و 0.8 ثابت شده است.

حجم نمونه به دست آمده مقدار کمی با فرمول بالا تفاوت دارد.

- ☒ Calculate Sample Size (for specified Power)
- ☐ Calculate Power (for specified Sample Size)

Enter a value for p1:

Enter a value for p2:

- ☐ 1 Sided Test
- ☒ 2 Sided Test

Enter a value for α (default is .05):

Enter a value for desired power (default is .80):

The sample size (for each sample separately) is:

روش مورگان

39

این روش در مواقعی استفاده می شود که مقیاس اندازه گیری رتبه ای و حجم جامعه کم باشد. با فرمول زیر می توان حجم نمونه را انتخاب کرد.

$$S = \frac{\chi^2 N \times pq}{(N - 1)d^2 + \chi^2}$$

توجه داشته باشید که در فرمول بالا:

S = تعداد نمونه مورد نیاز است این مقدار در جدول مشخص شده است.

N = تعداد اعضای جامعه است که در جدول آمده است.

P = نسبت جمعیت است (در جدول زیر برای تعیین واریانس از نسبت ۰.۵ استفاده شده است).

d = درجه دقت نسبت بیان شده است (که در این جدول $d = ۰/۰۵$ در نظر گرفته شده است).

χ^2 = مقداری از جدول کای اسکور با یک درجه آزادی در سطح اطمینان ۹۵٪ است (اگر به جدول مراجعه کنید این مقدار ۳/۸۴۱ است که جدول مذکور نیز با توجه به این مقدار تهیه شده است).

بر اساس فرمول فوق جدولی تهیه شده که می توانید با داشتن حجم جامعه، مقدار نمونه لازم را از این جدول معلوم کنید.

جدول مورگان

Population Table for Determining Sample Size from a Given

<i>S</i>	<i>N</i>	<i>S</i>	<i>N</i>	<i>S</i>	<i>N</i>
291	1200	140	220	10	10
297	1300	144	230	14	15
302	1400	148	240	19	20
306	1500	152	250	24	25
310	1600	155	260	28	30
313	1700	159	270	32	35
317	1800	162	280	36	40
320	1900	165	290	40	45
322	2000	169	300	44	50
327	2200	175	320	48	55
331	2400	181	340	52	60
335	2600	186	360	56	65
338	2800	191	380	59	70
341	3000	196	400	63	75
346	3500	201	420	66	80
351	4000	205	440	70	85
354	4500	210	460	73	90
357	5000	214	480	76	95
361	6000	217	500	80	100
364	7000	226	550	86	110
367	8000	234	600	92	120
368	9000	242	650	97	130
370	10000	248	700	103	140
375	15000	254	750	108	150
377	20000	260	800	113	160
379	30000	265	850	118	170
380	40000	269	900	123	180
381	50000	274	950	127	190
382	75000	278	1000	132	200
384	1000000	285	1100	136	210