بسم الله الرحمن الرحيم

تعیین حجم نمونه

گرداورنده: سمانه افتخاری مهابادی
هیئت علمی بخش آمار
دانشکده‌ی ریاضی، آمار و علوم کامپیوتر
دانشکدگان علوم
دانشگاه تهران

مقدمه

- طراحی آزمایش برای ضمانت کیفیت یک مطالعه‌ی زیست پزشکی امری ضروری است.
- محاسبه‌ی حجم نمونه جزء ضروری این طراحی محسوب می‌شود.
- محقق بایستی پیش از شروع مطالعه حجم نمونه را تعیین و دلیل قانع کننده‌ای برای آن داشته باشد.

هدف از محاسبه‌ی حجم نمونه

- تعیین **تعداد بهینه‌ی** شرکت کنندگان (انسان یا حیوان) برای گنجاندن در کارآزمایی است.
- نمونه‌هایی با حجم خیلی کوچک یا خیلی بزرگ **هم از نظر اقتصادی** و **هم از نظر اخلاقی** توجیه پذیر نیستند.
- هر دو حالت، حجم‌های نمونه‌ی بالا و حجم‌های نمونه‌ی ناکافی، سبب خواهند شد علاوه بر اتلاف وقت و انرژی، خطرات غیرضروری بر شرکت کنندگان در مطالعه تحمیل شود.
- در نتیجه باید سعی گردد **کوچکترین حجم نمونه‌ای که در عین حال توان کافی را برای یافتن تفاوت ها و یا ارتباط ها دارد**، محاسبه و مطالعه بر اساس آن هدایت شود.

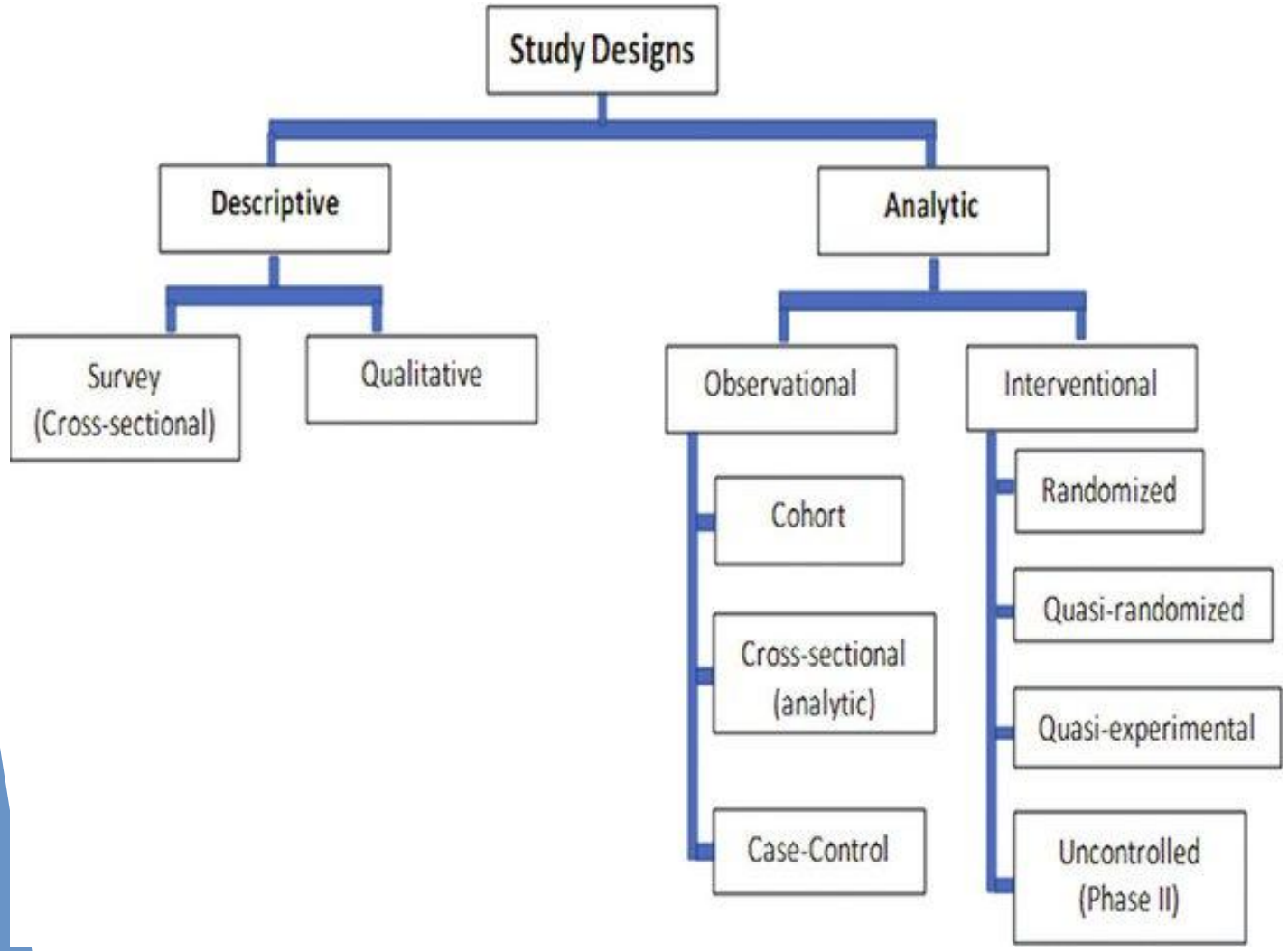
محاسبه‌ی حجم نمونه

- هرچند محاسبه‌ی حجم نمونه زیرشاخه ای از علم آمار است اما محققانی که پژوهش‌های زیست پزشکی را انجام می‌دهند بایستی در برنامه‌ریزی حجم نمونه مشارکت داشته باشند تا به کمک تخصص و دانش آنها از ادبیات کار، به تعیین درست حجم نمونه پرداخته شود.

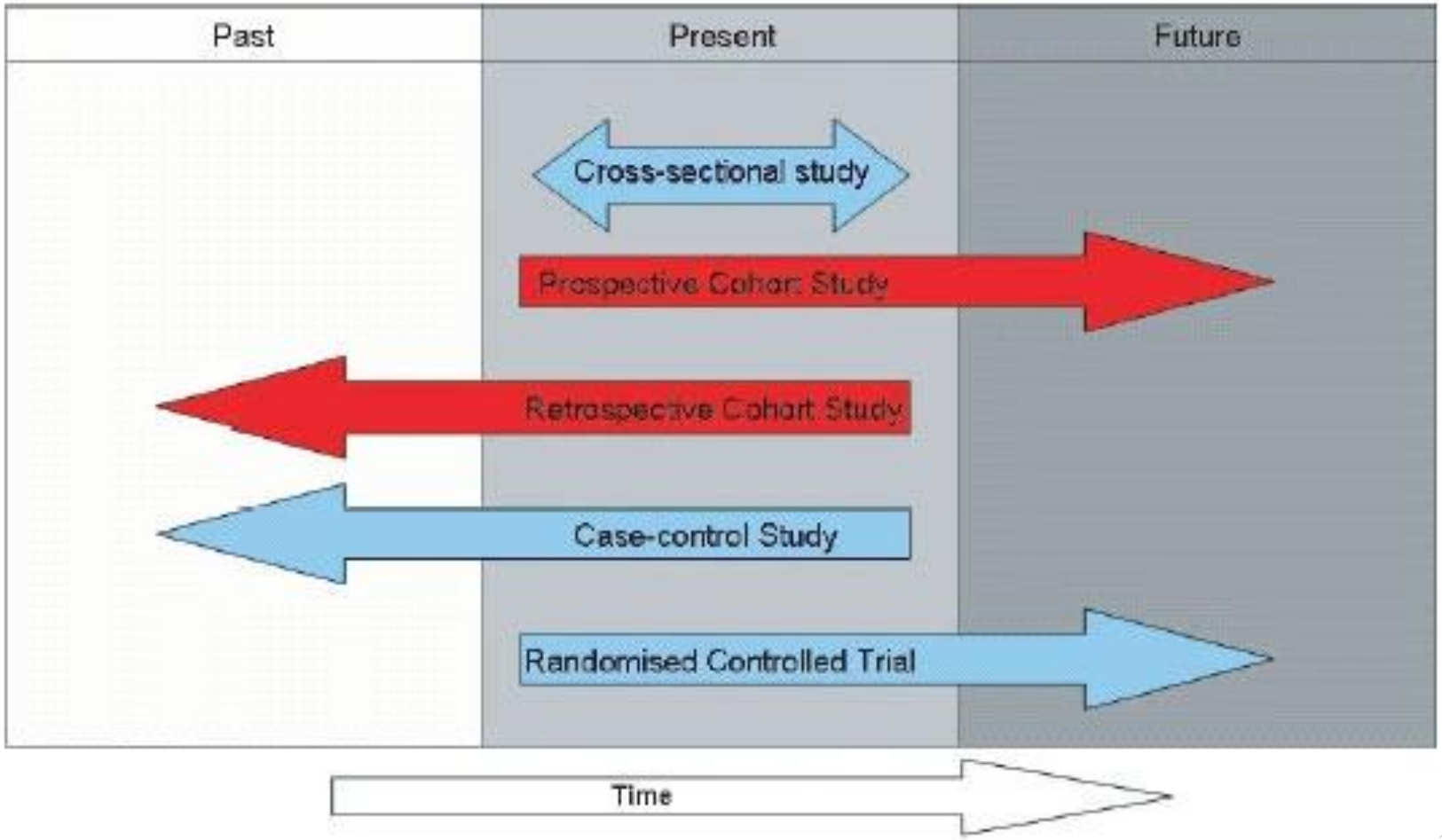
عوامل موثر در تعیین حجم نمونه:

۱. نوع مطالعه

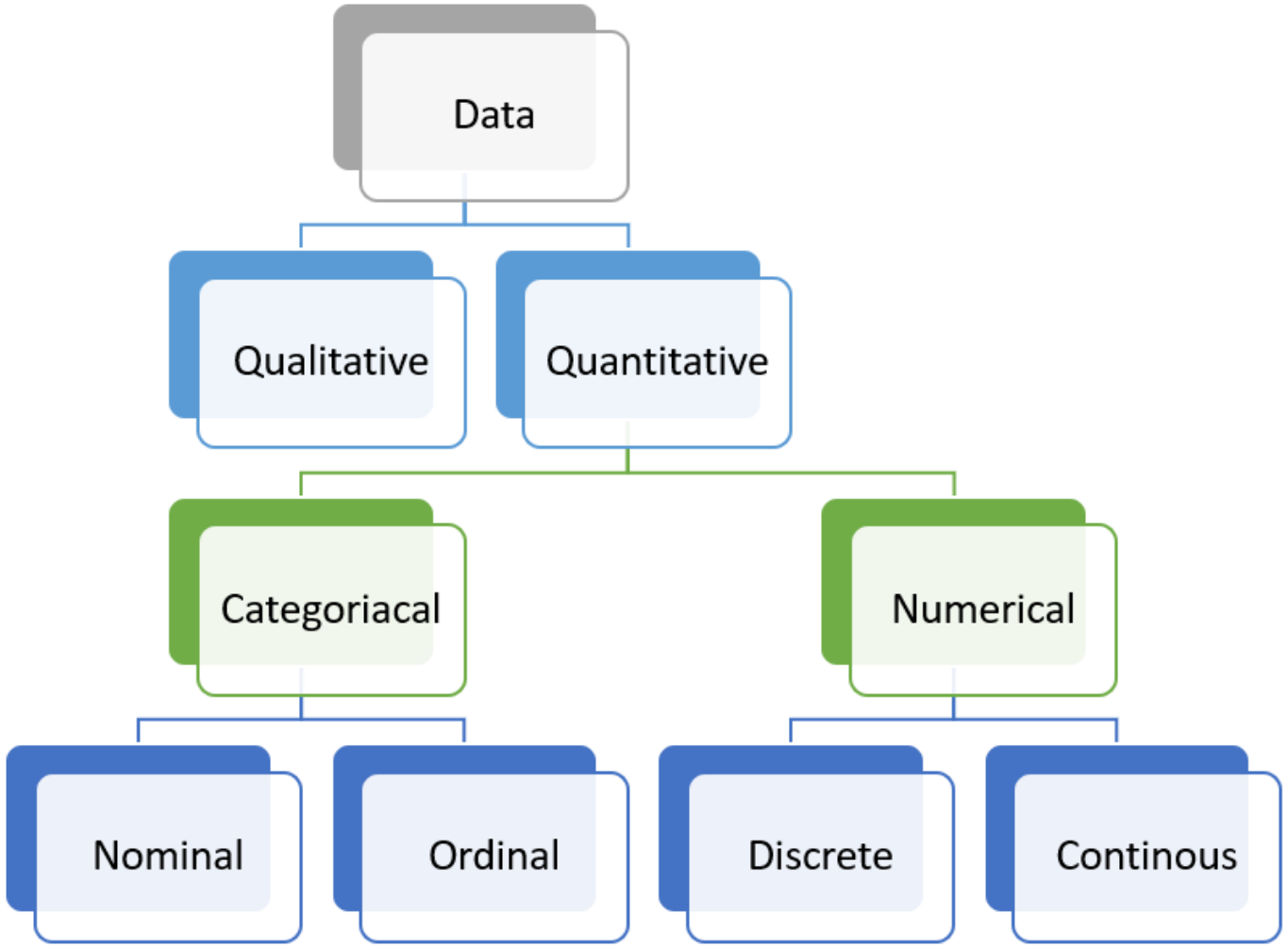
- توصیفی؟
- مشاهده‌ای؟
- تحلیلی؟
- .....



نوع مطالعه


Past
Present
Future
Cross-sectional study
Prospective Cohort Study
Retrospective Cohort Study
Case-control Study
Randomised Controlled Trial
Time

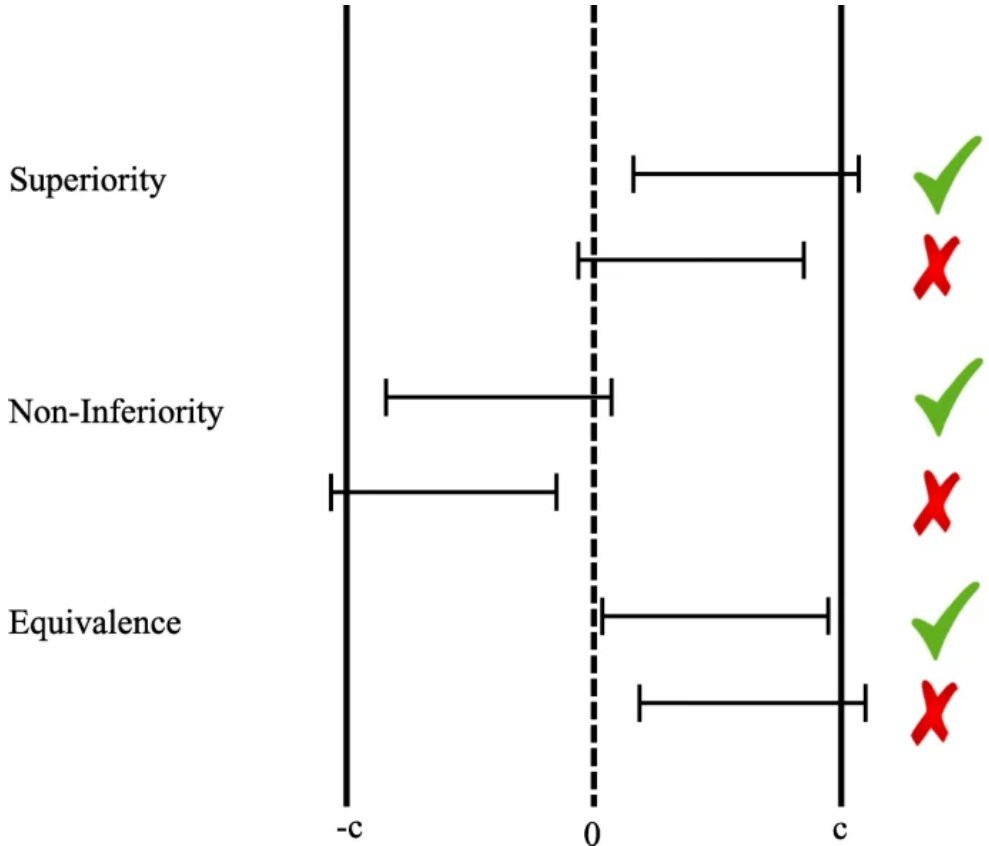
۲. نحوه اندازه‌گیری متغیرهای اصلی



- آیا اندازه گیری ویژگی مورد نظر به صورت دو وجهی، رتبه‌ای و یا به صورت متغیرهای پیوسته؟

۳. جهت مطالعه

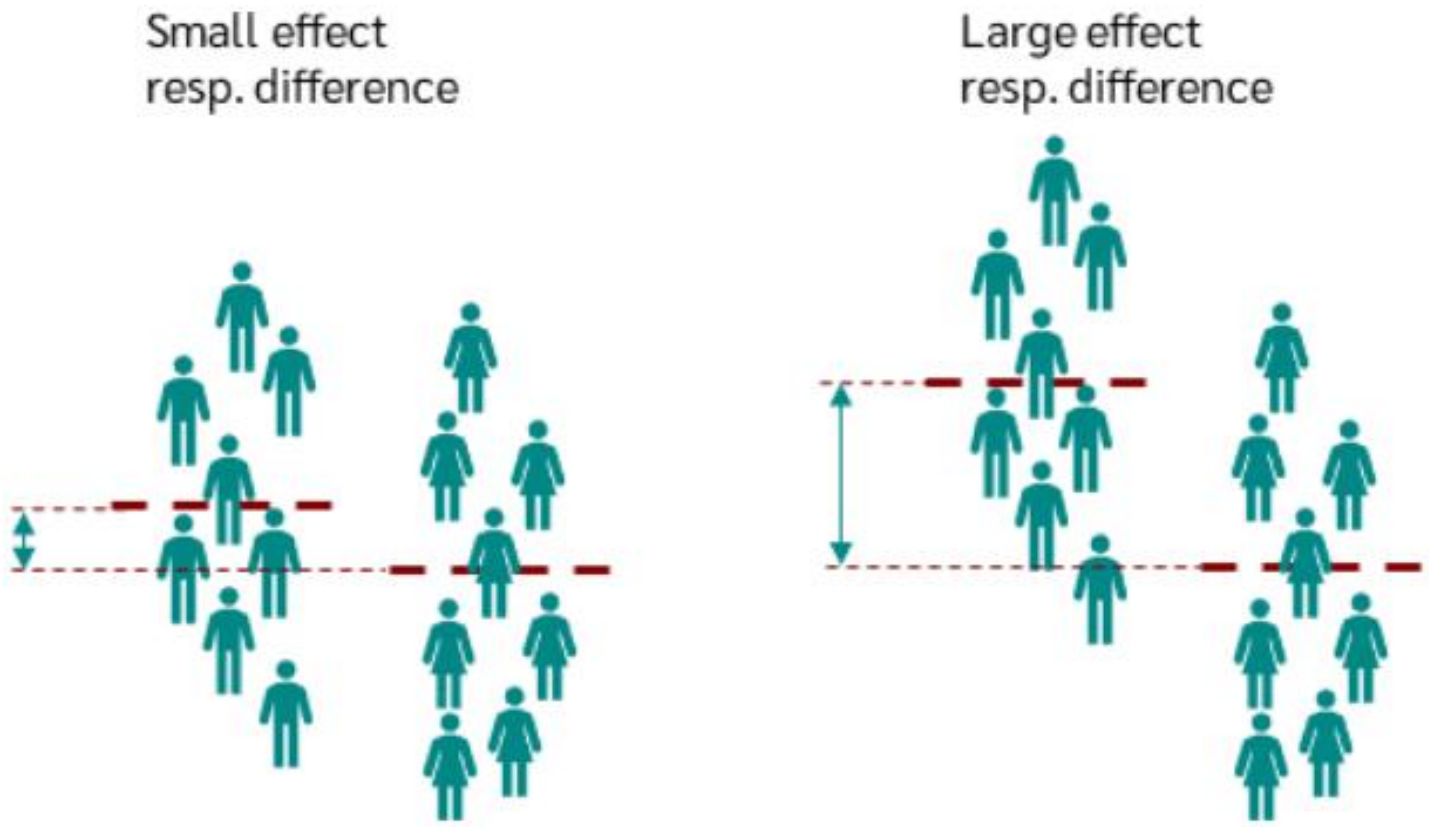
- آیا محقق از قبل فرضیه برتری، کم اثرتربودن و هم ارزی یک روش درمانی را در ذهن خود دارد یا خیر؟



۴. تعیین اندازه تاثیر (size effect)

- حداقل چه میزان از تغییرات، یا تفاوت از نظر بالینی می تواند اهمیت داشته باشد؟
- در مطالعاتی که به دنبال تعیین تفاوت‌ها یا میزان تاثیر یک مداخله می باشد، تاثیر مستقیم خواهد داشت.
- هرچه این عدد بزرگتر باشد، سبب خواهد شد که به حجم نمونه پایین‌تری نیاز باشد و برعکس.

فرض کنید محققی می‌خواهد تاثیر یک داروی ضد کلسترول را بر میزان کلسترول خون مشخص نماید. اگر او حداقل تغییر در کلسترول که اهمیت بالینی خواهد داشت را برابر ۵ میلی‌گرم دردسی لیتر فرض نماید به حجم نمونه بیشتری نیاز خواهد داشت، تا این که حداقل تغییر را برابر ۱۰ میلی‌گرم در دسی لیتر اختیار کند.



۵. واریانس (variance)

- هر چه واریانس متغیری که می‌خواهید اندازه گیری شود، بیشتر باشد، نیاز به حجم نمونه بالاتری برای اندازه گیری و برآورد دقیق‌تر متغیر مورد نظر خواهید داشت
- برآورد واریانس را یا از مطالعات گذشته که برروی جمعیت‌های مشابه صورت گرفته یا مطالعه‌ای پایلوت ترتیب داده شود و با استفاده از برآوردی که از آن به دست خواهد آمد، حجم نمونه تعیین گردد.

۶. احتمال خطای نوع اول

- **α، سطح معناداری** (significance level، Type I Error)
- احتمال رد فرض صفر، علی‌رغم صحت آن)نرخ مثبت کاذب، اگر فرض صفر را منفی بودن نتیجه آزمایش بگیریم(

  α کمتر ⟸ حجم نمونه مورد نیاز بیشتر

- معمولا در مطالعات پزشکی α را ۱ یا ۵ درصد در نظر می‌گیرند.

۷. احتمال خطای نوع دوم (مرتبط با توان آزمون)

(Type II Error، Power of Test)

- این خطا موقعی رخ می‌دهد که علی‌رغم اشتباه بودن فرضیه صفر، ما آن را رد نمی‌کنیم (نرخ منفی کاذب، اگر فرض صفر را منفی بودن نتیجه آزمایش بگیریم).
- این نوع خطا یک منهای توان مطالعه می‌باشد (توان: قدرت مطالعه در کشف یک تفاوت معنی دار).
- هر چه حجم نمونه بالاتر باشد این توان افزایش خواهد یافت.
- پس اگر بخواهیم خطای نوع دوم را در مطالعه خود کاهش دهیم (یا قدرت مطالعه را بالا ببریم) به حجم نمونه بالاتری نیاز خواهیم داشت.

		Actual	
		Positive	Negative
Predicted	Positive	**True Positive**	**False Positive**
	Negative	**False Negative**	**True Negative**

Hypothesis testing

	Truth In the population		
		H_o (False)	H_o(True)
Results in the study	Reject hypothesis	Correct	Type I error
	Accept hypothesis	Type II error	correct

تعیین حجم نمونه براساس نوع مطالعه

- برآورد یک میانگین یا نسبت
- مقایسه دو میانگین
- مقایسه چند میانگین
- طراحی های قبل و بعد و متقاطع (هر گروه، هر دو مداخله را دریافت میکند، هر فرد در مطالعه می‌تواند به عنوان شاهد خود محسوب گردد.)
- مقایسه دو نسبت
- مطالعات مورد شاهدی (یافتن عوامل خطر موثر بر بیماریها)
- مطالعات کوهورت (رابطه علیتی بین عوامل خطر و پیامدها)

تعیین حجم نمونه براساس اهداف مطالعه

- مطالعات همبستگی
- رگرسیون خطی
- رگرسیون لوژستیک
- داده‌های طبقه‌ای رتبه بندی شده
- طرح های خوشه‌ای
- تحلیل عاملی
- تحلیل بقا
- آنالیز واریانس یکطرفه و چند طرفه

حجم نمونه برای برآورد یک نسبت در جمعیت

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

p: برآورد نسبت در جامعه می باشد. (۱) از مطالعات گذشته (۲) یک مطالعه پایلوت (۳) گاهی دو راه مذکور امکانپذیر نیست که آن را برابر ۵۰% (متوسط تخمینی از صفر تا ۱۰۰ درصد) فرض می کنیم.

d: میزان دقت است. "دقت" این سئوال را مطرح می کند: در برآورد نسبت چقدر اشتباه جایز است؟ یا محقق چند درصد اشتباه در برآورد شیوع را می پذیرد؟

مثال:

محققی قصد دارد برآوردی از شیوع سنگ کیسه صفرا در کشور داشته باشد. در گذشته این شیوع در جمعیتی مشابه برابر ۰/۸ درصد برآورد شده است. برای هدایت این مطالعه به چند نمونه نیاز است اگر محقق خطای نوع اول را پنج درصد در نظر بگیرد؟

پاسخ:

اطلاعات زیر در اختیار ماست:

$$\alpha = 0.05 \Rightarrow z_{1-\alpha/2} = 1.96 \qquad P = 0.08 \qquad d = 0.1P = 0.1 \times 0.08 = 0.008$$

بر اساس رابطه ۲.۱ خواهیم داشت:

$$n = \frac{z_{1-\alpha/2}^2 \times p(1-p)}{d^2} = \frac{1.96^2 \times 0.008 \times 0.992}{0.0008^2} = 47616$$

حجم نمونه برای برآورد یک میانگین در جمعیت

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

δ: دقت است. یعنی تا چه حد اشتباه، جایز و از نظر بالینی تاثیرگذار نیست. به عنوان مثال در اندازه‌گیری فشار خون، چقدر تفاوت را می‌توان نادیده گرفت و اهمیتی ندارد؟ اگر یک میلی‌متر جیوه فشار خون بالاتر یا پایین‌تر برآورد شود، می‌تواند بر پیامدِ بیماری تاثیری بگذارد؟ در واقع حداقل تفاوتی که می‌تواند از نظر بالینی با اهمیت تلقی شود(و نباید در برآوردها، اشتباه بیشتر از آن باشد) را دقت یا همان δ گویند. این را محقق تعیین می‌کند.

مثال:

محققی می‌خواهد میانگین فشار خون سیستولیک زنان ۱۸ سال و بالاتر را در شمال ایران تخمین بزند. بر اساس مطالعه‌ای که در گذشته برروی همین جمعیت صورت گرفت برآورد فشار خون برابر 115.42 ± 17.60 بدست آمد. برای هدایت این مطالعه محقق به چند نمونه نیاز خواهد داشت، اگر دقت را در این مطالعه ۲ میلی‌متر جیوه در نظر گرفته باشد (البته با خطای آلفای یک درصد)؟

پاسخ:

داده‌هایی که از مثال فوق در اختیار ماست عبارتند از:

$$\alpha = 0.01 \Rightarrow z_{1-\alpha/2} = 2.57 \qquad \sigma^2 = 17.6^2 \qquad d = 2$$

در نتیجه حجم نمونه مورد نیاز بر اساس رابطه ۲.۲برابر خواهد بود با:

$$n = \frac{z_{1-\alpha/2}^2 \times \sigma^2}{\delta^2} = \frac{2.57^2 \times 17.6^2}{2^2} = 511.48 = 512$$

تعیین حجم نمونه برای مقایسه میانگین‌های دو گروه

- الف) تعیین حجم نمونه برای مقایسه میانگین‌های دو گروه با واریانس‌های برابر

$$\delta = \mu_A - \mu_B$$

φ: نسبت حجم نمونه در گروه B به گروه A است.

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

$$n_B = \varphi n_A$$

- ب) تعیین حجم نمونه برای مقایسه میانگین‌های دو گروه با واریانس‌های نابرابر

- دانلود رایگان نرم افزار R
  - https://cran.r-project.org/

محاسبه حجم نمونه در R

- پکیج pwr

- install.packages("pwr") نصب پکیج

- library("pwr") فراخوانی

حجم نمونه برای برآورد یک نسبت

- pwr.p.test(h =، sig.level =، power =، alternative =)

h	Effect size
sig.level	Significance level (Type I error probability)
power	Power of test (1 minus Type II error probability)
alternative	a character string specifying the alternative hypothesis، must be one of "two.sided" (default)، "greater" or "less"

Effect size اندازه تاثیر

0.2=small, 0. 5=medium, and 0.8 large effect sizes

Effect size calculation

- h= 2*asin(sqrt(p_1))-2*asin(sqrt(p_2))
  - p_1=proportion 1
  - p_2=proportion 2

Results:

```
> #sample number
> pwr.p.test(h=0.2, sig.level=0.05, power=0.80, alternative="two.sided")

     proportion power calculation for binomial distribution (arcsine transformation)

              h = 0.2
              n = 196.2215
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

→ Round up to 197

مثال ١

You are interested in determining if the male incidence rate proportion of cancer in North Dakota is higher than the US average (prop=0.00490). You find trial data cancer prevalence of 0.00495.

- h= 2*asin(sqrt(0.00495))-2*asin(sqrt(0.00490))=0.0007
- pwr.p.test(h=0.0007, sig.level=0.05, power=0.80, alternative="greater")
- n = 12617464 -> 12,617,464 samples

مثال ۲

You are interested in determining if the female incidence rate proportion of cancer in North Dakota is lower than the US average (prop=0.00420).

- Guess a very low effect size (0.001)
- pwr.p.test(h=-0.001, sig.level=0.05, power=0.80, alternative="less")
- n = 6182557 -> 6,182,557 samples

براورد یک میانگین

- pwr.t.test(d=,sig.level=,power=,type="one-sample",alternative=)

d=effect size
sig.level=significant level
power=power of test
type=type of test

For t-tests:
0.2=small, 0.5=medium, and 0.8 large effect sizes

Results:

```
> #sample number
> pwr.t.test(d=0.50, sig.level=0.05, power=0.80, type="one.sample", alternative="two.sided")

     One-sample t test power calculation

              n = 33.36713
              d = 0.5
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

→ Round up to 34

مثال ١

You are interested in determining if the average sleep time change in a year for college freshman is different from zero. You collect the following data of sleep change (in hours).

Sleep Change	-0.55	0.4	2.6	0.65	-0.23	0.21	-4.3	2	-1.7	-1.8

- Effect size $=(Mean_{H1}-Mean_{H0})/SD$
- Two-tailed test
- **pwr.t.test(d= -0.137, sig.level=0.05, power=0.80, type="one.sample", alternative="two.sided")**

حجم نمونه برای مقایسه میانگین‌های دو گروه

- pwr.t.test(d=,sig.level=,power=,type="two.sample",alternative=)

For t-tests:
0.2=small, 0.5=medium, and 0.8 large effect sizes

Results:

```
> #sample number
> pwr.t.test(d=0.5, sig.level=0.05, power=0.80, type="two.sample", alternative="greater")
```

Two-sample t test power calculation

n = 50.1508 → Round up to 51, per group

d = 0.5

sig.level = 0.05

power = 0.8

alternative = greater

NOTE: n is number in *each* group

مثال ١

You are interested in determining if the average protein level in blood different between men and women. You collected the following trial data on protein level (grams/deciliter).

Male Protein	1.8	5.8	7.1	4.6	5.5	2.4	8.3	1.2
Female Protein	9.5	2.6	3.7	4.7	6.4	8.4	3.1	1.4

- Effect size = $(Mean_{H1}-Mean_{H0})/ SD_{pooled} =(4.59-4.98)/ \sqrt{((2.58^2+ 2.88^2)/2)} = -0.14$
- two-tailed test
- **pwr.t.test(d=-0.14, sig.level=0.05, power=0.80, type="two.sample", alternative="two-sided")**
- n = 801.87 -> **802 samples per group**

مثال ٢

You are interested in determining if the average glucose level in blood is lower in men than women

- Guessed a small effect (0.20), then used a one-tailed test
- **pwr.t.test(d=-0.20, sig.level=0.05, power=0.80, type="two.sample", alternative="less")**
- n = 309.8 -> **310 samples per group**

مقایسه میانگین بیش از دو گروه

- pwr.anova.test(k = ، f = ، sig.level =، power =)

k	Number of groups
f	Effect size
sig.level	Significance level (Type I error probability)
power	Power of test (1 minus Type II error probability)

For f-tests:
0.1=small, 0.25=medium, and 0.4 large effect sizes

مثال ١

You are interested in determining there is a difference in weight lost between 4 different surgery options. You collect the following trial data of weight lost in pounds (shown on right)

- $\eta^2 = SS_{treat} / SS_{total} = 37.12/(37.12+57.22) = 0.39$
- $f = \sqrt{((0.39/(1-0.39)} = 0.81$
- 4 groups
- pwr.anova.test(k =4 , f =0.81, sig.level=0.05 , power =0.80)
- n = 5.29 -> 6 samples per group (24 total)

Option 1	Option 2	Option 3	Option 4
6.3	9.9	5.1	1.0
2.8	4.1	2.9	2.8
7.8	3.9	3.6	4.8
7.9	6.3	5.7	3.9
4.9	6.9	4.5	1.6

مثال ۲

You are interested in determining if there is a difference in white blood cell counts between 5 different medication regimes.

- Guessed a medium effect size (0.25)
- 5 groups
- **pwr.anova.test(k =5 , f =0.25 , sig.level=0.05 , power =0.80)**
- n = 39.15 -> **40 samples per group (200 total)**

#	Name of Test	in R?	Package	Function
1	One Mean T-test	Yes	pwr	pwr.t.test
2	Two Means T-test	Yes	pwr	pwr.t.test
3	Paired T-test	Yes	pwr	pwr.t.test
4	One-way ANOVA	Yes	pwr	pwr.anova.test
5	Single Proportion Test	Yes	pwr	pwr.p.test
6	Two Proportions Test	Yes	pwr	pwr.2p.test
7	Chi-Squared Test	Yes	pwr	pwr.chisq.test
8	Simple Linear Regression	Yes	pwr	pwr.f2.test
9	Multiple Linear Regression	Yes	pwr	pwr.f2.test
10	Correlation	Yes	pwr	pwr.r.test
11	One Mean Wilcoxon Test	Yes*	pwr	pwer.t.test + 15%
12	Mann-Whitney Test	Yes*	pwr	pwer.t.test + 15%
13	Paired Wilcoxon Test	Yes*	pwr	pwer.t.test + 15%
14	Kruskal Wallace Test	Yes*	pwr	pwr.anova.test + 15%
15	Repeated Measures ANOVA	Yes	WebPower	wp.rmanova
16	Multi-way ANOVA (1 Category of interest)	Yes	WebPower	wp.kanova
17	Multi-way ANOVA (>1 Category of interest)	Yes	WebPower	wp.kanova
18	Non-Parametric Regression (Logistic)	Yes	WebPower	wp.logistic
19	Non-Parametric Regression (Poisson)	Yes	WebPower	wp.poisson
20	Multilevel modeling: CRT	Yes	WebPower	wp.crt2arm/wp.crt3arm
21	Multilevel modeling: MRT	Yes	WebPower	wp.mrt2arm/wp.mrt3arm
22	GLMM	Yes^	Simr & lme4	n/a

***-parametric test with non-parametric correction**
^-detailed in future Module

منابع پیشنهادی:

- نیما معتمد، فرهاد زمانی (۱۳۹۵)، *حجم نمونه در تحقیقات پزشکی رویکردی کاربردی*، انتشارات عصر روشن‌بینی.

- Shein-Chung Chow، Jun Shao، Hansheng Wang، Yuliya Lokhnygina (2020) *Sample Size Calculations in Clinical Research*،3rd Edition، Chapman and Hall/CRC.

- Sample Size Calculation with R

  Dr. Mark Williamson، Statistician

  Biostatistics، Epidemiology، and Research Design Core