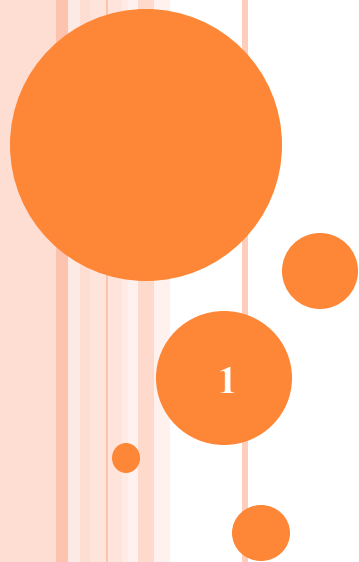


مفاهيم آماری

جلسه دوم



در پایان این بخش دانشجو باید بتواند:

- اهمیت علم آمار را با ذکر مثال توضیح دهد.
- تفاوت بین آمار توصیفی و استنباطی را بیان کند.
- متغیرهای کمی و کیفی را شناسایی کند.
- تفاوت مقیاسهای اندازه‌گیری را درک کرده باشد.
- تفاوت جامعه و نمونه را بداند.



چرا آمار؟

○ زیرا پاسخ دادن به سوالات و فرضیه ها مستلزم جمع آوری اصولی داده، خلاصه سازی آن و در نهایت انتخاب آزمون آماری درست میباشد. تنها در این صورت سیاستگذارو یا محقق میتواند تصمیم درست بگیرد.



مقدمه

آمار علم جمع آوری، سازماندهی و خلاصه کردن داده و در نهایت استنباط و نتیجه گیری درباره مجموعه ای از مشاهدات میباشد.

علم آمار کاربرد فراوانی در رشته های گوناگون علوم مانند: بازرگانی، آموزش و پرورش، روانشناسی، کشاورزی، اقتصاد، مهندسی، پزشکی، بهداشت و ... دارد. اگر در تجزیه و تحلیل آماری از داده های موجود در علوم زیستی و پزشکی استفاده کنیم، در این مورد اصطلاح آمار زیستی را بکار می بریم.

آمار

```
graph TD; A[آمار] --> B[آمار استنباطی]; A --> C[آمار توصیفی]; B --> D[آمار توصیفی]; B --> E[آزمونهای آماری]; C --> F[جداول و نمودارها]; C --> G[خلاصه سازی داده]
```

آمار استنباطی

آمار توصیفی

آزمونهای آماری

آمار توصیفی

جداول و نمودارها

خلاصه سازی داده

آمار توصیفی

شامل

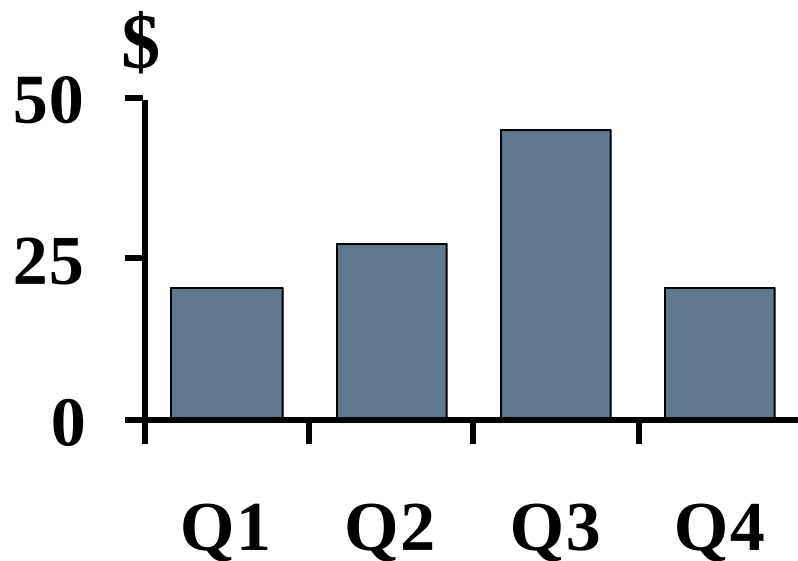
○ جمع آوری داده

○ خلاصه سازی داده

○ بیان ویژگیهای داده

هدف

○ توصیف داده



$$\bar{X} = 30.5 \quad S^2 = 113$$

آمار استنباطی

Population
?



شامل

○ برآورد

○ آزمون فرض

هدف

○ تصمیم گیری در مورد ویژگی جامعه مورد نظر

متغیرها

○ **متغیر:** هر گاه در مورد یک ویژگی دریابیم که آن ویژگی در رابطه با اشخاص، اماکن و یا اشیای گوناگون مقادیر مختلف را بخود می گیرد، آن ویژگی را متغیر می نامیم.

○ به عبارت دیگر، مشخصه ای که از یک واحد آماری به یک واحد آماری دیگر امکان تغییر دارد را متغیر می نامند.

○ **مثال:** قد مردان بالغ، وزن کودکان کودکستانی، سن بیماران مراجعه کننده به درمانگاه دندان پزشکی

انواع مقیاس ها

○ اسمی (Nominal)

○ رتبه ای (Ordinal)

○ فاصله ای (Interval)

○ نسبتی (Ratio)

انواع مقیاس ها...

○ اسمی (Nominal)

- بر اساس این مقیاس سنجش، تنها می توان نامی را برای فرد مورد مطالعه در رابطه با خاصیت مورد نظر انتخاب کرد.
- بدیهی است که بر اساس این سنجش نمی توان اندازه خاصیتی را در فردی از فرد دیگر کمتر یا بیشتر دانست.
- مثال : در مطالعه جنس و یا گروه خونی می توان شخص مورد نظر را تحت عنوان مرد یا زن و گروه خونی A، B، AB یا O مشخص کرد.

انواع مقیاس ها...

○ رتبه ای (Ordinal)

- در پاره ای از مشاهدات می توان نتیجه سنجش یک خاصیت را با بیان رتبه ی فرد یا شیئی در رابطه با سایر افراد بیان کرد.
- مثلاً برای سنجش یک بیماری ممکن است براساس ضوابطی فرد را در یکی از سه گروه خفیف، متوسط یا شدید قرار داد.
- بدیهی است که در این گونه موارد نمی توان گفت که شدت بیماری در گروه متوسط دو برابر گروه خفیف است.
- عبارت دیگر نمی توان فاصله ی دو رتبه متمایز و یا نسبت اندازه ی بین آنها را مشخص کرد ولی بهرحال این نوع صفت از خاصیت صفت اسمی برخوردار می باشد.

انواع مقیاس ها...

○ فاصله ای (Interval)

○ دارای مبدا قراردادی است

○ ولی نظر به اینکه این نوع اندازه گیری فاقد صفر ذاتی می باشد نمی توان نسبت اندازه خاصیت مورد مطالعه را در افراد مشخص کرد به همین دلیل این نوع اندازه ها را اندازه های فاصله ای می نامند.

○ مثلا برای اندازه گیری حرارت اشیا از درجات معین و ثابتی (سانتی گراد یا فارنهایت) استفاده می کنیم ولی به دلیل عدم وجود صفر حقیقی نمی توان مثلا مقدار حرارت ۸۰ درجه سانتی گراد را دو برابر حرارت جسمی با ۴۰ درجه سانتی گراد دانست..

انواع مقیاس ها...

○ نسبتی (Ratio)

○ بعضی از سنجش مشاهدات اندازه هایی هستند که بر اساس آن نه تنها می توان فرد مورد مطالعه را به رده ای منتسب کرد یا رتبه آن را معین نمود و یا فاصله آن را در رابطه با افراد دیگر معین کرد، بلکه می توان نسبت اندازه خاصیت مورد مطالعه را در دو فرد مشخص کرد زیرا برخلاف اندازه های فاصله ای این اندازه ها از صفر ذاتی برخوردار می باشند.

○ مثلاً در مورد طول یک شیء مفهوم ذاتی صفر کاملاً روشن است و بدین ترتیب طول یک شیء ۸۰ سانتیمتری دو برابر شیء ۴۰ سانتیمتری می باشد.

بهتر است به خاطر داشته باشیم:

- معمولاً اندازه های اسمی و رتبه ای را اطلاعات کیفی و اندازه های فاصله ای و نسبی را اطلاعات کمی نامگذاری می کنند.

انواع متغیرها

متغیرهای کمی :

متغیرهایی هستند که قابل اندازه گیری می باشند. مثل قد، وزن، سن و ...

متغیرهای کیفی:

متغیرهایی که قابل اندازه گیری نیستند ولی می توان آنها را طبقه بندی نمود.
اگرچه در مورد متغیرهای کیفی، امکان اندازه گیری در مفهوم معمولی کلمه وجود ندارد، اما می توان تعداد اشخاص، اماکن و یا اشیایی که به گروه ها و یا مقوله های گوناگون متعلق هستند را شمارش کرد.

مثال: طبقه بندی تعداد بیمارانی که در یک روز با انواع گوناگون بیماری در بیمارستان پذیرفته شده اند، بر حسب نوع بیماری.

متغیر تصادفی

وقتی که قد، وزن و یا سن فردی را تعیین می کنیم، نتیجه حاصل را **مقدار** **متغیر** مذکور می نامیم.

اگر مقادیر بدست آمده در نتیجه عوامل تصادفی و شانسی بدست آیند، متغیر را **متغیر تصادفی** می نامیم. در این رابطه بیشتر اوقات، مقادیری را که با روش های اندازه گیری بدست می آیند، **مشاهدات** و یا به طور ساده، **اندازه ها** می گوئیم.

انواع متغیرهای کمی

متغیر تصادفی گسسته: هرگاه مابین مقادیر قابل تصور یک متغیر، فاصله ای وجود نداشته باشد، آن متغیر را منفصل یا گسسته می نامیم. مثال: تعداد بیماران پذیرفته شده در یک روز ، تعداد دندان های فاسد در کودکان

متغیر تصادفی پیوسته: در متغیر تصادفی متصل یا پیوسته فاصله ای مابین هیچ کدام از دو مقدار قابل تصور وجود ندارد. به عبارت دیگر مابین هر دو مقدار قابل تصور متغیر مزبور، هر مقدار دیگری را می توان فرض کرد. اندازه های وزن، قد و ... از این نوع متغیرها هستند.

جمعیت

مجموعه افراد یا چیزهایی را که می‌خواهیم یک یا چند ویژگی را درباره آنها مطالعه کنیم یک جمعیت می‌نامند.

مثال: جمعیت نوزادانی که در سال گذشته به دنیا آمده‌اند از نظر مصرف شیرخشک

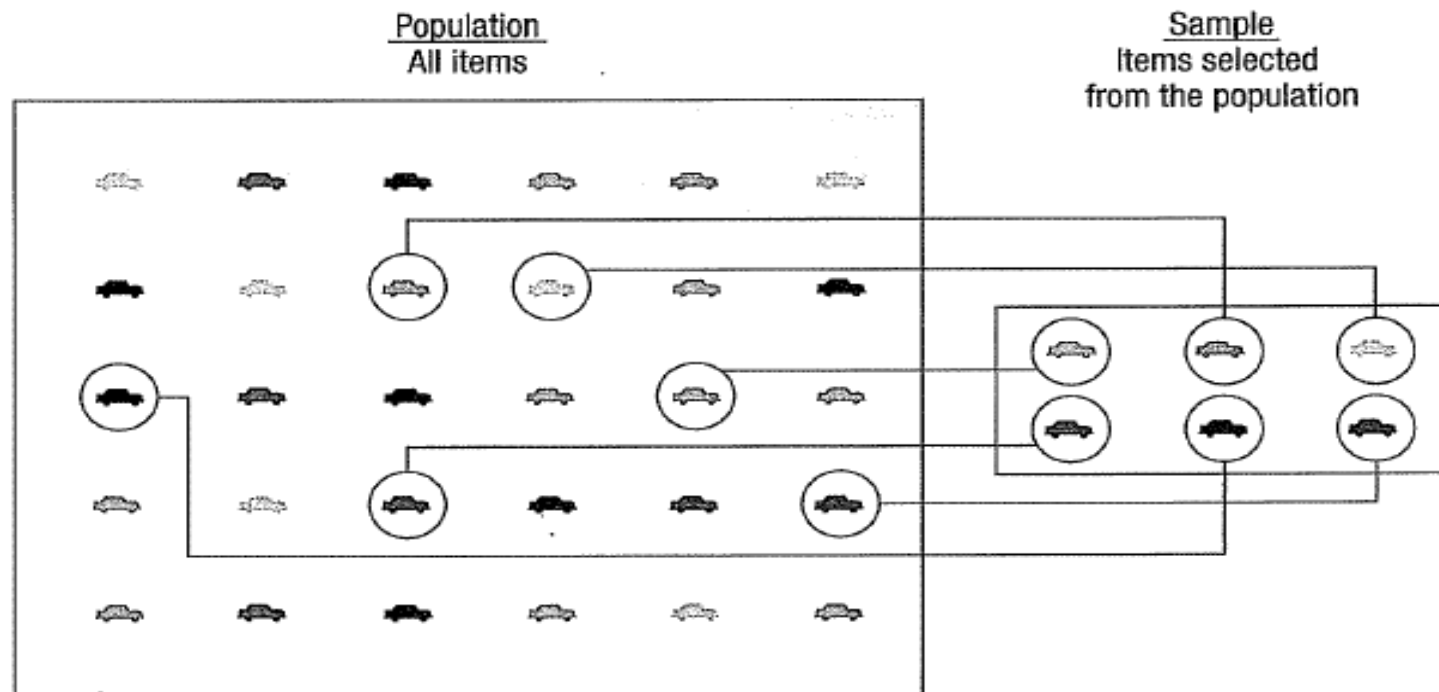
برای انجام هر کار آماری روی یک جمعیت، باید آن جمعیت و ویژگی مورد مطالعه، بدون هر گونه ابهام قبلاً مشخص شوند.

اگر جمعیت مقادیر، از تعداد ثابت و مشخصی از مقادیر تشکیل شود آن را جمعیت محدود یا با پایان می‌نامند. مثلاً جمعیت نوزادان به دنیا آمده در سال گذشته و اگر جمعیت از یک ردیف بی‌انتهای مقادیر تشکیل شده باشد آن را جمعیت نامحدود یا بی‌پایان می‌خوانیم.

گاهی جمعیت مورد مطالعه به قدری انبوه است که عملاً می‌توان آن را بی‌پایان تلقی نمود. مثل جمعیت دانه‌های گندم که از یک مزرعه برداشت می‌شود.

نمونه

در آمار، هر زیرمجموعه ای از یک جامعه را یک نمونه می گویند. (از یک جمعیت می توان نمونه های گوناگونی انتخاب نمود).
نمونه باید تصادفی و نماینده خوبی برای جامعه مورد مطالعه باشد.



خلاصه سازی داده ها



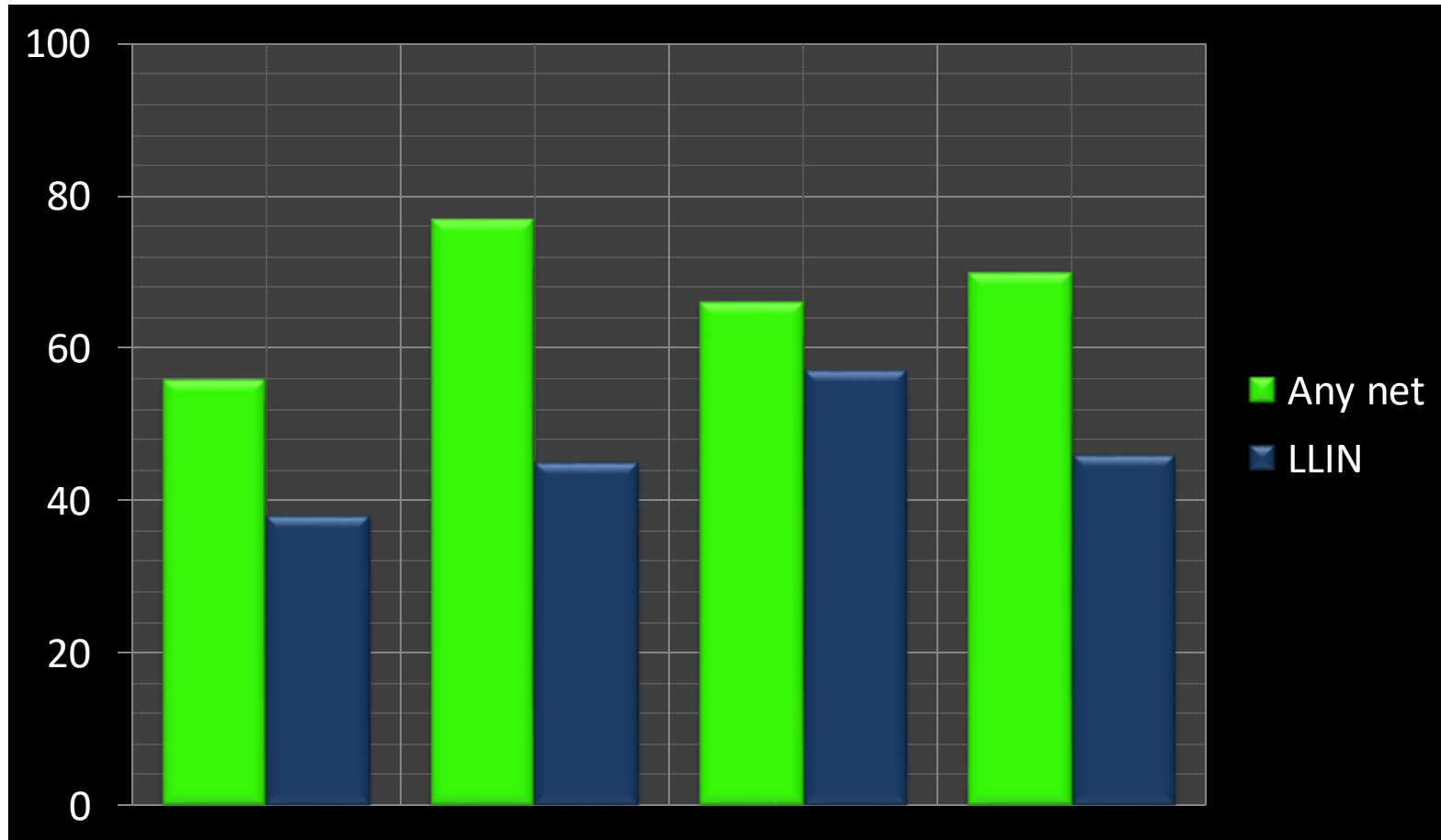
در پایان این بخش دانشجو باید بتواند:

- راه های متفاوت خلاصه سازی داده را نام ببرد.
- بتواند نمودار و یا جدول مناسب را با توجه به داده های خود انتخاب کند.
- بتواند نمودار و یا جدول گویا رسم کند.
- بتواند نمودار و یا جدول رسم شده را جذاب کند.

شما داده های خود را به این صورت معرفی میکنید؟

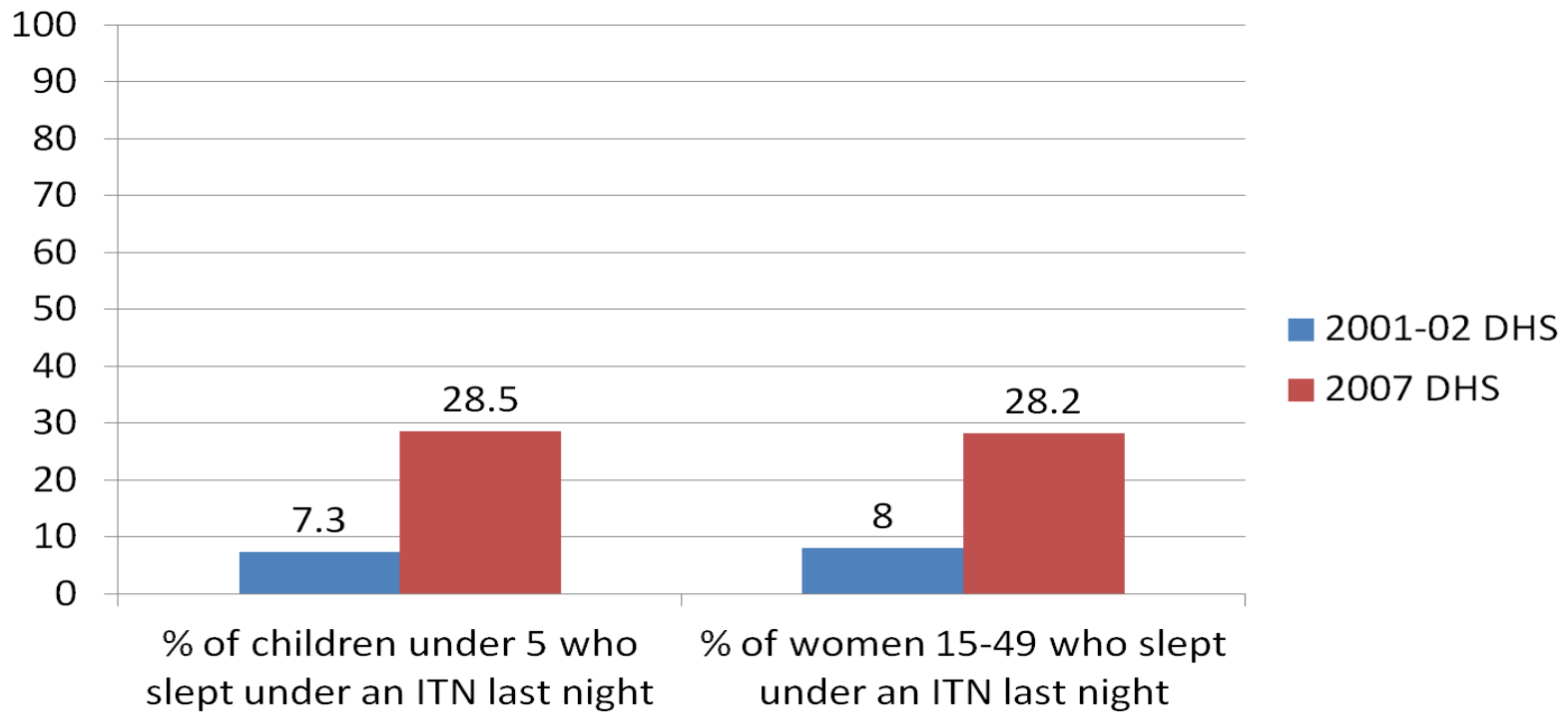
Year	Closing Value	Year	Closing Value
1980	964	1991	2633.7
1981	875	1992	3168.8
1982	1046.5	1993	3301.1
1983	1258.6	1994	3754.4
1984	1211.6	1995	5117.1
1985	1546.7	1996	6448.3
1986	1896	1997	7908.3
1987	1938.8	1998	9181.4
1988	2168.6	1999	11497.1
1989	2753.2	2000	10757.99
1990	2753.2	2001	10021.5

یا به این صورت؟



این بهتر است!

Use of ITNs in Zambia



خلاصه سازی داده ها

جداول

- ساده ترین روش برای خلاصه سازی داده ها میباشد.
- داده ها با فراوانی یا درصد فراوانی نشان داده میشوند.

نمودارها

- یک نمایش بری از داده ها میباشد.
- معمولاً داده ها با درصد بیان میشوند.

جدول فراوانی

Time Between Eruptions (seconds)	Tally	Frequency	Relative Frequency
670 – 679		2	$2/45 = 0.0444$
680 – 689		0	0
690 – 699		7	0.1556
700 – 709		9	0.2
710 – 719		9	0.2
720 – 729		11	0.2444
730 – 739		7	0.1556

جدول فراوانی...

Time Between Eruptions (seconds)	Tally	Frequency	Relative Frequency
670 – 674		1	$1/45 = 0.0222$
675 – 679		1	0.0222
680 – 684		0	0
685 – 689		0	0
690 – 694		0	0
695 – 699		7	0.1556
700 – 704		7	0.1556
705 – 709		2	0.0444
710 – 714		5	0.1111
715 – 719		4	0.0889
720 – 724		6	0.1333
725 – 729		5	0.1114
730 – 734		3	0.0667
735 – 739		4	0.0889

چند نکته:

○ اگر با یک متغیر کمی پیوسته روبرو باشیم، داده ها را به چند طبقه و با چه فاصله ی گروهی تقسیم کنیم؟

تقسیم داده ها در تعداد کمی از فواصل گروهی سبب از دست دادن اطلاعات زیادی می گردد. از سوی دیگر، اگر تعداد فواصل گروهی را زیاد کنیم، هدف ما در خلاصه کردن داده ها حاصل نمی شود.

○ در اغلب مطالعات بهتر است که فاصله گروهها یکسان انتخاب شوند با وجود این در مطالعات اپیدمیولوژی فاصله های نامساوی کاملاً متداول و مرسوم است.

دامنه تغییرات

○ طول بازه ای را که متغیر در آن تغییر می کند را دامنه تغییرات می گویند.

$$R=b-a$$

b = بزرگترین داده

a = کوچکترین داده

▶ بزرگی دامنه تغییرات نشان دهنده تفاوت زیاد در جامعه است.

▶ هرچقدر این دامنه کمتر باشد، یعنی افراد جامعه از لحاظ این متغیر به هم نزدیکترند.

▶ اگر دامنه تغییرات صفر باشد، تمام افراد با هم برابر و یکسانند.

مرکز و طول دسته

- برای دسته $[a_i, b_i)$ ، a_i را کران پایین و b_i را کران بالا و x_i را مرکز آن دسته می نامیم.

$$X_i = (a_i + b_i) / 2$$

- تفاضل دو کران پایین متوالی یا دو کران بالای متوالی را طول آن دسته می نامیم.

مثال

داده های زیر مربوط به ضریب هوشی ۳۴ دانش آموز پسر سال اول دبیرستان به شرح زیر می باشد:

۱۰۵	۱۰۵	۱۰۳	۹۱	۸۸	۱۰۱	۱۰۸	۱۰۳	۱۱۱	۱۱۱	۱۰۸	۱۰۸	۸۸
۱۰۵	۹۸	۹۵	۸۹	۹۲	۱۱۲	۱۱۲	۱۰۰	۹۲	۹۶	۱۰۷	۱۱۱	۹۶
				۷۷	۹۰	۱۰۶	۷۷	۱۰۸	۹۵	۱۱۱		

$$R = 112 - 77 = 35 \quad \text{دامنه تغییرات برابر است با :}$$

با تقسیم این دامنه به ۷ دسته ، دسته های حاصل برابرند با :

$[77, 82)$, $[82, 87)$, $[87, 92)$, $[92, 97)$, $[97, 102)$, $[102, 107)$, $[107, 112]$

انواع فراوانی

- **فراوانی مطلق:** فراوانی مطلق دسته i ام را که با f_i نشان می دهیم برابر تعداد اعضایی است که در دسته i ام قرار گرفته اند.
- **فراوانی نسبی:** اگر f_i فراوانی دسته i ام و تعداد داده ها n باشد کسر f_i/n را فراوانی نسبی دسته i ام می گوییم.
- **فراوانی تجمعی:** فراوانی تجمعی هر دسته برابر تعداد اشیائی است که مقدار آن ها از کران بالای آن دسته کمتر اند.

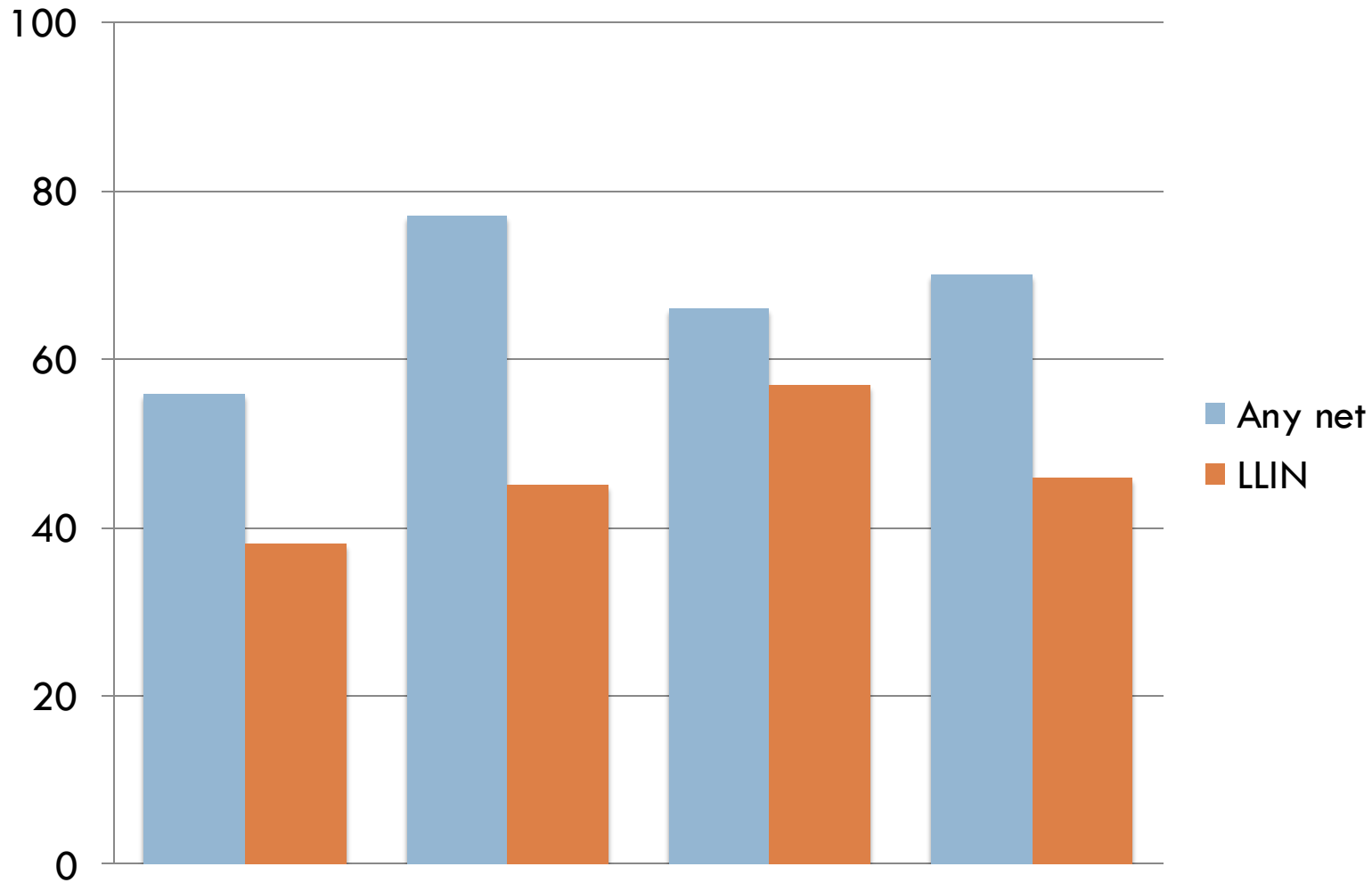
مثال

حدود دسته ها	مرکز دسته ها	فراوانی مطلق f_i	فراوانی نسبی	فراوانی تجمعی
77 - 82	79/5	2	0/059	2
82 - 87	84/5	0	0	2
87 - 92	89/5	5	0/147	7
92 - 97	94/5	6	0/176	13
97-102	99/5	1	0/03	14
102-107	104/5	8	0/235	22
107-112	109/5	12	0/353	34
جمع		34	%100	

راهنمای انتخاب درست نمودارها:

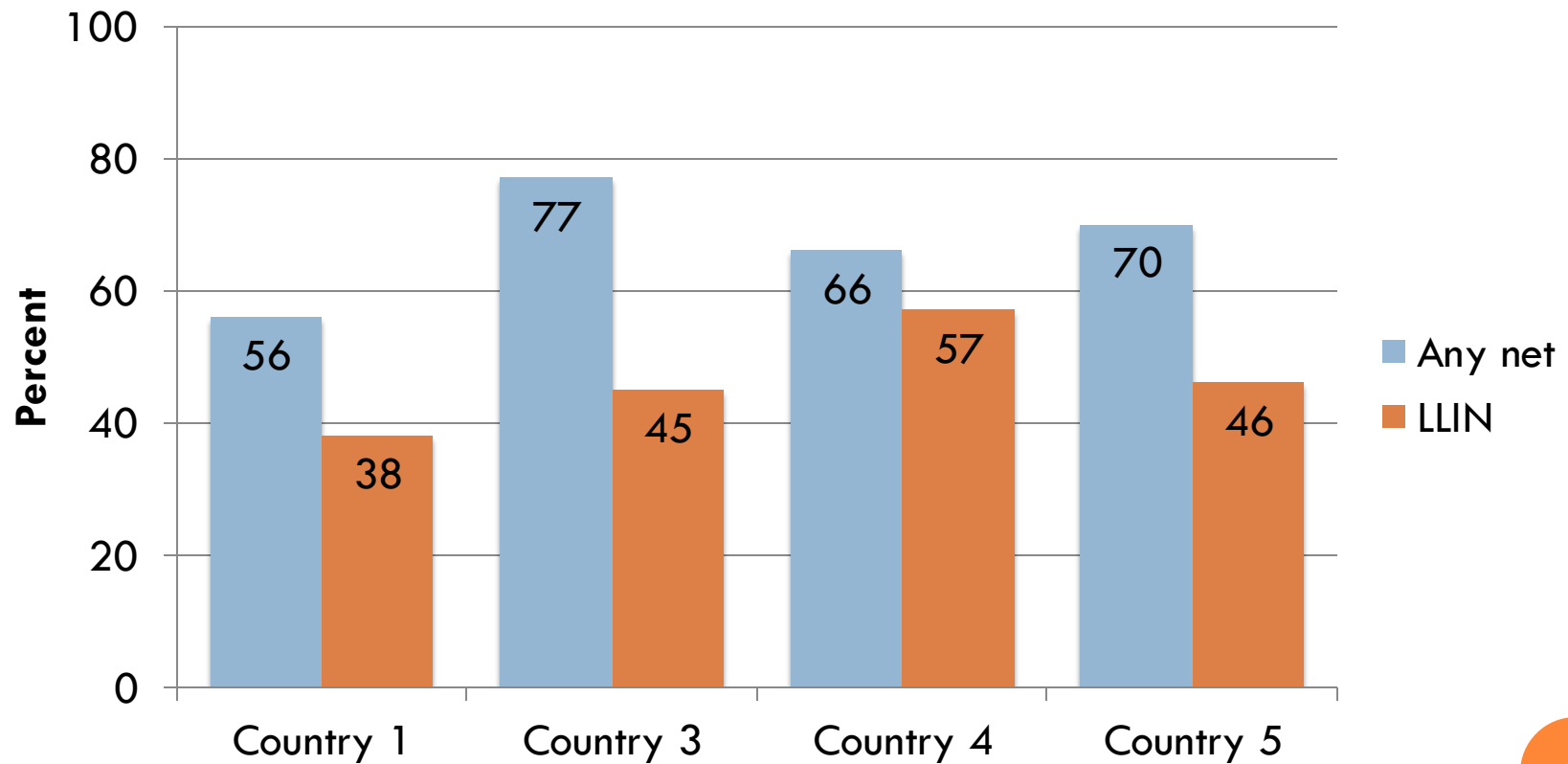
- نمودار میله ای: برای مقایسه داده های طبقه ایی
- نمودار دایره ای: نمایش چگونگی توزیع ویژگی کیفی در داده مورد نظر
- هیستوگرام: نمایش فراوانی در گروه های مختلف داده کمی
- نمودار خطی: نمایش روند در زمانهای مختلف داده کمی
- نمودار ساقه وبرگ: نشان دادن فراوانی داده کمی با کمک داده ها

نمودار میله ای



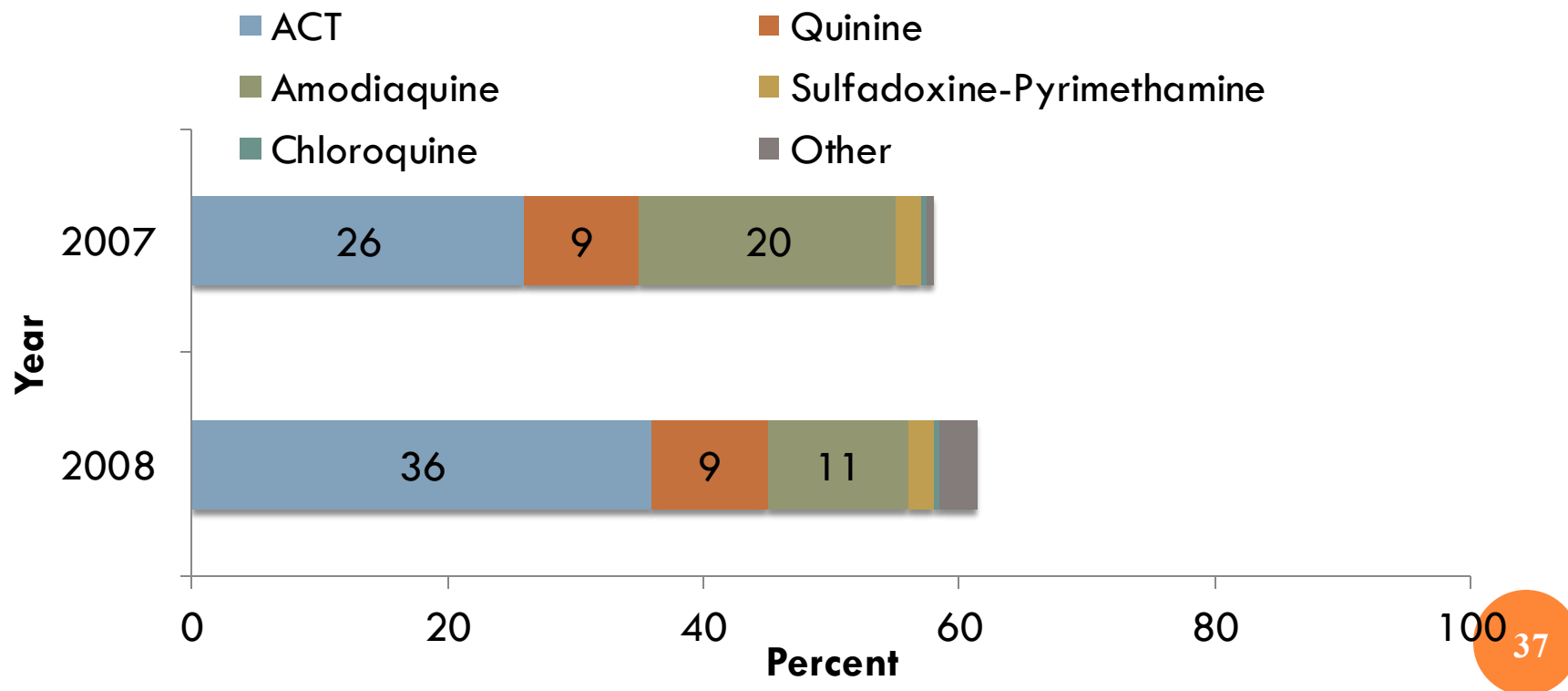
نمودار میله ای

Household Ownership of at Least 1 Net or ITN, 2008



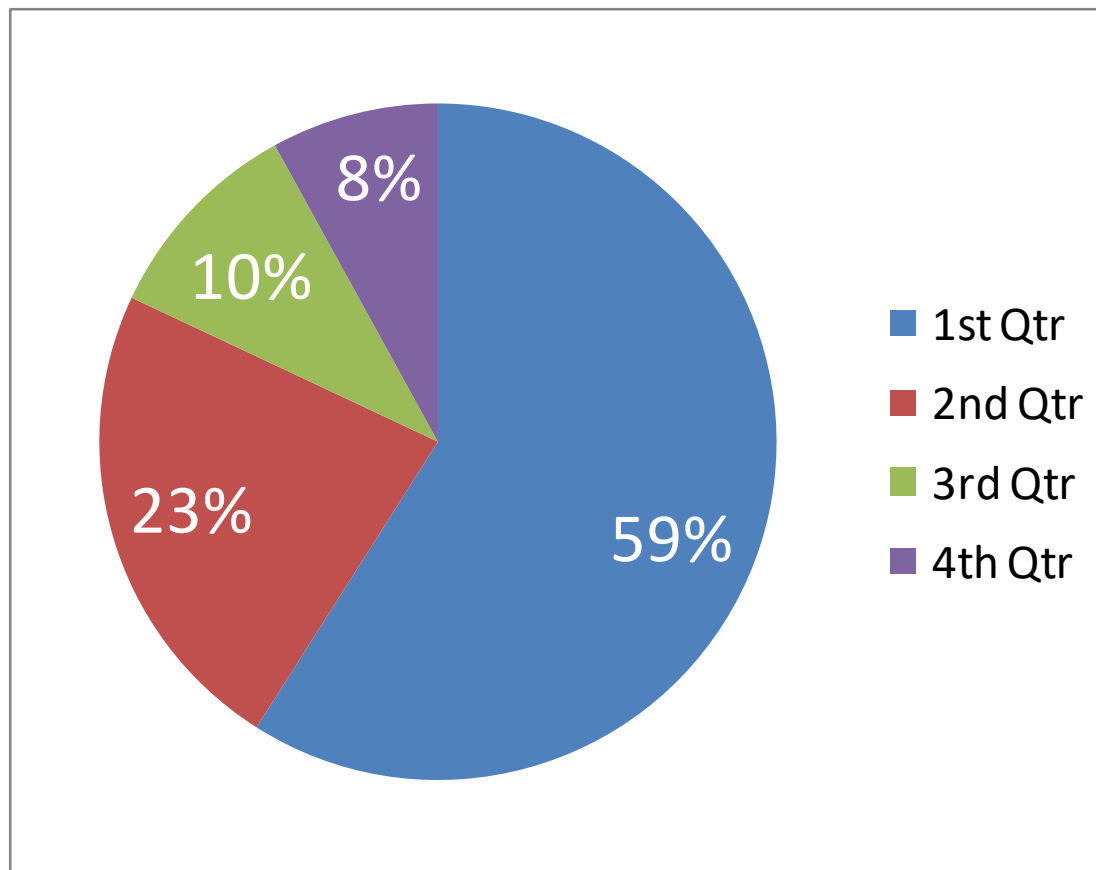
نمودار میله ای

Children <5 with Fever who Took Specific Antimalarial, 2007-2008 %



نمودار دایره ایی

Percentage of all confirmed malaria cases treated by quarter, Country X, 2011



محاسبه زاویه در نمودار دایره ایی

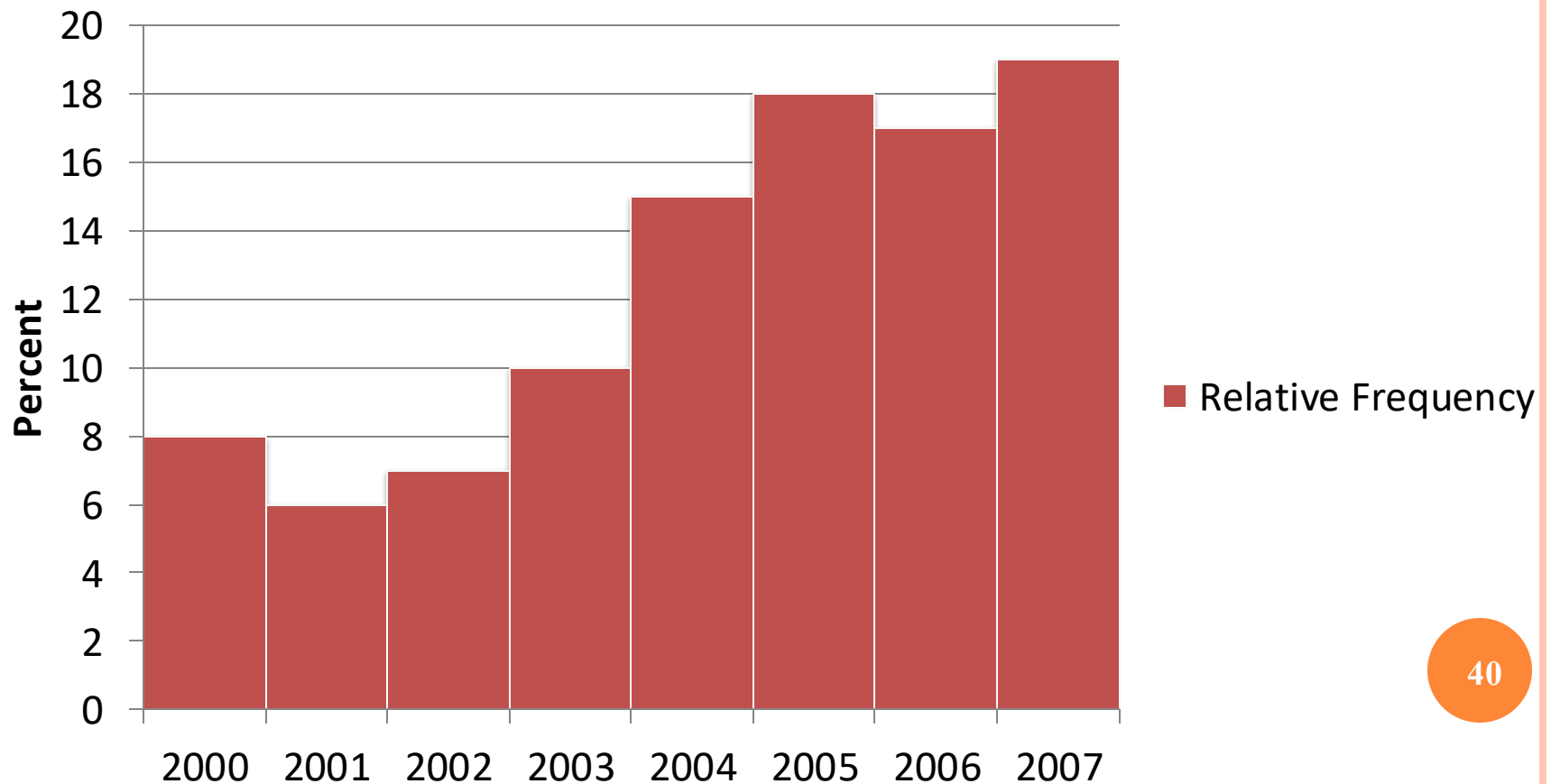
○ فرض کنید متغیر تصادفی مورد مطالعه ما دارای k حالت باشد $(f_1, f_2, \dots, f_k)$ در این صورت دایره ای به شعاع دلخواه را به وسیله زاویه های مرکزی به k قسمت تقسیم می کنیم به قسمی که اندازه زاویه مرکزی هریک از این قسمت ها متناسب با فراوانی آن قسمت باشد.

○ پس زاویه مرکزی نظیر دسته اول عبارت است از :

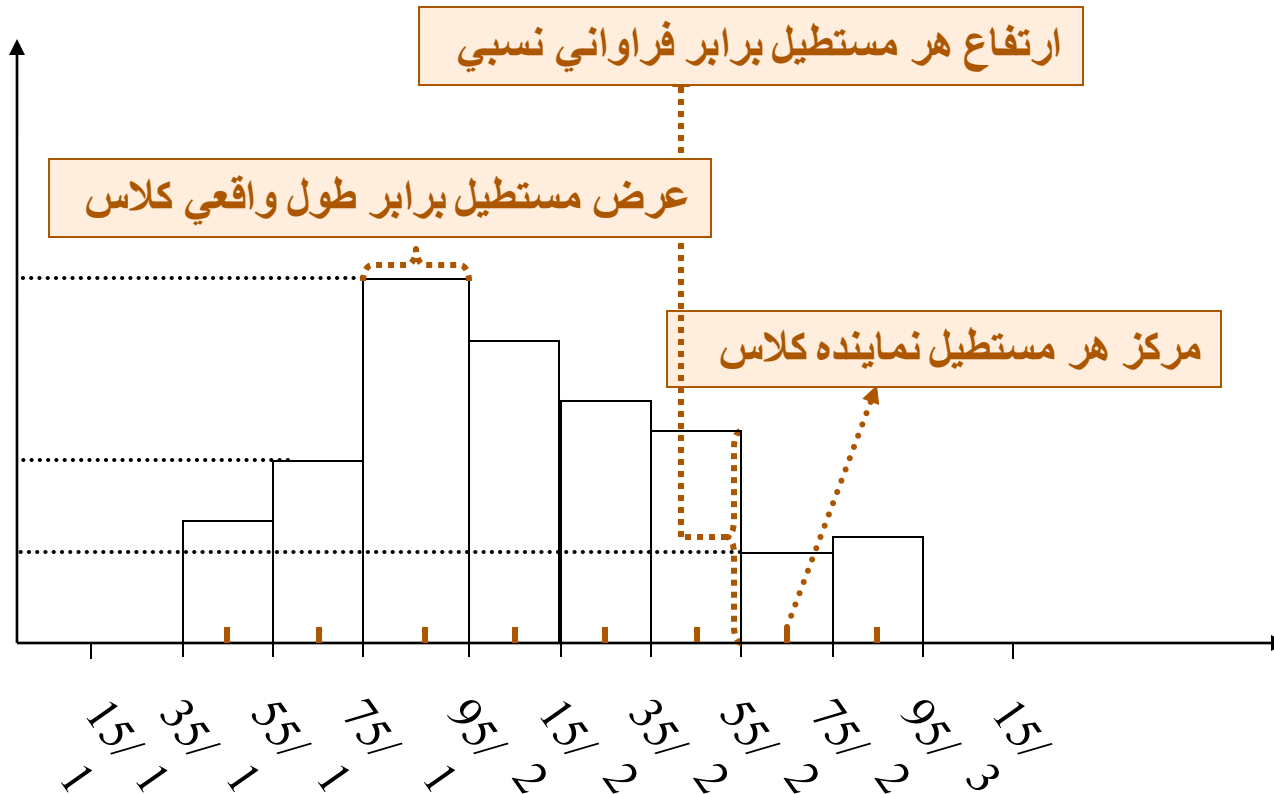
$$\frac{360}{f_1 + f_2 + \dots + f_k} \times f_1 = 360 \times \frac{f_1}{n}$$

هستوگرام

**Percent contribution of reported malaria cases by year
between 2000 and 2007, Kenya**



هیستوگرام...



نکته

- چنانچه در هیستوگرام فواصل گروه ها مساوی نباشد، از آنجا که باید سطح بالای گروه متناسب با فراوانی گروه باشد، لازم است با توجه به افزایش و یا کاهش قاعده مستطیل (فاصله گروه) ارتفاع آنرا چنان تغییر داد که نسبت مذکور مراعات شود.

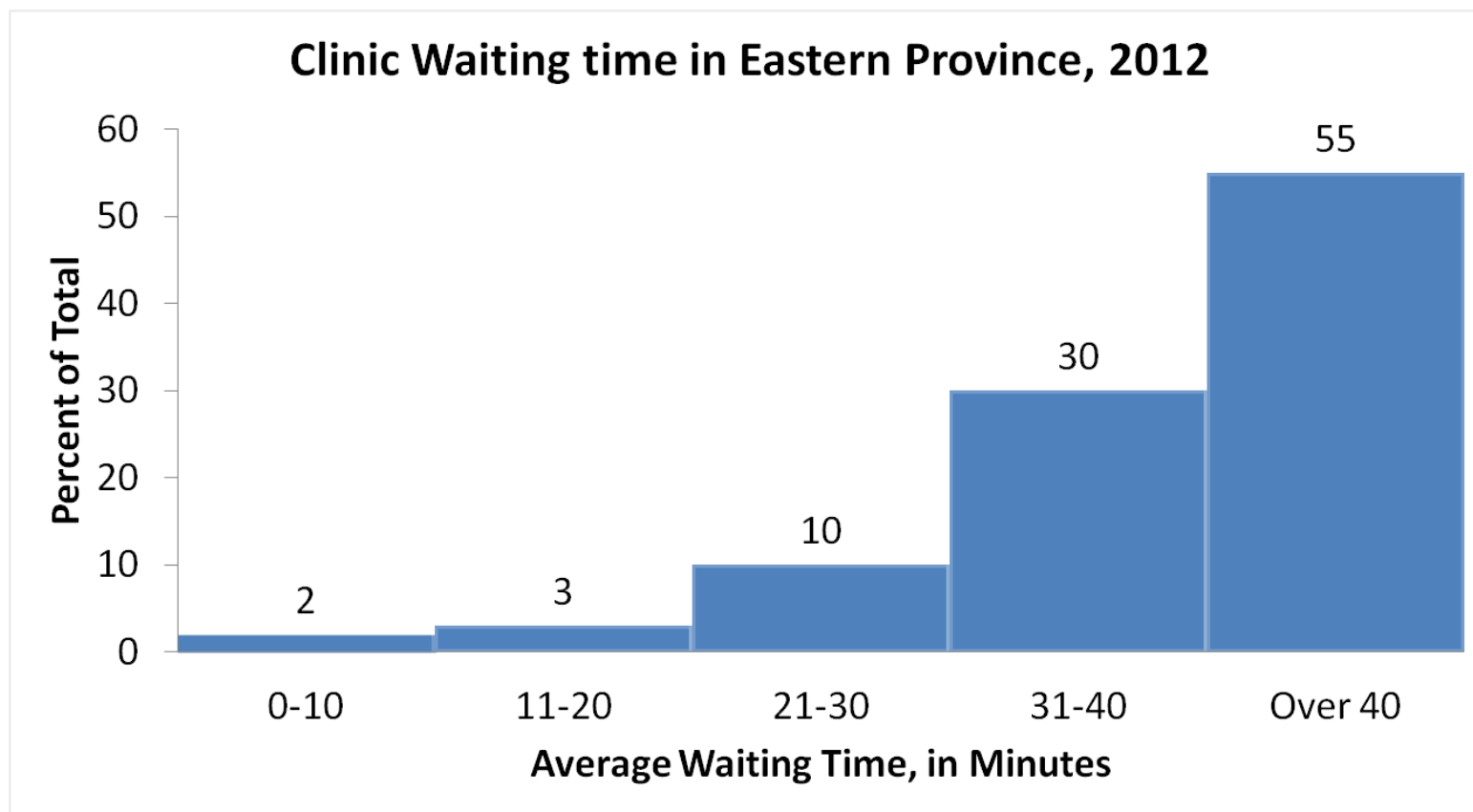
مثال

سن بر حسب سال	تعداد مرگ
0-1	18
1-2	43
2-3	50
3-4	60
4-5	36
5-6	24
6-7	22
7-8	21
8-9	6
9-10	5
10-15	14
15-20	3
جمع	302

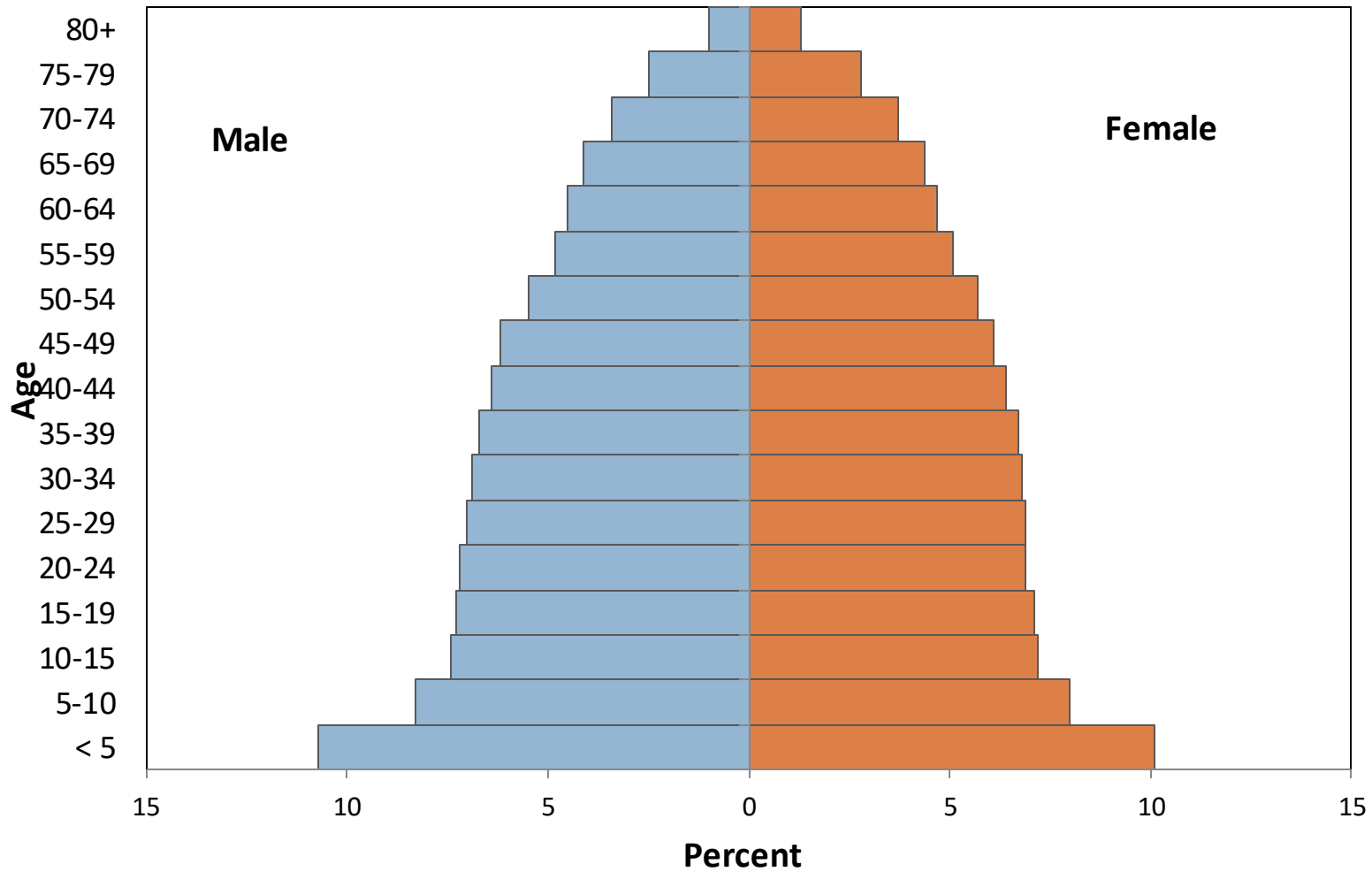
طول دسته دو گروه آخر ۵ برابر گروه های دیگر است. برای جبران آن لازم است ارتفاع مستطیل ها را برای این دو گروه به ترتیب $\frac{2}{8} = 5 \div 14$ و $\frac{0}{6} = 5 \div 3$ انتخاب نمود. به عبارت دیگر فراوانی را برای واحد فاصله حساب می کنیم.

هیستوگرام این داده ها را رسم کنید.

هیستوگرام داده چوله

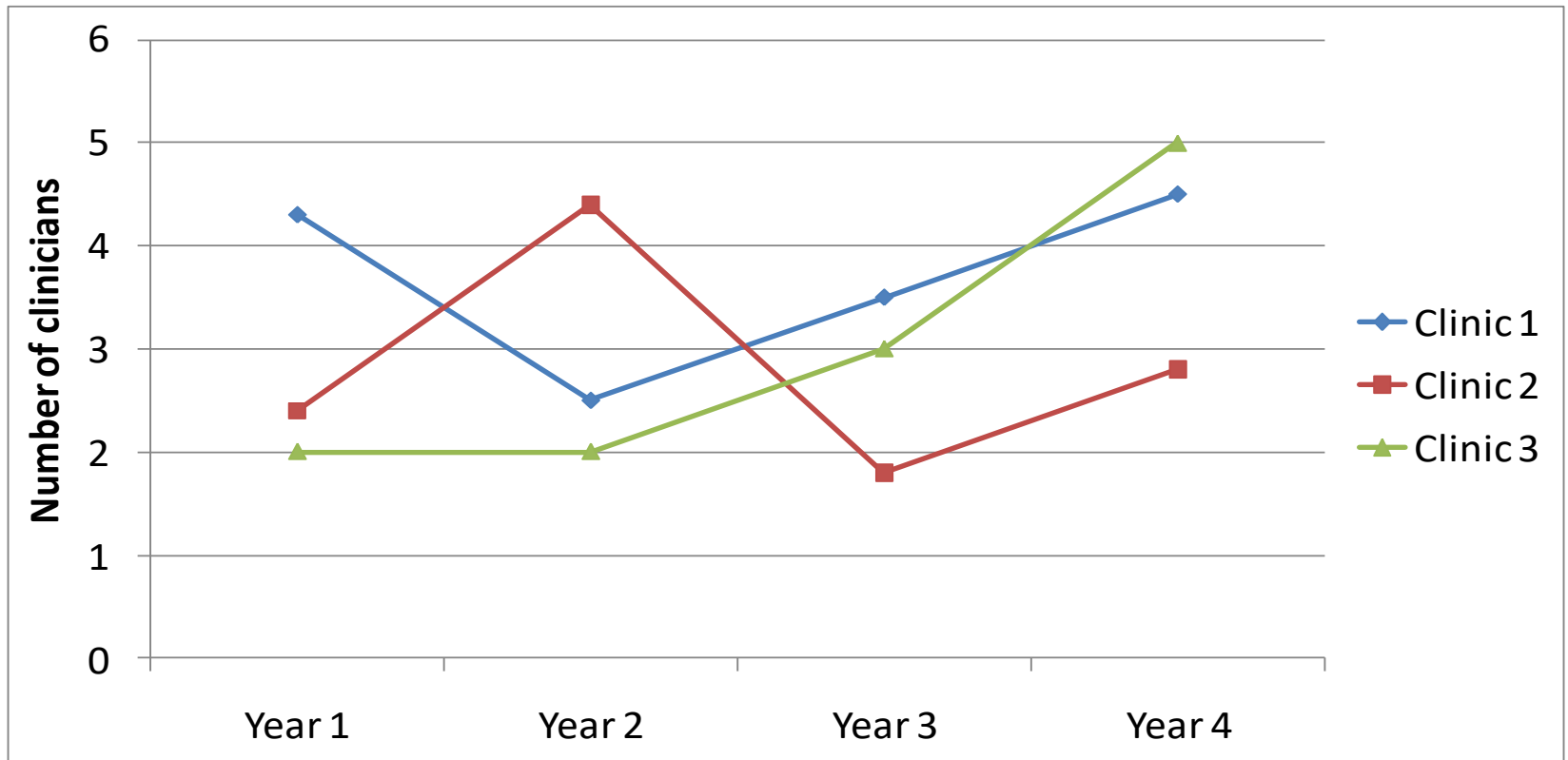


هرم جمعیتی



نمودار خطی

Number of Clinicians* Working in Each Clinic During Years 1-4, Country Y



*Includes doctors and nurses.

نمودار ساقه و برگ

ساقه

برگ

2	5																		
3	0	0	1	1	2	2	2	5	5	7	7	8	8	9					
4	0	0	0	0	1	2	4	4	4	5	5	6	6	6	7	7	7	9	9
5	0	0	0	0	0	1	2	2	2	4	6	6	6	6	7	7			
6	1	1	4	6	7														
7	1	8																	

تمرین

نمودارهای مناسب برای داده زیر را رسم کنید.

Method of delivery of 600 babies born in a hospital.

Method of delivery	No. of births	Percentage
Normal	478	79.7
Forceps	65	10.8
Caesarean section	57	9.5
Total	600	100.0

تمرین

نمودارهای مناسب برای داده زیر را رسم کنید.

Haemoglobin levels in g/100 ml for 70 women.

10.2	13.7	10.4	14.9	11.5	12.0	11.0
13.3	12.9	12.1	9.4	13.2	10.8	11.7
10.6	10.5	13.7	11.8	14.1	10.3	13.6
12.1	12.9	11.4	12.7	10.6	11.4	11.9
9.3	13.5	14.6	11.2	11.7	10.9	10.4
12.0	12.9	11.1	8.8	10.2	11.6	12.5
13.4	12.1	10.9	11.3	14.7	10.8	13.3
11.9	11.4	12.5	13.0	11.6	13.1	9.7
11.2	15.1	10.7	12.9	13.4	12.3	11.0
14.6	11.1	13.5	10.9	13.1	11.8	12.2