«جلسه دوم درس روش های آماری»

انواع مقیاس های اندازه گیری: جلسه قبل مقیاس اسمی و ترتیبی ارائه شد اکنون به مقیاس فاصله ای و نسبتی می پردازیم.

- مقیاس کمی فاصله ای: در این نوع اندازه گیری عناصر مورد اندازه گیری نه تنها از نظر ترتیب اندازه می توانند مقایسه شوند بلکه برحسب طول فاصله ها بین اندازه های عناصر نیز می توانند مقایسه و طبقه بندی گردند.

Example: مقیاسی که با آن شدت حرارت را نشان می دهیم. «واحد فاصله» در اینجا برای اندازه گیری «درجه» است.

در این مقیاس انجام بعضی از عملیات جبری روی اعداد بدست آمده از اندازه گیری امکان پذیر است.

Example: در اندازه گیری حرارت با مقیاس سانتیگراد نمی توان گفت که درجه حرارت ۵۰ دو برابر درجه حرارت ۲۵ می باشد.

- مقیاس کمی نسبتی: زمانیکه اعداد بدست آمده از اندازه گیری، از نظر ترتیب و فاصله بین دو مقدار، دارای اهمیت هستند، بکار می رود و همچنین وقتی که نسبت دو عدد بدست آمده از اندازه گیری ها، دارای اهمیت باشند. به عبارت دیگر، مقیاس نسبتی امکان می دهد که تعیین کنیم یک اندازه از اندازه دیگر چقدر بیشتر یا کمتر است و نیز می توان تعیین کرد این اندازه برای یک عنصر چند برابر اندازه عنصر دیگر است. وقتی که اندازه گیری به منظور مقایسه دو عنصر، بصورت نسبت اندازه های آن ها باشد، استفاده از اندازه گیری با مقیاس نسبتی اجتناب ناپذیر است. در این مقیاس نیز مانند مقیاس فاصله ای، واحد فاصله بین دو اندازه برای عناصر بطور دلخواه تعیین می شود. برخلاف مقیاس فاصله ای که انجام عمل تقسیم روی اعداد بدست آمده از اندازه گیری غیر ممکن می باشد، در مقیاس نسبتی انجام این عمل امکان پذیر است و می توان بین نسبت هایی از اعداد که از اندازه گیری ها بدست می آیند، تساوی برقرار کرد. (به عنوان مثال: $\frac{1}{2} = \frac{6}{12}$)

تفاوت اساسی مقیاس نسبتی و مقیاس فاصله ای: برای مقیاس نسبتی اندازه طبیعی بنام «صفر مطلق» وجود دارد، در صورتیکه در اندازه گیری بر طبق مقیاس فاصله ای، اندازه صفر به طور دلخواه تعیین می گردد.

نمادها:

- n ← حجم یا تعداد نمونه
- N ← حجم یا تعداد جامعه آماری
- x_i ← حدود طبقات
- x_i' ← مرکز (نماینده) طبقات
- T ← چوب خط
- F_i ← فراوانی مطلق
- f_i ← فراوانی نسبی
- $f_i\%$ ← درصد فراوانی نسبی
- c_i ← فراوانی تجمعی
- $c_i\%$ ← درصد فراوانی تجمعی

روش های طبقه بندی داده ها:

۱) داده های کیفی

۲) داده های کمی: گسسته / پیوسته

جدول توزیع فراوانی داده های کیفی: Example: رشته تحصیلی افراد

Example: ۱۰ نفر دانشجو بصورت تصادفی از لحاظ رشته تحصیلی بررسی شد که نتایج آن به شرح ذیل می باشد:

حسابداری - مدیریت - حسابداری - تربیت بدنی - مدیریت - تربیت بدنی - حسابداری - تربیت بدنی - حسابداری - مدیریت.

متغیر کیفی

جدول توزیع فراوانی آن به شرح ذیل می شود:

شماره ردیف i	رشته تحصیلی افراد x_i	چوب خط T	فراوانی مطلق F_i	فراوانی نسبی $f_i = \frac{F_i}{n}$	درصد فراوانی نسبی $\%f_i = f_i \times 100$	S_i	$\%S_i = \frac{S_i}{n} \times 100$
۱	مدیریت	III	۳	۰/۳	٪۳۰	$S_1 = F_1 = ۳$	٪۳۰
۲	حسابداری	IIII	۴	۰/۴	٪۴۰	$S_2 = F_1 + F_2 = ۷$	٪۷۰
۳	تربیت بدنی	III	۳	۰/۳	٪۳۰	$S_3 = F_1 + F_2 + F_3 = ۱۰$	٪۱۰۰
جمع			$n = \sum F_i = ۱۰$	$\sum f_i = 1$	٪۱۰۰		

نکات: ۱- مجموع فراوانی مطلق برابر n است ($\sum F_i = n$) ۲- مجموع فراوانی نسبی برابر ۱ است ($\sum f_i = 1$)

۳- فراوانی تجمعی طبقه اول برابر با فراوانی مطلق طبقه ی اول است و فراوانی تجمعی بقیه طبقات برابر است با مجموع فراوانی مطلق همان طبقه با طبقات ما قبل آن ۴- فراوانی تجمعی طبقه آخر برابر است با مجموع فراوانی مطلق کل طبقات (n)

۵- فراوانی نسبی برابر است با فراوانی مطلق تقسیم بر n ($f_i = \frac{F_i}{n}$) ۶- درصد فراوانی نسبی عبارت است از: $f_i \times 100$ (فراوانی نسبی)

نماد

۷- چوب خط برای تعیین تعداد فراوانی مطلق هر طبقه براساس داده های خام استفاده می شود (T)

۸- درصد فراوانی تجمعی برابر است با: $\%S_i = \frac{S_i}{n} \times 100 \Leftarrow \frac{\text{فراوانی تجمعی}}{n} \times 100$

جدول توزیع فراوانی داده های کمی گسسته (جدا): Example: تعداد فرزندان

متغیر کمی گسسته

شماره ردیف i	تعداد فرزندان x_i	F_i	$f_i = \frac{F_i}{n}$	$\%f_i$	S_i	$\%S_i = \frac{S_i}{n} \times 100$
۱	۰	۲۰	٪۲۹	٪۲۹	۲۰	٪۲۹
۲	۱	۴۰	٪۵۷	٪۵۷	۶۰	٪۸۶
۳	۲	۲	٪۰۳	٪۳	۶۲	٪۸۹
۴	۳	۸	٪۱۱	٪۱۱	۷۰	٪۱۰۰
جمع		$n = ۷۰$	۱	٪۱۰۰		

«جلسه سوم درس روش های آماری ۲»

جدول توزیع فراوانی داده های کمی پیوسته :

برای تشکیل جدول توزیع متغیر کمی پیوسته معمولاً سؤالاتی مطرح می شود که عبارتند از :

الف) تعیین تعداد طبقات (گروه های جدول) ← k (k از طریق قاعده استروجس بصورت زیر بدست می آید)

قاعده استروجس : $k = ۱ + ۳/۳۲۲ \log n$

ب) تعیین فاصله طبقات یا طول طبقات (طول هر دسته) ← h

بطوریکه $R = \underbrace{\max}_{\text{بزرگ ترین عدد}} - \underbrace{\min}_{\text{کوچک ترین عدد}}$

$h = \frac{R}{k}$

یا

$h = \frac{R}{k-1}$

در بعضی موارد آخرین ردیف جدول بزرگ ترین عدد را در بر نمی گیرد در چنین مواردی استفاده از فرمول دوم بهتر است.

Example : اگر کوچک ترین و بزرگ ترین عدد ها به ترتیب و ۲۰ باشد و $n=۳۰$ آنگاه طول طبقات چقدر است؟

$k = ۱ + ۳/۳۲۲ \log ۳۰ \Rightarrow \boxed{k = ۵/۹ \simeq ۶}$ $\qquad R = \max - \min = ۲۰ - ۸ \Rightarrow \boxed{R = ۱۲}$

طول طبقات : $h = \frac{R}{k} = \frac{۱۲}{۶} \Rightarrow \boxed{h = ۲}$

ج) تعیین حدود طبقات جدول ← x_i

* حدود طبقات جدول به دو دسته واقعی و ظاهری تقسیم می شوند

Example : اگر کوچک ترین عدد ۲۲ باشد حدود طبقات واقعی و ظاهری را تا ۲ طبقه بصورت زیر مشخص می کنیم :

حدود طبقات واقعی	حدود طبقات ظاهری
[۲۱/۵ - ۲۳/۵)	[۲۲ - ۲۴)
[۲۳/۵ - ۲۵/۵)	[۲۴ - ۲۶)

* اگر از کران پایین طبقات ظاهری ۰/۵ واحد کم کنیم و بعد به عدد بدست آمده به اندازه h اضافه کنیم حدود طبقات واقعی بدست می آید.

نکته : حدود طبقات ظاهری و واقعی وقتی مطرح می شود که داده های آماری را گرد کرده باشیم.

Example : ۲۱/۶ ، ۲۱/۷ ، ۲۱/۸ ، ۲۱/۹ $\simeq$ ۲۲ → اگر حدود طبقات واقعی را تشکیل دهیم ۲۲ همه اعداد روبرو را در بر می گیرد

نکته: ۱- حدود طبقات ممکن است بصورت فواصل نیم بسته - نیم باز باشد

Example: $\begin{cases} [۲۰-۲۵) \\ [۲۵-۳۰) \end{cases}$ ⇐ دراین روش نمودارها پیوسته هستند

همانطور که مشاهده می‌کنید ابتدای هر طبقه بسته و انتهای آن ها باز هستند به همین دلیل فواصل نیم بسته - نیم باز نامیده شدند. (چون عدد ۲۵ شامل طبقه دوم است پس در انتهای طبقه اول بسته نمی‌شود)

۲- فواصلی که در هر دو طرف بسته باشند

Example: $\begin{cases} [۲۰-۲۵] \\ [۲۶-۳۱] \end{cases}$ دراین روش نمودارها گسسته بوجود می‌آیند

(همانطور که مشاهده می‌کنید ابتدا و انتهای هر طبقه بسته است بنابراین در طبقه دوم دیگر از عدد ۲۵ شروع نمی‌شود و از [۲۶-۳۱] می‌شود)

ج) مرکز طبقه (نماینده طبقه):

$$X_i' = \frac{\text{کران بالا} + \text{کران پایین}}{2}$$

Example: برای طبقه [۲۰-۲۵) مرکز طبقه را بدست آورید!

$$X_i' = \frac{20+25}{2} = 22/5$$

Example: برای متغیر سن افراد جدول توزیع فراوانی را بدست آورید! با توجه به اینکه متغیر سن یک متغیر کمی پیوسته است داریم:

شماره ردیف (i)	حدود طبقات (X_i)	F_i	f_i	$f_i\%$	C_i	$C_i\%$	نماینده طبقات (X_i')
۱	[۲۲-۲۴)	۴۰	٪۴	٪۴۰	۴۰	٪۴۰	$X_1' = \frac{22+24}{2} = 23$
۲	[۲۴-۲۶)	۱۵	٪۱۵	٪۱۵	۵۵	٪۵۵	$X_2' = 25$
۳	[۲۶-۲۸)	۲۰	٪۲	٪۲۰	۷۵	٪۷۵	$X_3' = 27$
۴	[۲۸-۳۰)	۱۰	٪۱	٪۱۰	۸۵	٪۸۵	$X_4' = 29$
۵	[۳۰-۳۲)	۱۵	٪۱۵	٪۱۵	۱۰۰	٪۱۰۰	$X_5' = 31$
جمع		$n = 100$	۱	٪۱۰۰			

دراین حالت به جدول توزیع فراوانی حدود طبقات و نماینده طبقات نیز اضافه می‌شود.

نکته: سن افراد را می‌توان بصورت گسسته طبقه بندی کرد مثلاً گروه ۲۵ ساله، ۲۷ ساله و غیره یا می‌توان بصورت

پیوسته طبقه بندی کرد. مثلاً ۲۲-۲۰ و ۲۴-۲۲

* اگر متغیر گسسته باشد فقط بصورت گسسته می توان طبقه بندی کرد ولی اگر متغیر پیوسته باشد می توان بصورت پیوسته یا گسسته طبقه بندی کرد و بخاطر اینکه تعداد طبقات که مربوط به داده های زیاد می باشد کمتر باشد بهتر است که سون را بصورت پیوسته طبقه بندی کرد.

* توجه شود که در این حالت ما می توانیم تعداد طبقات را بدست آوریم ولی برای دو جدول قبل تعداد طبقات دیگر مطرح نبوده

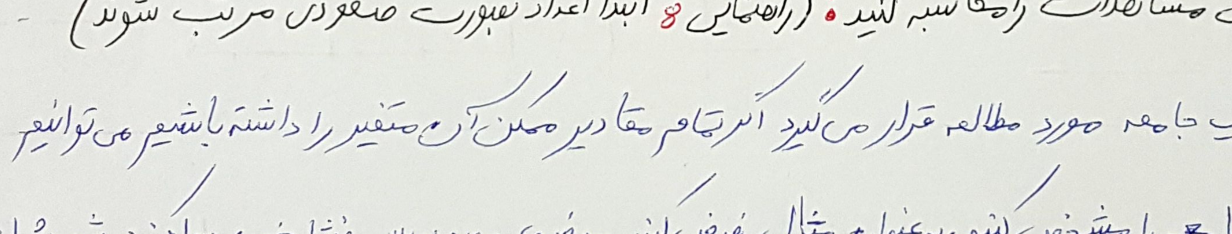
مهم تمرین: فرض کنید معدل ۲۲ دانش آموز به شرح زیر باشد:

۱۶/۴ - ۱۶/۷ - ۱۶/۵۶ - ۱۵/۳۴ - ۱۶/۸ - ۱۲/۷ - ۱۹/۵ - ۱۴/۵ - ۱۸/۲ - ۱۲ - ۱۷/۲

۱۶/۳۳ - ۱۵/۹۶ - ۱۳/۷ - ۱۴/۳ - ۱۸/۴ - ۱۷/۸ - ۱۶/۳ - ۱۶/۴ - ۱۵/۷ - ۱۸/۳ - ۱۳/۵

جدول فراوانی این مشاهدات را محاسبه کنید؟ (راهنمایی: ابتدا اعداد بصورت صعودی مرتب شوند) -

* هرگاه متغیری در یک جامعه مورد مطالعه قرار می گیرد اگر تمام مقادیر ممکن آن متغیر را داشته باشیم می توانیم نحوه توزیع دقیق جامعه را مشخص کنیم. به عنوان مثال فرض کنید موضوع مورد بررسی فشارخون ساکنین شهر شیراز باشد اگر تمامی اطلاعات مربوط به فشارخون اعضای این جامعه را داشته باشیم آنگاه می توانیم شکل توزیعی و دقیق فشارخون جامعه را در بین افراد سالم و دارای فشارخون بالا و پایین بصورت دقیق معین نماییم که این عمل از طریق اتصال انتهای میله های فراوانی نسبی مشاهدات ← در هر نقطه بین کوچک ترین و بزرگ ترین فشارخون ها به دست می آید، همانند نمودار زیر:



در عمل بدست آوردن این شکل دقیق توزیعی جامعه امکان پذیر نمی باشد «به دلیل نبودن تمامی اطلاعات جامعه»

« جلسه چهارم درس روش های آماری »

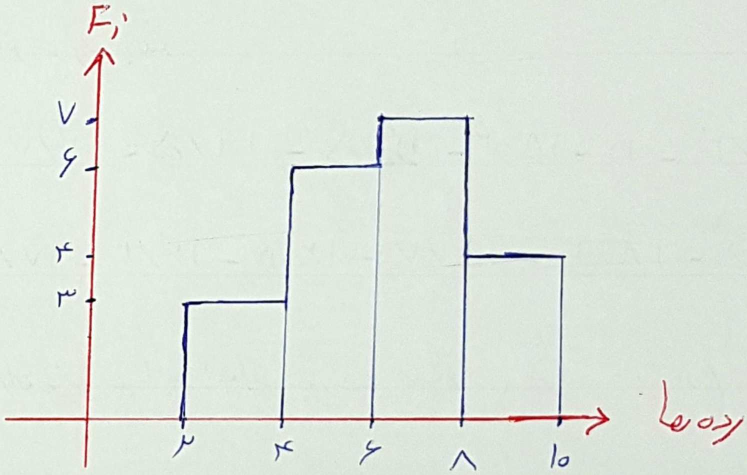
و هرگاه بخواهیم از نمونه ی استخراج یک جامعه درباره نحوه ی توزیع جامعه نتیجه گیری کنیم می بایست ابزارهای ترسیمی نیز داشته باشیم که نتیجه حاصل از این ابزارها تعیین توزیع تقریبی جامعه می باشد.

انواع نمودار :

۱) هیستوگرام « نمودار مستطیلی » : این نمودار تنها برای داده های کمی جدول بندی شده قابل استفاده می باشد و بر اساس ستون فراوانی نسبی یا فراوانی مطلق قابل رسم می باشد. به عنوان مثال برای جدول توزیع فراوانی زیر نمودار هیستوگرام بصورت زیر رسم می شود :

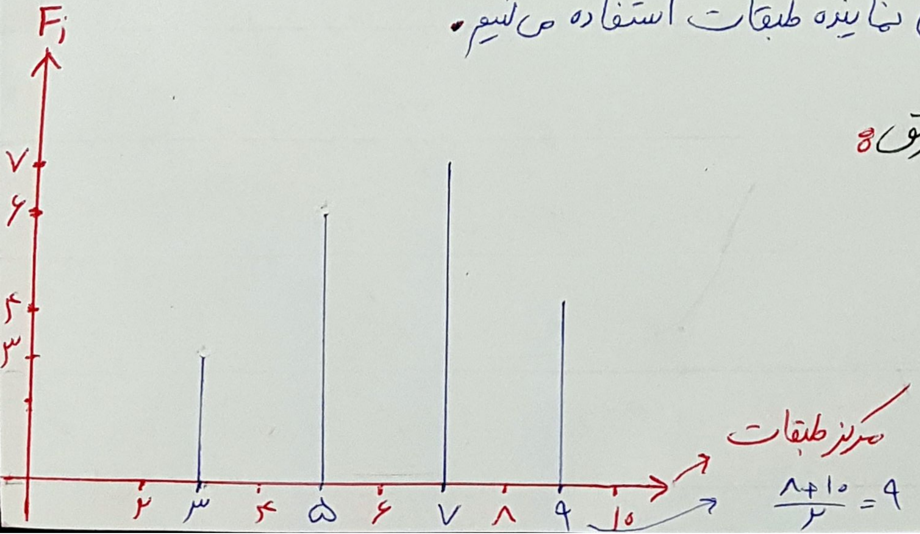
رده ها	F_i	f_i	ز ک
۲-۴	۳	$\frac{۳}{۲۰}$	۳
۴-۶	۶	$\frac{۶}{۲۰}$	۹
۶-۸	۷	$\frac{۷}{۲۰}$	۱۶
۸-۱۰	۴	$\frac{۴}{۲۰}$	۲۰
جمع	$n=۲۰$	۱	

نمودار هیستوگرام



۲) نمودار میله ای : این نمودار قابل استفاده برای داده های کیفی و کمی می باشد و بر اساس ستون فراوانی نسبی و مطلق قابل رسم است. هرگاه بخواهیم این نمودار را برای داده های کمی استفاده کنیم می بایست از نماینده ی طبقات بهره ببریم. خطی عمود بر محور مختصات و به ارتفاع فراوانی نسبی یا مطلق بر آن نقاط رسم نماییم و در متغیرهای کیفی از سطوح خود متغیر به عنوان (به جای) نماینده طبقات استفاده می کنیم.

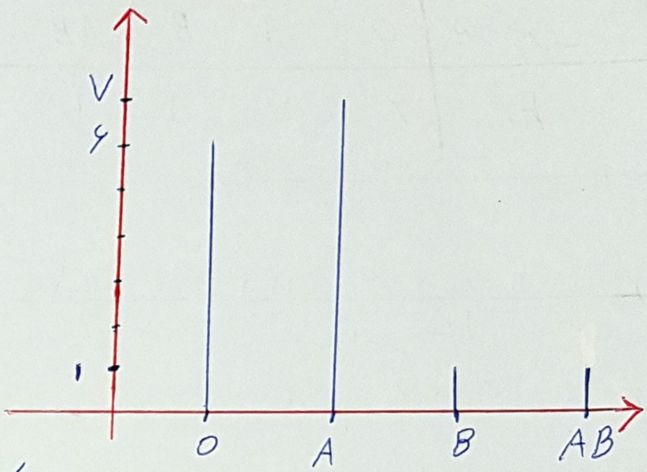
مثال برای داده های کمی جدول توزیع فراوانی فوق :



مثال برای داده های کیفی :

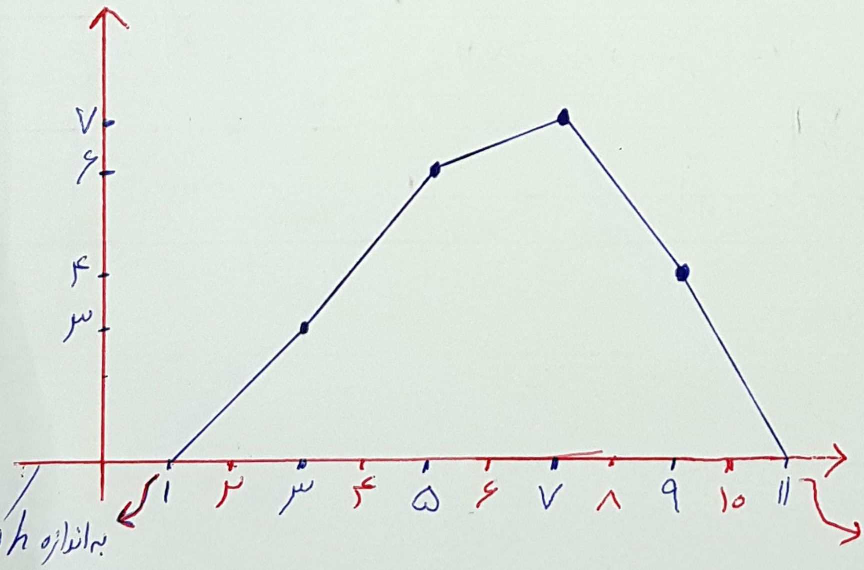
گروه خونی	F_i	f_i
O	۶	$\frac{6}{15}$
A	۷	$\frac{7}{15}$
B	۱	$\frac{1}{15}$
AB	۲	$\frac{2}{15}$

نمودار میله ای



۳) چندبر فراوانی : این نمودار برای داده های کمی و براساس ستون فراوانی مطلق قابل رسم می باشد چندبر فراوانی براساس اتصال نقاط انتهایی نمودار میله ای واتصال آن به محور افقی رسم می گردد.

EX : برای جدول توزیع فراوانی ارائه شده در قسمت نمودار هیستوگرام نمودار چندبر فراوانی بصورت زیر می شود



به اندازه h کم می کنیم یعنی ۳-۲=۱

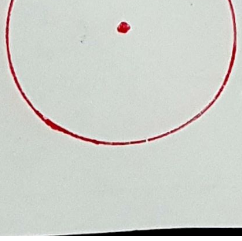
$h=2 \Rightarrow$

به اندازه h اضافه می کنیم یعنی ۹+۲=۱۱

چون حدود طبقات ۲تا۲تا جلو رفته

۴) نمودار دایره ای : این نمودار برای داده های کمی و کیفی قابل استفاده است و بصورت زیر محاسبه ورسم می گردد :

k سطح یا طبقه



قطاع مربوط به طبقه iام یا سطح iام متغیر کیفی : $\gamma_i = 360^\circ \times f_i = 360^\circ \times \frac{F_i}{n}$

Example : نمودار دایره‌ای دو مثال قبل را بدست آورید؟

مثال ۱ :

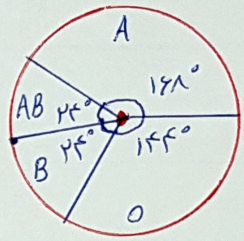
گروه خونی	O	A	B	AB
F_i	۶	۷	۱	۱

$$\sum_{i=1}^{n} F_i = n = ۶ + ۷ + ۱ + ۱ = ۱۵$$

$$r_O = \frac{۳۶۰ \times ۶}{۱۵} = ۱۴۴$$

$$r_A = \frac{۳۶۰ \times ۷}{۱۵} = ۱۶۸$$

$$r_B = r_{AB} = \frac{۳۶۰ \times ۱}{۱۵} = ۲۴$$



مثال ۲ :

طبقه‌ها	F_i	$f_i = \frac{F_i}{n}$	r_i
طبقه اول ۲–۴	۳	$\frac{۳}{۲۰}$	$r_1 = \frac{۳}{۲۰} \times ۳۶۰ = ۵۴$
طبقه دوم ۴–۶	۶	$\frac{۶}{۲۰}$	$r_2 = \frac{۶}{۲۰} \times ۳۶۰ = ۱۰۸$
طبقه سوم ۶–۸	۷	$\frac{۷}{۲۰}$	$r_3 = \frac{۷}{۲۰} \times ۳۶۰ = ۱۲۶$
طبقه چهارم ۸–۱۰	۴	$\frac{۴}{۲۰}$	$r_4 = \frac{۴}{۲۰} \times ۳۶۰ = ۷۲$
جمع	$n = ۲۰$	$\sum f_i = \frac{۲۰}{۲۰} = ۱$