آمار و احتمالات مهندسی

جلسه دوم

مسئله:

جدول فراوانی زیر مربوط به وزن ۵۰ دانش آموز است. جدول را کامل کنید.

G	F	r	f	X	رده
		0.06			**30.5 - 35.5**
			7		**35.5 – 40.5**
	19				
0.64					
		0.16			
	46				
			50		جمع

حل مسئله:

جدول فراوانی زیر مربوط به وزن ۵۰ دانش آموز است.

G	F	r	f	X	رده
0.06	3	0.06	3	33	30.5 - 35.5
0.2	10	0.14	7	38	35.5 – 40.5
0.38	19	0.18	9	43	40.5 – 45.5
0.64	32	0.26	13	48	45.5 – 50.5
0.8	40	0.16	8	53	50.5 – 55.5
0.92	46	0.12	6	58	55.5 – 60.5
1	50	0.08	4	63	60.5 – 65.5
		1	50		جمع

نمودارها

نمودارها کمک میکنند تا اطلاعات بدست آمده از جدول فراوانی با ساده ترین بیان و در کمترین زمان و با تصوری واضح از توزیع، ارائه شوند.

مهمترین نمودارهای داده های کیفی

نمودار ستونی: مقایسه فراوانی در بین رده ها

نمودار دایره ای: مقایسه فراوانی نسبی رده ها

نمودار پارتو (Pareto): مقایسه فراوانی رده ها از فراوانی بیشتر تا فراوانی کمتر

مهمترین نمودارهای داده های کمی

نمودار فراوانی (Histogram): مقایسه فراوانی ها یا فراوانی نسبی در بین رده ها

نمودار چندبر فراوانی (Polygon): از اتصال نقاط فراوانی مربوط به هر نماینده رده در نمودار فراوانی به کمک خط مستقیم ایجاد می شود.

نمودار چند بر فراوانی تجمعی: از اتصال نقاط فراوانی تجمعی مربوط به هر نماینده رده در نمودار فراوانی به کمک خط مستقیم ایجاد می شود.

منحنی فراوانی: از اتصال نقاط فراوانی به کمک خطوط منحنی

شاخص های آماری

ارقام و اعداد بدست آمده از داده های جمع آوری شده که بیانگر ویژگی هایی از نمونه آماری هستند.

مهمترین شاخص ها

شاخص های مرکزی (Central Tendency Indexes)

شاخص های پراکندگی (Dispersion Indexes)

شاخص های توزیع یا تقارن (Distribution Indexes)

شاخص های آماری

شاخص های مرکزی

- میانگین (Mean)
  - میانگین حسابی (Arithmetic mean)
  - میانگین اصلاح شده (Trimmed mean)
  - میانگین وزنی (Weighted mean)
  - میانگین هندسی (Geometric mean)
  - میانگین همساز یا توافقی (Harmonic mean)
- میانه (Median)
- نما (Mode)
- چندک ها (Quintile)
  - چارک (Quartiles)
  - دهک ها (Deci les)
  - صدک ها (Percentiles)

شاخص های آماری

شاخص های مرکزی

- **میانگین (Mean)**

➢**میانگین حسابی (Arithmetic mean)**

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

همه مقادیر دارای اهمیت یکسانی در محاسبه میانگین هستند.

میانگین حسابی نسبت به مقادیر خیلی زیاد یا خیلی کم داده ها حساسیت داشته و وجود چنین مواردی موجب می شود تا میانگین حسابی نماینده خوبی از جامعه مورد مطالعه نباشد. از این رو در چنین مواردی می بایست داده هایی که نسبت به عموم جامعه مقادیر پرتی دارند کنار بگذاریم زیرا احتمال بروز اشتباه در جمع آوری داده ها وجود دارد.

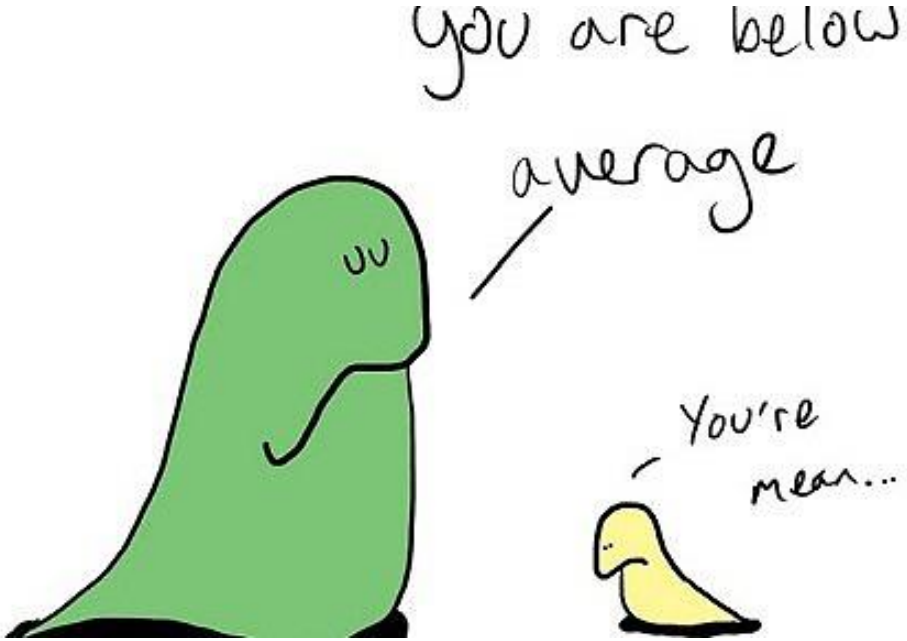
نکته: اگر میانگین یک سری اعداد برابر با $\bar{X}$ باشد، میانگین مقادیر Y=ax+b برابر خواهد بود با : $\bar{Y} = a\bar{X} + b$

شاخص های آماری

شاخص های مرکزی

- **میانگین (Mean)**
  - **میانگین اصلاح شده (Trimmed mean)**

داده های پرت را از مجموعه داده ها حذف کرده و از بقیه داده ها میانگین حسابی گرفته می شود.



شاخص های آماری

شاخص های مرکزی

- **میانگین (Mean)**

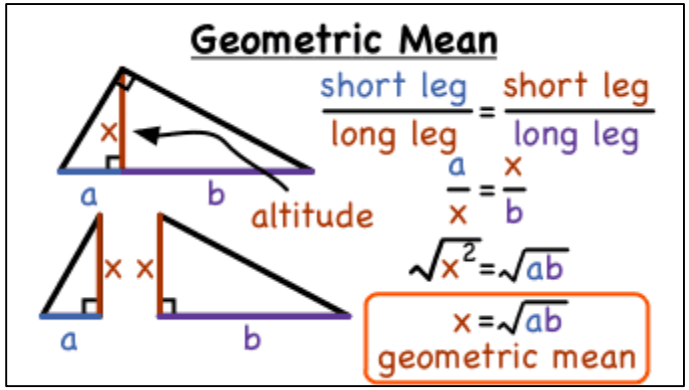
$$\overline{X}_w = \sum_{i=1}^{n} w_i X_i$$

➢**میانگین وزنی(Weighted mean)**

برخی از داده ها نسبت به سایرین از اهمیت بیشتر یا نقش موثرتری برخوردار هستند که این اهمیت به صورت وزن در محاسبه میانگین در نظر گرفته می شود.

قابل ذکر است که وزن هر داده، عددی بین صفر تا یک بوده به طوری مجموع وزن ها می بایست یک شود.

یک مثال خوب برای این مورد همان میانگین کل است که باتوجه به وزن هر درس (که همان تعداد واحد درس تقسیم بر کل واحدها است) محاسبه می شود.



شاخص های آماری

شاخص های مرکزی

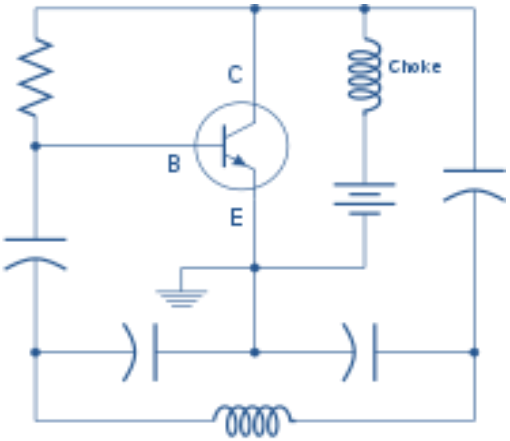
- **میانگین (Mean)**

➢ **میانگین هندسی (Geometric mean)**

$$G = \sqrt[n]{X_1 . X_2 \ldots X_n}$$

این میانگین برای داده های درصدی با مقادیر مثبت بکار می رود. در این نوع میانگین نیز همه داده ها اهمت یکسانی در محاسبه میانگین دارند. در واقع این میانگین برابر است با ریشه nام حاصل ضرب تمامی n داده در هم.

مثال: داده ها $\{0.1, 0.2, 0.3\}$ ⇦ $G = \sqrt[n]{X_1 . X_2 \ldots X_n} = \sqrt[3]{0.1 \times 0.2 \times 0.3} = 0.1817$



شاخص های آماری

شاخص های مرکزی

- **میانگین (Mean)**

➢**میانگین همساز (Harmonic mean)**

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} 1/x_i}, \qquad x_i > 0 \text{ for all } i.$$

این میانگین نوعی سنجش گرایش به مرکز است، و معمولاً هنگامی کاربرد دارد که محاسبه میانگین نرخ‌ها اهمیت داشته باشد. این مقدار عبارت است از تعداد مقادیر تقسیم بر مجموع معکوس مقادیر موجود در یک مجموعه داده‌ها. یکی از کاربردهای آن در محاسبات مربوط به خازن ها و مقاومت ها در الکترونیک است.

در موارد خاص، بخصوص مواردی که شامل نرخ‌ها و نسبت‌ها است، میانگین هارمونیک صحیح‌ترین مقدار میانگین را به ما می‌دهد. برای مثال: اگر یک وسیله فاصله مشخصی را با سرعت **x مثلا ۶۰ کیلومتر بر ساعت** طی کند و سپس همان فاصله را دوباره با سرعت **y مثلا ۴۰ کیلومتر بر ساعت** طی کند، مقدار سرعت متوسط، میانگین هارمونیک x , y است، یعنی: **۴۸ کیلومتر بر ساعت**.

شاخص های آماری

شاخص های مرکزی

- **مقایسهٔ میانگین ها**

اگر تمامی داده ها مثبت باشند:

میانگین توافقی < میانگین هندسی < میانگین حسابی

مثال:

برای داده های از **۱** الی **۲۰**، میانگین حسابی برابر با **۱۰/۵** ، میانگین هندسی برابر با **۸/۳** و میانگین توافقی برابر با **۵/۵۶** است.

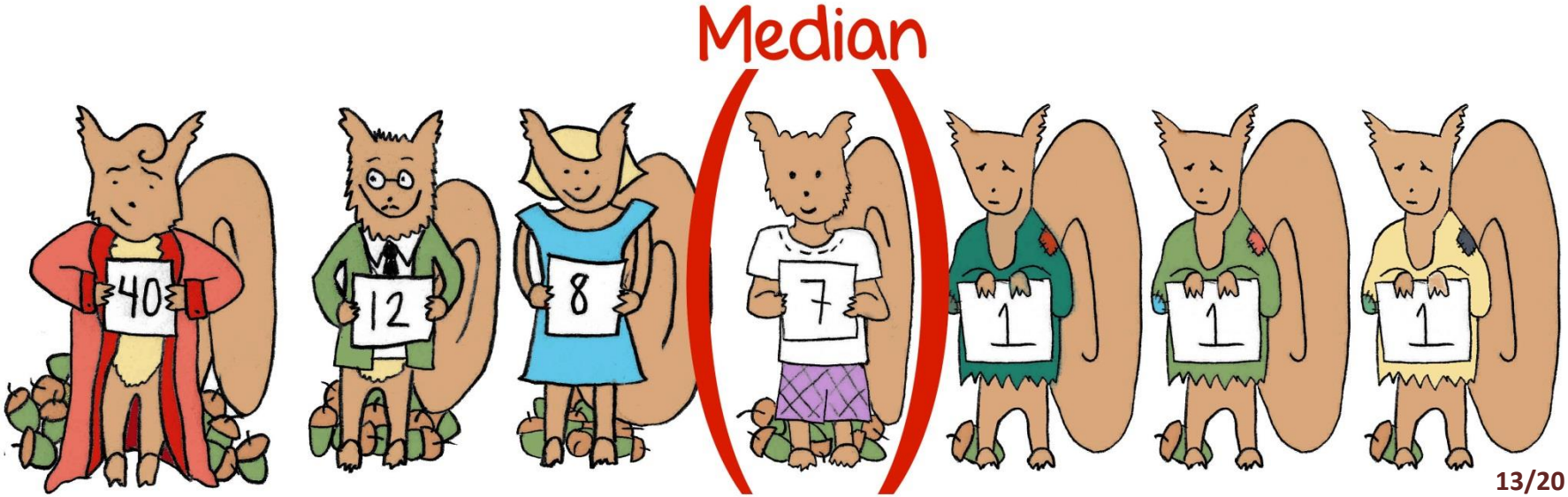
شاخص های آماری

شاخص های مرکزی

- **میانه (Median)**

برای بدست آوردن میانه داده ها بایستی از کم به زیاد مرتب سازی شوند. مقداری که در وسط لیست قرار گرفته است میانه خواهد بود.

$$m = \begin{cases} X_{(n+1)/2} & n = 2k+1 \\ (X_{(n/2)} + X_{(n/2)+1})/2 & n = 2k \end{cases}$$



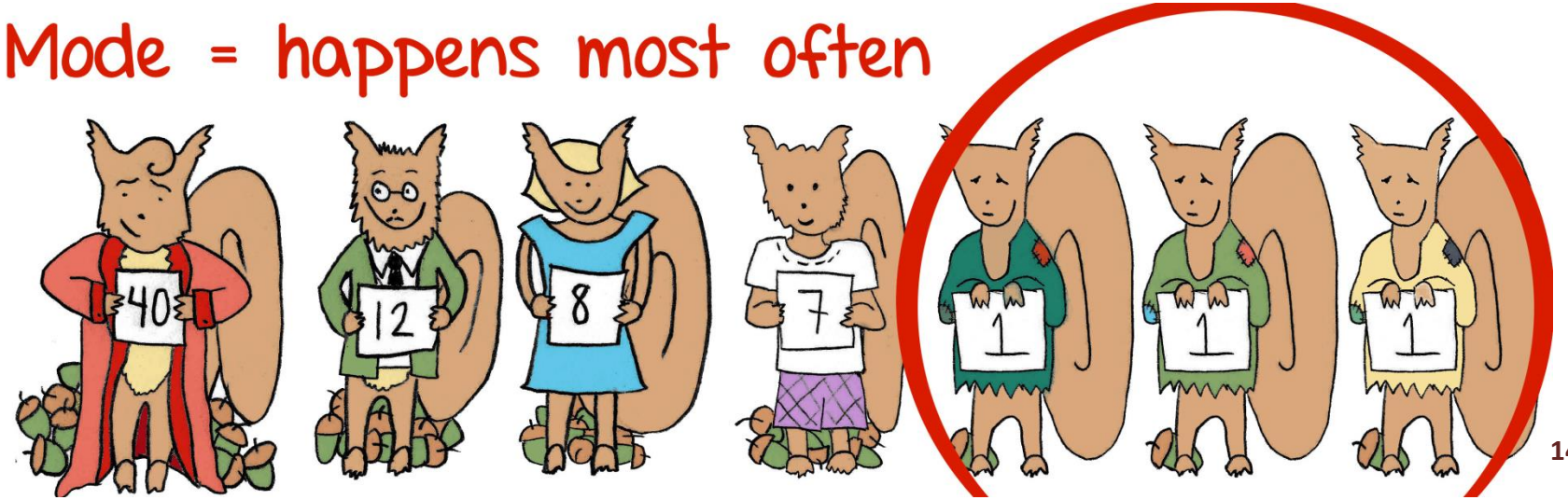
شاخص های آماری

شاخص های مرکزی

- **نما (Mode)**

داده ای با بیشترین تعداد تکرار در یک مجموعه را **نما** یا **مد** می گویند.

ممکن است داده ها دارای دو نما باشند. اگر دو رده که در مجاورت هم نیستند دارای فراوان یکسانی باشند، داده ها را دو نمایی گویند.



شاخص های آماری

شاخص های مرکزی

- **چندک ها (Quintile)**

برای استفاده از این شاخص، داده ها از کم به زیاد مرتب شوند.

مقداری که p درصد داده ها گوچکتر از آن باشند را به عنوان چندک q در نظر میگیریم. به طوری که $(n+1)*p=q$، بنابراین $X(q)$ چندک مورد نظر ما خواهد بود.

چندک های معروف:

- چارک $p=25\%, 50\%, 75\%$
- دهک $p=10\%, 20\%, \ldots, 90\%$
- صدک $p=1\%, 2\%, \ldots, 98\%, 99\%$

میانه همان چارک دوم، دهک پنجم یا صدک پنجاهم است.

مثال: در مجموعه {۱۲و۸و۷و۷و۵و۴و۴و۳و۲و۱و۰} صدک هفتادو پنجم (چارک سوم) کدام است؟

$(n+1)*p=q \rightarrow (11+1)*0.75= 9 \rightarrow X(9)=7$

شاخص های آماری بر اساس جدول فراوانی

میانگین

در جدول فراوانی f_i تعداد فراوانی هر رده، k تعداد رده و n تعداد کل داده است.

1) میانگین حسابی

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{n}$$

2) میانگین هندسی

$$G = \sqrt[n]{X_1^{f_1} . X_2^{f_2} \ldots X_k^{f_k}}$$

3) میانگین توافقی

$$H = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

شاخص های آماری بر اساس جدول فراوانی

میانه

$$m = L_{0.5} + \left(\frac{0.5 \times n - G_{0.5}}{f_{0.5}} \right) \times w$$

$L_{0.5}$ کران پایین رده میانه است.

$G_{0.5}$ فراوانی تجمعی رده قبل از میانه

$f_{0.5}$ فراوانی رده میانه

w طول رده

مثال:

رده	حدود رده	نماینده رده X	فراوانی f	فراوانی نسبی r	فراوانی تجمع F	فراوانی نسبی تجمعی G
۱	[۱۱.۵۱-۱۳.۵۱)	۱۲.۵۱	۲	۰.۱	۲	۰.۱
۲	[۱۳.۵۱-۱۵.۵۱)	۱۴.۵۱	۸	۰.۴	۱۰	۰.۵
۳	[۱۵.۵۱-۱۷.۵۱)	۱۶.۵۱	۴	۰.۲	۱۴	۰.۷
۴	[۱۷.۵۱-۱۹.۵۱]	۱۸.۵۱	۶	۰.۳	۲۰	۱
			۲۰	۱		

$m=13.51+2*(0.5*20-2)/(8)=15.51$

شاخص های آماری بر اساس جدول فراوانی

چندک

در رابطه ارائه شده برای میانه، مقدار 0.5 بیانگر درصدی از داده ها بود که از میانه کوچکتر هستند. بنابراین رابطه فوق برای محاسبه چندک ها به صورت زیر ارائه می شود. که در آن کافی است به جای مقدار 0.5 از p استفاده شود.

$$Q_p = L_p + \left(\frac{p \times n - G_p}{f_p} \right) \times w$$

اگر بخواهیم در مثال قبلی چارک سوم را محاسبه کنیم داریم:

n*0.75 = 20*0.75 = 15 ➔ فراوانی تجمعی ردیف چهارم جدول بزرگتر مساوی۱۵ است

کران پایین این رده ۱۷/۵۱ است. پس داریم:

17.51+(15-14)*2/6=17.84

شاخص های آماری بر اساس جدول فراوانی

نما

$$M = L_M + \left(\frac{d_1}{d_1 + d_2}\right) \times w$$

L_M کران پایین رده نما است.

d_1 اختلاف فراوانی نسبی رده نما با رده قبلی از آن

d_2 اختلاف فراوانی نسبی رده نما با رده بعد از آن

w طول رده

مثال:

رده	حدود رده	نماینده رده X	فراوانی f	فراوانی نسبی r	فراوانی تجمع F	فراوانی نسبی تجمعی G
۱	[۱۱.۵۱-۱۳.۵۱)	۱۲.۵۱	۲	۰.۱	۲	۰.۱
۲	[۱۳.۵۱-۱۵.۵۱)	۱۴.۵۱	۸	۰.۴	۱۰	۰.۵
۳	[۱۵.۵۱-۱۷.۵۱)	۱۶.۵۱	۴	۰.۲	۱۴	۰.۷
۴	[۱۷.۵۱-۱۹.۵۱]	۱۸.۵۱	۶	۰.۳	۲۰	۱
			۲۰	۱		

$$d_1 = 0.4 - 0.1 = 0.3$$
$$d_2 = 0.4 - 0.2 = 0.2$$

$$M = 13.51 + \left(\frac{0.3}{0.5}\right) 2 = 14.71$$

امکان استفاده از شاخص ها با توجه به نوع داده ها

شاخص / نوع داده	نما	میانه	میانگین
کیفی (بدون ترتیب	دارد	ندارد	ندارد
کیفی (ترتیبی)	دارد	دارد	ندارد
کمی	دارد	دارد	دارد