

فصل چهارم، آمار استنباطی

درس دوم (برآورد)

یاد آوری مطالب مورد نیاز

انحراف معیار: میانگین فاصله داده ها از میانگین

نمونه گیری: فرایند انتخاب نمونه ای از یک جامعه به منظور تعمیم آن به جامعه

پارامتر: یک مشخصه عددی توصیف کننده جنبه ای خاص از جامعه

آماره: مشخصه ای خاص توصیف کننده جنبه ای خاص از نمونه

برآورد نقطه ای پارامتر جامعه: مقدار عددی آماره

مثال:

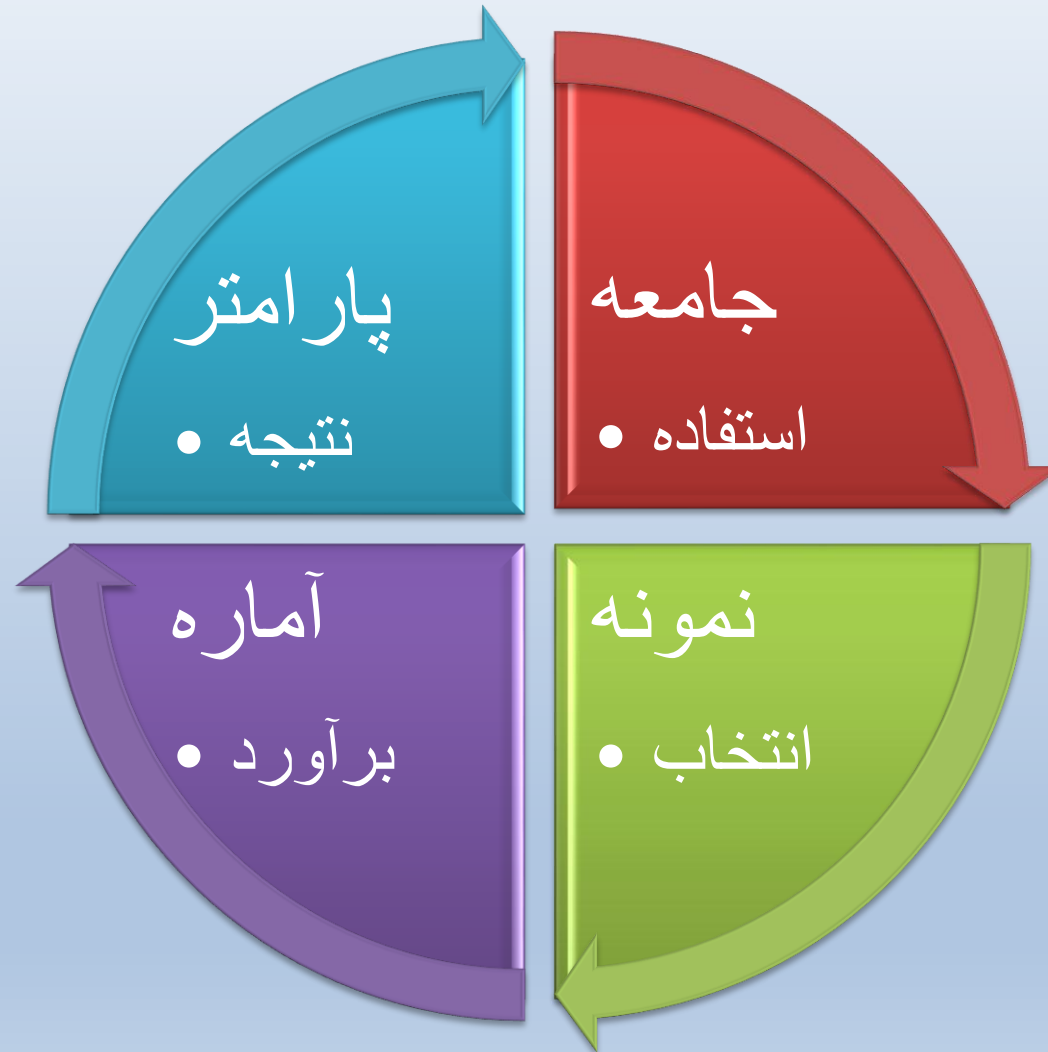
یک شرکت تولید لیوان شیشه ای می خواهد تعداد لیوان هایی را که در یک بسته قرار میدهد مشخص کند. تعداد لیوان های هر بسته به میانگین تعداد اعضای خانوارهای کشور بستگی دارد که بعد خانوار نامیده می شود. بعد خانوار در ایران پس از یک سرشماری که هر ده سال یک بار انجام میشود به دست می آید. پس این شرکت اطلاعات جدید و مورد نیاز را تا سه سال آینده در اختیار نخواهد داشت. از آنجا که سرشماری روش مقرون به صرفه ای برای گردآوری داده ها نیست، این شرکت تصمیم می گیرد که بعد خانوار خریدار را به روش نمونه گیری انجام دهد.

4	1	3	3	5	2	7	2	3
---	---	---	---	---	---	---	---	---

3.3 = میانگین

در این مثال تعداد میانگین اعضای خانوار پارامتر، آماره بعد خانوار نمونه و برآورد نقطه ای 3.3 است.

فرآیند رسیدن به برآورد



فرض کنید جامعه ای از که 6 نفر تشکیل شده باشد با در آمد ماهیانه بر حسب میلیون تومان به صورت زیر است.

4	1	0	3	5	2
---	---	---	---	---	---

میخواهیم به دو روش نمونه گیری و میانگین گیری پارامترها و برآورد های جامعه بالا را به دست بیاوریم:

ابتدا میانگین جامعه را به دست می آوریم که عدد _____ می شود که یک _____ است.

بعد از به دست آوردن میانگین میخواهیم به روش نمونه گیری و به صورت تقریبی برآوردی نزدیک میانگین به دست آوریم. ابتدا نمونه ها را یکی یکی میگیریم که احتمال انتخاب هر عضو $1/6$ است. اگر احتمالاً شخصی که در آمدش صفر است را انتخاب کنیم، میانگین کل را صفر در نظر میگیریم که به اندازه کافی به پارامتر جامعه نزدیک نیست.

به نظر شما این روش چه نقاط ضعف و قوتی دارد؟

شما چه روشی برای احتمال نزدیک تر شدن برآورد و پارامتر دارید؟

سپس نمونه ها را دوتا دوتا میگیریم که امکان مشاهده _____ نمونه است و میانگین دو نمونه را به عنوان برآورد در نظر می گیریم. جدول زیر احتمال مشاهده هر یک از برآورد ها برای نمونه های دوتایی است.

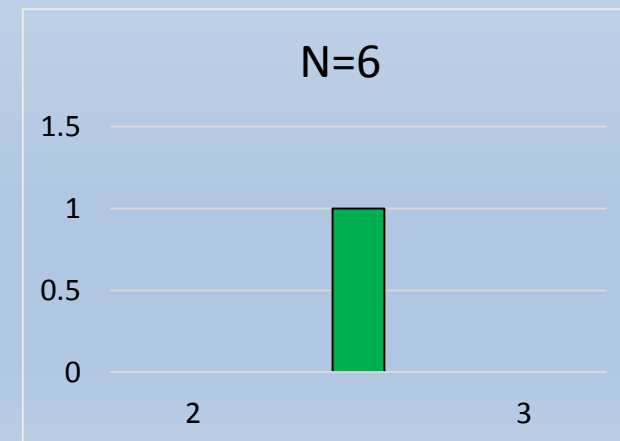
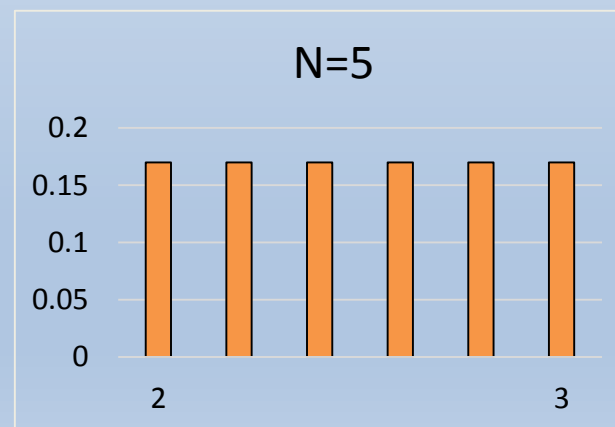
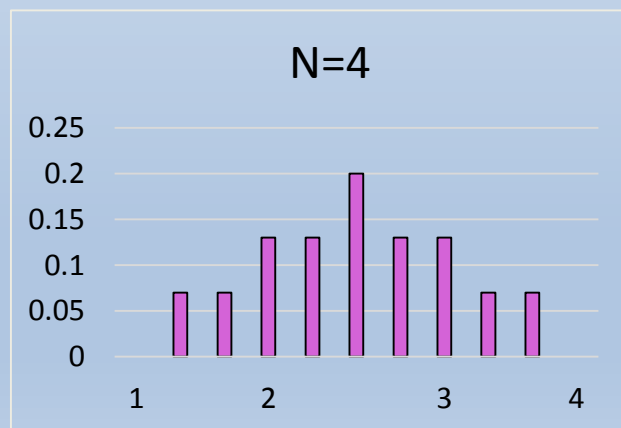
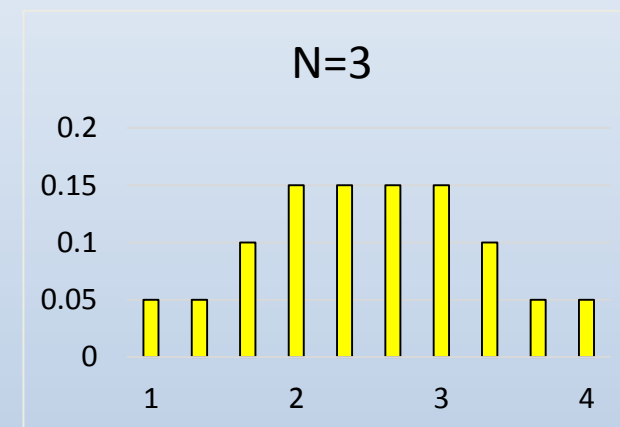
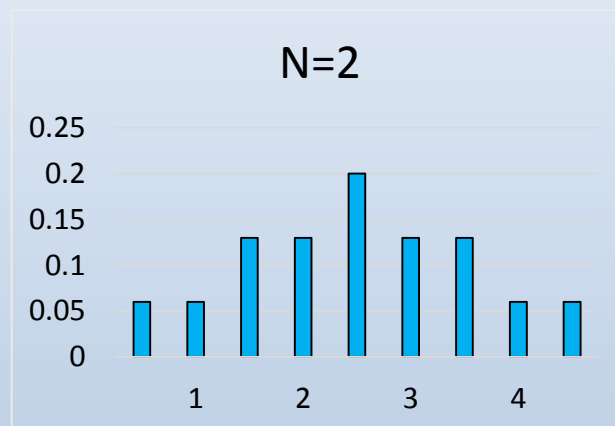
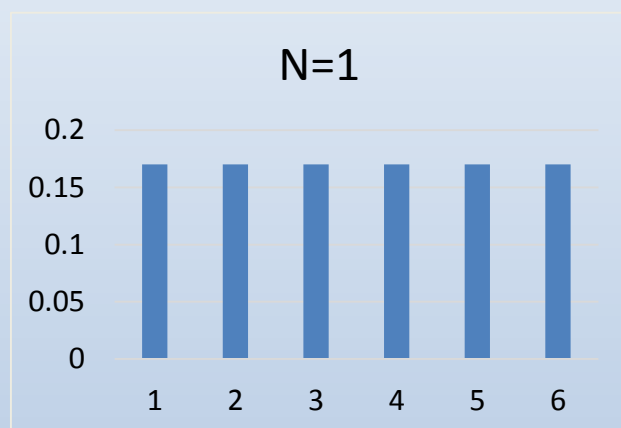
نمونه	0,1	0,2	0,3 1,2	0,4 1,3	0,5 1,4 2,3	1,5 2,4	2,5 3,4	3,5	4,5
میانگین	0.5	1	1.5	2	2.5	3	3.5	4	4.5
احتمال	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{1}{15}$

اکنون مانند قبل
نمونه گیری را
انجام می‌دهیم. اما
اینبار هر نمونه
سه یا چهار تایی
هست.

نمونه	0,1,2	0,1,3	0,1,4 0,2,3	0,1,5 0,2,4 1,2,3	0,2,5 0,3,4 1,2,4	0,3,5 1,4,3 1,2,5	0,4,5 1,3,5 1,2,4	1,4,5 2,3,5	2,4,5	3,4,5
میانگین	$\frac{3}{3}$	$\frac{4}{3}$	$\frac{5}{3}$	$\frac{6}{3}$	$\frac{7}{3}$	$\frac{8}{3}$	$\frac{9}{3}$	$\frac{10}{3}$	$\frac{11}{3}$	$\frac{12}{3}$
احتمال	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{3}{20}$	$\frac{3}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{1}{20}$

نمونه	0,1,2,3	0,1,2,4	0,1,2,5 0,1,3,4	0,1,3,5 0,2,3,4	0,1,4,5 0,2,3,5 1,2,3,4	0,2,4,5 1,2,3,5	0,2,4,5 1,2,4,5	1,3,4,5	2,3,4,5
میانگین	$\frac{6}{4}$	$\frac{7}{4}$	$\frac{8}{4}$	$\frac{9}{4}$	$\frac{10}{4}$	$\frac{11}{4}$	$\frac{12}{4}$	$\frac{13}{4}$	$\frac{14}{4}$
احتمال	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{1}{15}$

اکنون جدول های کشیده شده در اسلاید های قبل را به صورت نمودار در می آوریم تا بتوانیم بهتر آن ها را تحلیل کنیم. در نمودار های زیر محور عرض ها احتمال و محور طول ها میانگین را نشان میدهد.



از نمودار های اسلاید قبل نتیجه می گیریم که با افزایش اندازه نمونه برآورد ها به میانگین جامعه که پارامتر جامعه است نزدیک تر میشوند.

به بیان دیگر در هر نمودار با افزایش اندازه نمونه، انحراف معیار برآورد های پارامتر کمتر می شود. در نتیجه هرچقدر انحراف معیار کمتر باشد برآورد بهتر است.
سوال اساسی این است که انحراف معیار برآورد جامعه چقدر است؟
خوشبختانه آمار شناسان جواب این سوال را داده اند. البته برای سادگی محاسبات، فرض شده که جامعه نامتناهی است.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

هر چند انحراف معیار جمعه معمولاً معلوم نیست، ولی این رابطه حدس ما را ثابت کرده است: با افزایش اندازه نمونه انحراف معیار کاهش میابد. به عبارتی دیگر برآورد دقیق تر یا خطای کمتری برای برآورد میانگین جامعه داریم.

برآورد بازه ای

فرض کنید بعد از یک امتحان ساده از شما پرسیده میشود که نمره شما چند میشود. در این صورت شما بدون تردید نمره خود را اعلام می کنید (مثلا 15)

حال فرض کنید که امتحان کمی دشوار بوده و نمی توانید نمره خود را به صورت دقیق بیان کنید. در نتیجه سعی میکنید آن را به صورت یک بازه همراه با یک درصد بیان کنید. مثلا با احتمال (یا اطمینان) نود درصد بین نوزده تا شانزده. طبیعتا هرچقدر فاصله دو عدد بازه نزدیک تر و درصد اطمینان بیشتر باشد، برآورد دقیق تر است.

برآورد بازه ای یا بازه اطمینان پارامتر جامعه عبارت است از بازه ای عددی برای پارامتر به همراه یک درصد اطمینان که به ضریب اطمینان شهرت دارد.

فهمیدیم که دقت برآورد با انحراف معیار جامعه رابطه عکس و با تعداد اعضای نمونه رابطه مستقیم دارد. با استفاده از این دو کمیت می توان بازه ای برای پرامتر مشخص کرد که به احتمال نود و پنج درصد در آن قرار دارد و از رابطه زیر حساب میشود:

برآورد بازه ای برای میانگین جامعه

اگر نمونه ای به تصادف در اختیار داشته باشیم، با اطمینان بیش از 95 درصد میتوانیم بگوییم:

$$\bar{x} - \frac{2\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{2\sigma}{\sqrt{n}}$$

μ میانگین جامعه و σ انحراف معیار جامعه است.

مثال:

اگر نمونه ای با داده های 0,5,2,1 داشته باشیم آنگاه:

$$\sigma \approx 2.74 \quad \bar{x} = 4$$

$$1.26 \leq \mu \leq 6.74$$

فرض کنید برای به دست آوردن نسبتی از جامعه که ویژگی خاصی را دارند نمونه ای n تایی به تصادف انتخاب کنیم سپس اگر m تا از آنان ویژگی مورد مطالعه را داشته باشند. می توان بگوییم:

«تقریباً $\frac{m}{n}$ از افراد جامعه آن ویژگی را دارند»

آمارشناسان برای این گونه از نسبت ها برآورد را به صورت زیر تعیین کرده اند.

برآورد بازه ای نسبت، با اطمینان 95 درصد

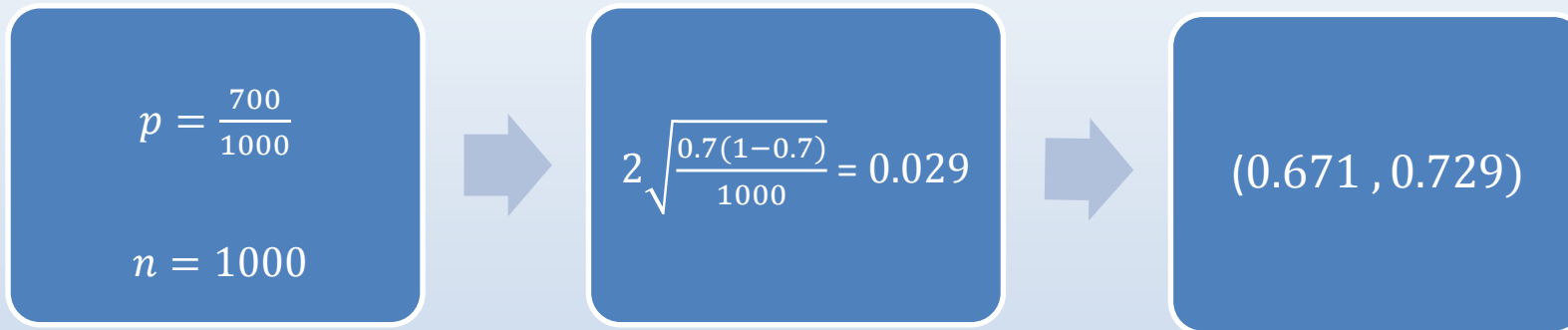
اگر از جامعه ای نسبتاً بزرگ، n نمونه تصادفی ساده انتخاب کنیم و m تای آنها ویژگی مورد مطالعه را داشته باشد، آنگاه نسبت واقعی افراد از جامعه که ویژگی مورد نظر را دارند، با اطمینان 95% در بازه زیر قرار دارد:

$$\left(p - 2\sqrt{\frac{p(1-p)}{n}}, \quad p + 2\sqrt{\frac{p(1-p)}{n}} \right)$$

سه نکته:

1. بازه به تعداد نمونه بستگی دارد، نه اعضای جامعه
2. میتوانیم با افزایش ضریب رادیکال ها و بدون تغییر تعداد نمونه ها، اطمینان را از 95 درصد بالاتر ببریم.
3. بازه دو سر باز است.

یک موسسه نظرسنجی معتبر، قبل از برگزاری انتخابات ریاست جمهوری از یک نمونه 1000 نفری واجد شرایط یک نظر سنجی میکند که آیا در انتخابات شرکت میکنید؟ اگر جواب 700 نفر مثبت بوده باشد یک بازه اطمینان 95 درصدی برای درصد شرکت کنندگان به وجود آورد. حل:



حال فرض کنید میخواهیم طوری نظرسنجی را انجام دهیم که طول بازه اطمینان کمتر از یک درصد باشد؛ یعنی نمونه گیری را دقیق تر کنیم. حال سوال اینجاست که باید تقریباً چند نمونه بگیریم؟

مشکل اساسی ما در این مسئله این است که مقدار p را نداریم. پس بازه ی محافظه کارانه ی $\left(p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right)$ را تشکیل میدهیم و از آن برای بدست آوردن n استفاده میکنیم. این بازه از جایگذاری مقدار $\frac{1}{2}$ بجای p بدست می آید که در این صورت مقدار $p(1-p)$ به بیشینه ی خود میرسد. در این بازه فاصله ی دو سر بازه $\frac{2}{\sqrt{n}}$ است. پس با برابر قرار دادن مقدار مذکور با فاصله مورد نظر میتوان n را بدست آورد.

منابع

- خیلی سبز
- کتاب درسی

پایان

حامد نکویی 

فاروق کریمی زاده شغویی 

امیرحسین نورنیا 

سینا نصرتی 