بسم الله الرحمن الرحيم

روشهای نمونه گیری

مجتبی جهانی فر
دانشکده روانشناسی و علوم تربیتی
نیم سال اول سال تحصیلی ۹۹-۹۸

درس سوم

نمونه گیری تصادفی ساده

برآورد میانگین و مجموع واحدهای جامعه

- ساده ترین روش نمونه گیری احتمالاتی نمونه گیری تصادفی ساده است.
- درک این روش، کمک زیادی به فهم سایر نمونه گیری ها خواهد کرد، به همین خاطر به آن نمونه گیری مادر می گویند.
- روش نمونه گیری ساده به دو روش با جایگذاری و بدون جایگذاری انجام می شود.
- در این روش، n واحد از جامعه ای به حجم N واحد انتخاب می شود.(n<N)
- از میان N واحد جامعه، می توان k نمونه به حجم n انتخاب کرد به طوری که $k = \binom{N}{n}$
- شانس انتخاب همه k نمونه ممکن یکسان و برابر با $\frac{1}{k}$ است.(بدون جایگذاری)
- شانس انتخاب همه واحدهای جامعه یکسان و برابر با $\frac{1}{N}$ است. (استخراج در نوبت اول، نوبت دوم و....با هم برابر است)(بدون جایگذاری)
- بیشتر برای جوامع کوچک، و یکنواخت به کار می رود که در دسترس هم باشد.

یک مثال آموزنده

$U = \{A, B, C, D\}$

$n = ۳, N = ۴, k = ۴$

- نمونه ها به حجم ۳ (ترتیب انتخاب مهم نیست)عبارتند از :

ABC ABD ACD BCD

- نمونه ها به حجم ۳ (ترتیب انتخاب مهم است) عبارتند از :

ABC	ABD	ACD	BCD
ACB	ADB	ADC	BDC
BAC	BAD	CAD	CBD
BCA	BDA	CDA	CDB
CAB	DAB	DAC	DBC
CBA	DBA	DCA	DCB

واحدهای نمونه گیری را چگونه انتخاب کنیم؟

- قرعه کشی
- نرم افزارهای رایانه ای
- جدول اعداد تصادفی

جدول اعداد تصادفی

- هر کدام از رقم های جدول از اعداد تصادفی از ۰ تا ۹ تشکیل شده اند.
- همه رقمها به طور تصادفی و مستقل از توزیع تولید شده اند.
- رقم های ۰ تا ۹ شانس یکسانی برای انتخاب شدن در هر خانه را دارند.
- ساختار جدول به گونه ای است که اگر بخواهیم اعداد چند رقمی بسازیم احتمال انتخاب ثابت باقی می ماند.
- در این روش باید همه عناصر جامعه شماره گذاری شوند.

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

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

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

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

تولید اعداد تصادفی با استفاده از رایانه

- برخی برنامه های رایانه ای می توانند با نوشتن چند خط برنامه ساده برای شما اعداد تصادفی ایجاد کنند.
- برنامه ای R و MATLAB از جمله آنها هستند.

https://www.r-project.org/

https://www.mathworks.com/products/matlab.html

▪ دستور زیر در برنامه R می تواند اعداد تصادفی دلخواه بین دو عدد مشخص را به صورت اعشاری تولید کند.

rannums <- runif(50, 100, 250)

rannums

247.4847 140.2535 239.0994 177.0807 157.9139 102.1751 246.3902 228.2672 233.7613 185.9351 147.4063 172.4284 133.1633

226.1966 187.5667 102.3940 191.0745 221.3734 169.0484 232.0205 233.8510 243.9901 123.3032 149.8102 172.8068 216.7875

150.8160 116.2874 213.3296 187.7297 193.2595 239.1499 189.3068 103.7160 225.8157 199.8060 130.2217 216.7823 156.0345

223.6371 170.4888 188.6956 247.1873 231.8264 217.4117 113.7333 239.1044 186.9935 147.5004 176.3961

floor(rannums)

247 140 239 177 157 102 246 228 233 185 147 172 133 226 187 102 191 221 169 232 233 243 123 149 172 216 150 116 213 187 193 239 189 103 225 199 130 216 156 223 170 188 247 231 217 113 239 186 147 176

▪ دستور زیر در برنامه R می تواند اعداد تصادفی دلخواه بین دو عدد مشخص را به صورت صحیح تولید کند.

```
rannums1 <- sample(20:100,25, replace=T)
rannums2 <- sample(20:100,25, replace=F)
```

rannums1

35 45 46 45 34 47 29 73 66 31 64 77 28 81 81 68 91 86 30 96 86 20 24 61 95

Rannums2

75 78 28 31 26 92 33 56 77 100 73 66 43 53 88 99 61 69 81 97 54 86 20 93 57

▪ دستور مشابه در برنامه MATLAB به صورت زیر است:

```
a = 20;
b = 100;
rannums = (b-a).*rand(25,1) + a
floor(rannums)
```

معرفی چند نماد

- هر ویژگی قابل اندازه گیری را با Y نمایش می دهیم.(قد، وزن، درآمد، نمره فیزیک و....)
- حجم جامعه را با N نمایش می دهیم و اندازه مشخصه تحت بررسی واحدهای جامعه به صورت زیر است:

$$Y_۱, Y_۲, Y_۳, \ldots, Y_N$$

- حجم نمونه انتخاب شده به روش تصادفی ساده n است و اندازه مشخصه تحت بررسی واحدهای نمونه به صورت زیر است

$$Y_۱, Y_۲, Y_۳, \ldots, Y_n$$

بدین ترتیب میانگین های جامعه و نمونه و مجموع واحدهای جامعه و نمونه به صورت زیر تعریف می شوند :

$$\overline{Y}_N = \frac{۱}{N}\sum_{۱}^{N} Y_i \qquad \overline{Y}_n = \frac{۱}{n}\sum_{۱}^{N} Y_i \qquad T_N = \sum_{۱}^{N} Y_i \qquad T_n = \sum_{۱}^{n} Y_i$$

❑ قضیه ۱ : میانگین نمونه تصادفی ساده بدون جایگذاری، برآورد کننده ناریب میانگین جامعه است. اگر k نمونه متفاوت از جامعه استخراج شود، برای هر کدام میانگین محاسبه شود و سپس از همه میانگین‌ها میانگین بگیریم، عدد بدست آمده همان میانگین جامعه است.

$$E(\overline{Y}_n) = \overline{Y}_N$$

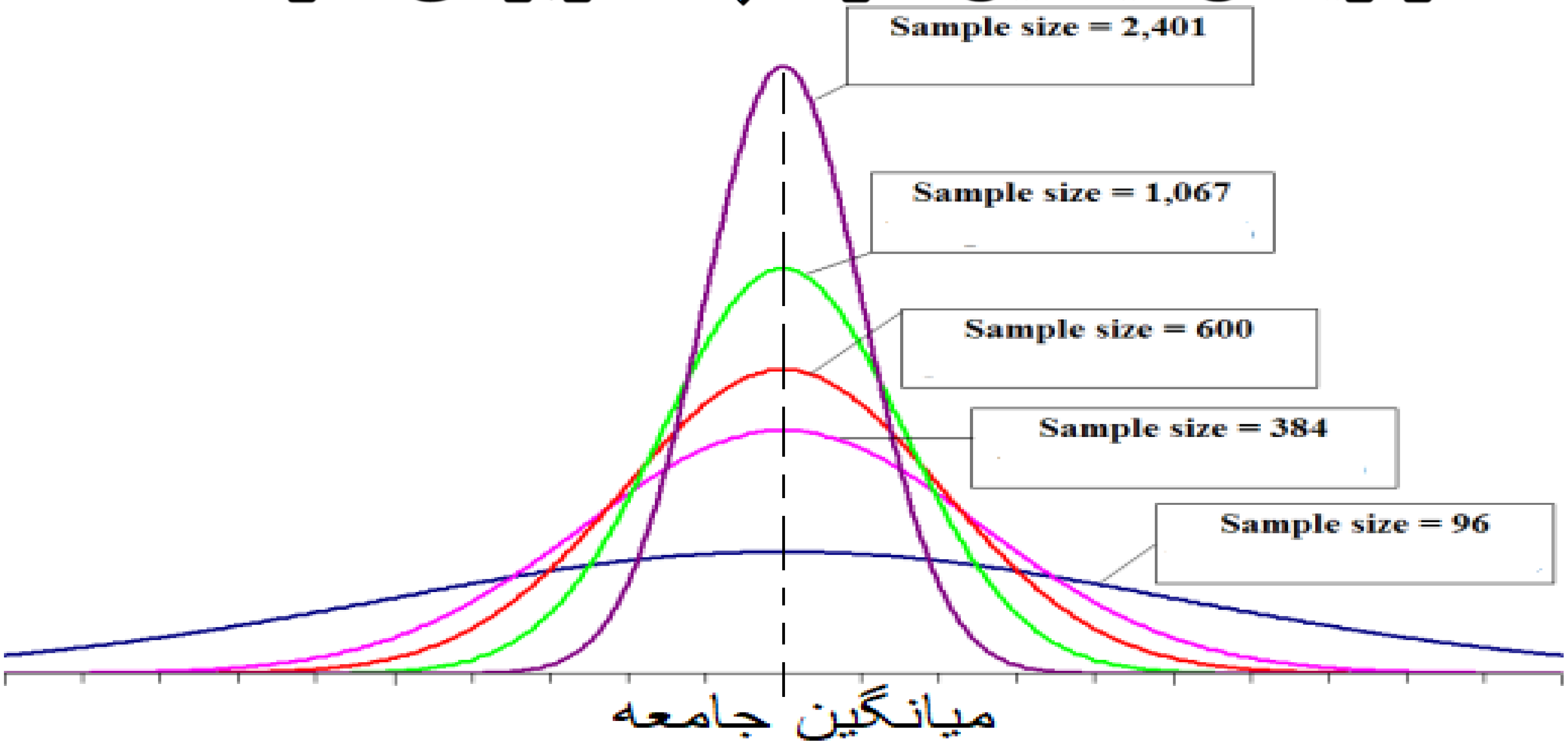
❑ قضیه ۲ : هرگاه نمونه ای به حجم n از جامعه ای به حجم N بدون جایگذاری انتخاب شود، واریانس میانگین از رابطه زیر بدست می آید:

$$V(\overline{Y}_n) = \left(\frac{N-n}{n}\right)\frac{S^2}{n} = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}$$

$$S^2 = \frac{1}{N-1}\sum_{1}^{N}(Y_i - \overline{Y}_N)^2$$

تغییرات جامعه

واریانس میانگین نمونه چه کاربردی دارد؟



- واریانس میانگین نمونه، میزان دقت ما در نمونه گیری را نشان می دهد.
- این واریانس نشان می دهد که چه مقدار از میانگین جامعه دور و یا به آن نزدیک شدیم.
- پس هر چه کوچکتر باشد نمونه گیری دقیق تر است.

واریانس میانگین نمونه را چگونه محاسبه کنیم؟

$$V(\overline{Y}_n) = (\frac{N-n}{n})\frac{S^۲}{n}$$

مجهول

$$S^۲ = \frac{۱}{N-۱}\sum_{۱}^{N}(Y_i - \overline{Y}_N)^۲$$

معلوم

واریانس نمونه به صورت زیر تعریف می شود:

$$s^۲ = \frac{۱}{n-۱}\sum_{۱}^{n}(Y_i - \overline{Y}_n)^۲$$

<u>واریانس نمونه</u> برآوردگر <u>ناااریب</u> برای <u>تغییرات جامعه</u> است پس می تواند <u>جایگزین مناسبی</u> برای آن باشد.

$$\widehat{V}(\overline{Y}_n) = (\frac{N-n}{N})\frac{s^۲}{n}$$

❑ قضیه ۳ : حاصل ضرب حجم جامعه و میانگین نمونه، برآورد گری ناریب برای مجموع واحدهای جامعه است.

$$\hat{T}_n = N\bar{Y}_n \qquad \text{مجهول} \longrightarrow T_N = \sum_{1}^{N} Y_i$$

❑ همچنین واریانس مجموع واحدهای جامعه به صورت زیر است:

$$V(\hat{T}_n) = N^{۲}\left(\frac{N-n}{N}\right)\frac{S^{۲}}{n} \longleftarrow \text{مجهول}$$

پس باید برآورد شود، بنابرین برآورد واریانس مجموع واحدهای جامعه به صورت زیر است:

$$\hat{V}(\hat{T}_n) = N^{۲}\left(\frac{N-n}{N}\right)\frac{s^{۲}}{n}$$

حدود اطمینان میانگین جامعه و مجموع واحدهای جامعه

- طبق قضیه حد مرکزی میانگین های نمونه توزیع طبیعی دارند

$$\bar{Y}_n \sim N(\bar{Y}_N, V(\bar{Y}_n))$$

- درصورتی که از نمره های Z استفاده کنیم باز هم توزیع طبیعی است اما با میانگین ۰ و ۱

$$Z = \frac{Y_n - \bar{Y}_n}{\sigma} \sim N(۰,۱)$$

- حال می توانیم در سطح اطمینان α حد بالا و حد پایین را برای میانگین جامعه تعریف کنیم

$$\bar{Y}_n - zs\sqrt{(\frac{۱}{n} - \frac{۱}{N})} < \bar{Y}_N < \bar{Y}_n + zs\sqrt{(\frac{۱}{n} - \frac{۱}{N})}$$

طول متناظر با رتبه درصدی $1-\frac{\alpha}{2}$

میانگین نمونه

$$\left[\bar{Y}_n - zs\sqrt{\left(\frac{1}{n}-\frac{1}{N}\right)},\ \bar{Y}_n + zs\sqrt{\left(\frac{1}{n}-\frac{1}{N}\right)}\right]$$

حد بالا برآورد میانگین جامعه

حد پایین برآورد میانگین جامعه

- برای حجم نمونه های کمتر از ۵۰ بهتر است از توزیع t با درجه آزادی n-1 استفاده کرد
- برای محاسبه حدود اطمینان مجموع واحدهای جامعه حدود بالا را در N ضرب کنید.

- کمیت $\frac{N-n}{n}$ ضریب تصحیح برای جامعه محدود نامیده می شود.
- هرگاه نسبت حجم نمونه به حجم جامعه کمتر از ۰/۰۵ باشد این ضریب تبدیل به ۱ خواهد شد و برآورد واریانس به صورت زیر خواهد بود:

$$\hat{V}(\overline{Y}_n) = \frac{s^۲}{n}$$

مثال ۲–۴

مثال ۳–۴

مثال ۴–۴

نمونه گیری تصادفی با جایگذاری