جزوه ی روش های نمونه گیری یک

۲۹ اردیبهشت ۱۴۰۳

منبع:

نظریه نمونه گیری و کاربردهای آن،تألیف:علی عمیدی

استاد درس:

استاد فیروزه حقیقی

نویسنده:

محمدرضا شهریارکشه

مباحث میان ترم اول:

هدف نمونه گیری

θ: پارامتر $\leftarrow$ مقدار مشخصی دارد ولی نامعلوم است. برای مثال:

$$\mu \text{ میانگین } \overset{estimator^{1}}{\longleftarrow} \bar{y_n}$$

$$P \text{ نسبت } \overset{estimator}{\longleftarrow} \bar{p}$$

انتخاب نمونه با حجم زیاد اگر چه دقت بیشتر همراه است، ولی باعث هزینه بیشتر نیز می شود.برای مثال به تکثیر پرسشنامه یا هزینه طبقه بندی و ثبت اطلاعات نمونه می توان اشاره کرد.

چند تعریف

جامعه:

مجموعه ای از اشیاء/افراد که قرار است استنباط خاصی در مورد آنها صورت بگیرد.

استنباط:

بررسی شواهد موجود و رسیدن به یک حقیقت

نمونه

بخشی از جامعه که به روش معیین تعیین شده است.

صفت متغییر

ویژگی از شی/فرد به شی/فرد دیگر تغییر کند. انواع متغییرها:

۱. کمی:

(آ) گسسته:تعداد فرزندان یک خانواده

(ب) پیوسته:وزن افراد

۲. کیفی:

(آ) اسمی:گروه خونی

(ب) رتبه ای:مدارک دانشگاهی

[1] برآوردگر

داده

مقادیر مشخص شده ی متغییر، داده است. تقسیم بندی داد ها بر اساس نقش پژوهشگر در تولید آن:

۱. داده های مشاهده ای[۲] مبتنی بر مشاهده، پژوهشگر صرفاً داد ها را مشاهده و ثبت می کند.

۲. داده های آزمایشی [۳] پژوهشگر داده ها را تولید می کند. مثل:پرتاب تاس

آمارگیری:

فعالیتی است که در آن اطلاعات مورد نیاز برای یک ویژگی بر اساس بخشی یا تمام جامعه و به روشی معین انجام می شود.
انواع آمارگیری:

۱. سرشماری: براساس کل جامعه انجام می شود.

۲. نمونه گیری: براساس بخشی لز جامعه انجام می شود.

مزایای نمونه گیری

۱. سرعت بالاتر

۲. هزینهٔ کمتر

۳. کیفیت بالاتر در جمع داده ها

۴. محاسبات کمتر و در نتیجه دقت بیشتر

۵. حفظ اعضای جامعه برای مثال در نمونه گیری از خط تولید

خطاهای آمارگیری:

۱. خطای نمونه گیری:
ناظر بر این است که برآورد پارامتر بر اساس بخشی از داده(نمونه) انجام می شود.

۲. خطای غیر نمونه گیری:
هر خطایی به غیر از خطای نمونه گیری. برای مثال: اشتباه در پرکردن پرسشنامه

انواع نمونه گیری:

۱. نمونه گیری احتمالاتی: هر فرد جامعه احتمال مشخص و غیر صفر برای ورود به نمونه دارد.

۲. نمونه گیری غیر احتمالاتی: هر فرد جامعه احتمال مشخص و غیر صفر برای ورود به نمونه دارد.

[۲] Survey Study
[۳] Experiment Study

جامعه هدف:

۱. Who

۲. When

۳. Where

مثال: دانشجویان نمونه گیری ۱ ترم دوم ۱۴۰۲ دانشگاه تهران
چارچوب: لیست تمام واحدهای جامعه
ویژگی های یک برآورد خوب:

۱. نااریبی:

$$E(\hat{\theta}) = \theta$$

۲. مقدار واریانس با افزایش حجم نمونه کاهش بیابد.

۳.

$$CV^{۴} = \frac{Var(\hat{\theta})}{E(\hat{\theta})}$$

نمونه گیری تصادفی ساده (بدون جایگزاری و با جایگزاری)

[۵]

$$\begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ \vdots \\ y_N \end{pmatrix} \xrightarrow{SRS} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

(آ) N: حجم جامعه متناهی است.

(ب) n: حجم نمونه

این روش برای مواقعی است که می خواهیم از جامعه ای به حجم N نمونه ای به حجم n بگیریم به گونه ای که تمام نمونه های ممکن شانس یکسان در انتخاب شدن داشته باشند.

[۴] ضریب تغییرات
[۵] SRS: Simple Random Sample

بدون جایگزاری:

مثال: تمام نمونه های سه بعدب از جامعه زیر به همرا احتمال انتخاب شدن آن بنویسید.

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

تعداد نمونه های ممکن برابر است با:

$$\binom{4}{3} = 4$$

نمونه های ممکن در یک نمونه تصادفی ساده بدون جایگزاری:

$$(1,2,3) \to p = \frac{1}{4} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{2} \times 6 = \frac{1}{4}$$
$$(1,3,4) \to p = \frac{1}{4}$$
$$(2,3,4) \to p = \frac{1}{4}$$
$$(1,2,4) \to p = \frac{1}{4}$$

تعداد نمونه های n تایی به روش SRS از جامعه ی N عضوی به شرط حضور فرد خاص در جامعه:

$$\binom{N-1}{n-1}$$

برآورد میانگین جامعه:

$$\begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ \vdots \\ y_N \end{pmatrix} \xrightarrow{n} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

n : حجم نمونه

$\bar{y_N}$: میانگین جامعه
$\bar{y_N} = \frac{1}{N}\sum_{i=1}^{N} y_i$

S^2 واریانس جامعه
$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y_N})^2$:

σ^2 واریانس جامعه
$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y_N})^2 \Rightarrow N\sigma^2 = (N-1)S^2$

$\bar{y_N}$: میانگین نمونه ای
$\bar{y_n} = \frac{1}{n}\sum_{i=1}^{n} y_i$

s^2 : واریانس نمونه ای
$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y_n})^2$

اثبات کتید که $\bar{y_n}$ یک برآوردگر مناسب برای $\bar{y_N}$ است.

$$E(\bar{y_n}) = \bar{y_N}$$

متغییر تصادفی Z_i را به شکل زیر تعریف می کنیم.

$$Z_i = \begin{cases} 1 & y_i \; appear \; in \; sample \\ 0 & Other \; where \end{cases}$$

$$Z_i \sim B(1, \frac{1}{N})$$

$$E(Z_i) = \frac{n}{N}$$

$$Var(Z_i) = \frac{n}{N}(1 - \frac{n}{N})$$

$$Cov(Z_i, Z_j) = \frac{n}{N}\frac{n-1}{N-1} - \frac{n^2}{N^2}$$

می توان گفت که:

$$\bar{y_n} = \frac{Z_1y_1 + Z_2y_2 + \cdots + Z_Ny_N}{n}$$

برای اثبات:

$$E(\bar{y_n}) = E[\frac{\sum_{i=1}^{N} Z_iy_i}{n}] = \frac{1}{n}E[\sum_{i=1}^{N} Z_iy_i] = \frac{1}{n}\sum_{i=1}^{N} y_iE(Z_i)$$

Z	1	0
$P(Z=z)$	$P(Z=1)$	$P(Z=0)$

نتیجه می دهد که:

$$E(Z_i) = P(Z=1) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$ [۶]

ئر نهایت:

$$E(\bar{y_n}) = \frac{1}{n}\sum_{i=1}^{N} y_i P(Z_i=1) = \frac{1}{\cancel{n}}.\frac{\cancel{n}}{N}\sum_{i=1}^{N} y_i = \frac{\sum_{i=1}^{N} y_i}{N} = \bar{y_N}$$

$$Var(\bar{y_n}) = (1-\frac{n}{N})\frac{S^2}{n}$$

اثبات:

میدانیم که:

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y_N})^2 = \frac{1}{N-1}\{\sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N}\bar{y_N}^2 - 2\sum_{i=1}^{N} y_i\bar{y_N}\} =$$
$$= \frac{1}{N-1}\{\sum_{i=1}^{N} y_i^2 + N\bar{y_N}^2 - 2\bar{y_N}\sum_{i=1}^{N} y_i\} = \frac{1}{N-1}\{\sum_{i=1}^{N} y_i^2 - N\bar{y_N}^2\}$$

همچنین:

$$(\sum X_i)^2 = \sum X_i^2 + \sum_{i\neq j}\sum X_i X_j$$

[۶] احتمال انتخاب نمونه با وجود یک عضو خاص در آن

$$Var(\bar{y_n}) = Var[\frac{1}{n}\sum_{i=1}^{N} Z_i y_i] = \frac{1}{n^2}\{\sum_{i=1}^{N} Var(Z_i y_i) + \sum\sum_{i\neq j} Cov(Z_i y_i, Z_j y_j)\} =$$

$$= \frac{1}{n^2}\{\sum_{i=1}^{N} y_i(E(Z_i^2) - (E(Z_i))^2) + \sum\sum_{i\neq j} y_i y_j Cov(Z_i, Z_j)\} =$$

$$= \frac{1}{n^2}\{\sum_{i=1}^{N} y_i(P(Z_i = 1) - (P(Z_i = 1))^2) + \sum\sum_{i\neq j} y_i y_j Cov(Z_i, Z_j)\} =$$

$$= \frac{1}{n^2}\{\sum_{i=1}^{N} \frac{n}{N}(1 - \frac{n}{N}) + \sum\sum_{i\neq j} y_i y_j \frac{n}{N}(\frac{n-1}{N-1} - \frac{n}{N})\} =$$

$$= \frac{1}{n^2}.\frac{n}{N}.\frac{N-n}{N}\{\sum_{i=1}^{N} y_i^2 - \frac{1}{N-1}\sum\sum_{i\neq j} y_i y_j\}$$

$$=$$

$$\frac{1}{n^2}.\frac{n}{N}.\frac{N-n}{N}\{\sum_{i=1}^{N} y_i^2 - \frac{1}{N-1}(\sum_{i=1}^{N} y_i)^2 + \frac{1}{N-1}\sum_{i=1}^{N} y_i^2\}$$

$$=$$

$$\frac{1}{n^2}.\frac{n}{N}.\frac{N-n}{N}\frac{1}{N-1}\{(N-1)\sum_{i=1}^{N} y_i^2 - (\sum_{i=1}^{N} y_i)^2 + \sum_{i=1}^{N} y_i^2\} =$$

$$= \frac{1}{n^2}.\frac{n}{N}.\frac{N-n}{N}\frac{1}{N-1}\{N\sum_{i=1}^{N} y_i^2 - (\sum_{i=1}^{N} y_i)^2\} =$$

$$= \frac{1}{n^2}.\frac{n}{N}.\frac{N-n}{N}\frac{1}{N-1}\{N\sum_{i=1}^{N} y_i^2 - N^2\bar{y_N}^2\} =$$

$$= \frac{1}{n^2}.\frac{n}{N}.\frac{N-n}{N}.N\{\frac{1}{N-1}(\sum_{i=1}^{N} y_i^2 - N\bar{y_N}^2)\} = \frac{1}{n^2}.\frac{n}{N}.\frac{N-n}{N}.NS^2 =$$

$$= \frac{1}{\cancel{n^2}}.\frac{\cancel{n}}{\cancel{N}}.\frac{N-n}{N}.\cancel{N}S^2 = (1 - \frac{n}{N})\frac{S^2}{n}$$

از طرفی S یک آماره ی مربوط به جامعه است و مقدار آن نامعلوم است. کافی است ثابت کنیم که $(1-\frac{n}{N})\frac{s^2}{n}$ یک برآوردگر خوب برای $(1-\frac{n}{N})\frac{S^2}{n}$ است.یعنی:

$$E[(1-\frac{n}{N})\frac{s^2}{n}] = (1-\frac{n}{N})\frac{S^2}{n} \Longrightarrow \cancel{\frac{1}{n}(1-\frac{n}{N})}E(s^2) = \cancel{\frac{1}{n}(1-\frac{n}{N})}S^2 \Rightarrow E(s^2) = S^2$$

می دانیم که:

$$s^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y_n})^2}{n-1}=\frac{\sum_{i=1}^{n}y_i^2-n\bar{y_n}^2}{n-1}$$

اثبات:

$$E(s^2)=E[\frac{\sum_{i=1}^{n}y_i^2-n\bar{y_n}^2}{n-1}]=\frac{1}{n-1}\{E[\sum_{i=1}^{n}y_i^2-n\bar{y_n}^2]\}=\frac{1}{n-1}\{\frac{n}{N}\sum_{i=1}^{N}y_i^2-nE(\bar{y_n}^2)\}$$

$$=\frac{1}{n-1}\{\frac{n}{N}\sum_{i=1}^{N}y_i^2-\frac{N-n}{N}S^2-n\bar{y_N}^2\}=\frac{1}{n-1}\frac{n}{N}\{\sum_{i=1}^{N}y_i^2-N\bar{y_N}^2-\frac{N-n}{n}S^2\}=$$

$$=\frac{1}{n-1}\frac{n}{N}\{(N-1)S^2-\frac{N}{n}S^2-S^2\}=\frac{1}{n-1}\frac{n}{N}\{NS^2-\frac{N}{n}S^2\}=$$

$$=\frac{N}{n-1}\frac{n}{N}\{S^2-\frac{1}{n}S^2\}=\frac{\cancel{N}}{\cancel{n-1}}\frac{\cancel{n}}{\cancel{N}}\frac{\cancel{n-1}}{\cancel{n}}S^2=S^2$$

در نتیجه:

$$\hat{Var}(\bar{y_n})=(1-\frac{n}{N})s^2$$

برآورد مجموع جامعه

T_N مجموع جامعه

T_n مجموع نمونه

می خواهیم بررسی کنیم که برآوردگر nT_n یک برآوردگر نااریب هست یا نه

$$E(T_n)=nE(\bar{y_n})=n\bar{y_N}=\frac{n}{N}T_N$$

که مشاهده کردیم نااریب نیست . با استفاده از خاصیت خطی امید در می یابیم که برآوردگر $\frac{N}{n}T_n$ یک برآوردگر نااریب است.
زیرا:

$$E(\frac{N}{n}T_n)=NE(\bar{y_n})=N\bar{y_N}=T_N$$

$$Var(\frac{N}{n}T_n)=N^2Var(\bar{y_n})=N^2(1-\frac{n}{N})\frac{S^2}{n}$$

$$\hat{Var}(\frac{N}{n}T_n)=N^2Var(\bar{y_n})=N^2(1-\frac{n}{N})\frac{s^2}{n}$$

مثال

دندانپزشک های A و B دندان های ۲۰۰ کودک یک دهکده را بررسی می کنند. دندانپزشک A یک نمونه ی تصادفی مستقل با اندازه ی ۲۰ و تعداد دندان هلی خرااب را مشخص می کند.

تعداد دندان های خراب	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
تعداد بچه ها	۸	۴	۲	۲	۱	۴	۰	۰	۰	۱	۱

دندانپزشک B با همان تکنیک معاینه ۲۰۰ کودک را معاینه می کند وآنهایی را متمایز می کند که هیچ دندان خرابی ندارند. ۶۰ کودک هیچ دندان خرابی ندارند. تعداد دندان های خراب کودکان این دهکده را با استفاده از روش های زیر برآورد کنید.

الف: با استفاده از نتیجه ی معاینات دندانپزشک A

$$T_n = N\bar{y_n} = 200 \times (\frac{0 \times 8 + 1 \times 4 + \ldots 10 \times 1}{20}) = 200 \times 2.1 = 420$$

ب: با استفاده از نتایج دندانپزشک A و B

$$T_n = 140 \times (\frac{0 \times 8 + 1 \times 4 + \ldots 10 \times 1}{12}) = 140 \times 3.5 = 490$$

ج: کدام روش دقیق تر است؟
قسمت ب دقیق تر است زیرا داده های بیشتری در مورد جامعه ی خود داریم و براورد ما در این حالت دقیق تر از حالت الف است.

قضیه ها:

۱ یک زیر نمونه SRS به حجم n_1 از یک نمونه به حجم n_2 از یک جامعه یه حجم N ،خود یک نمونه SRS است.

۲ ترکیب یک نمونه ی SRS به حجم n_1 از یک جامعه به حجم N با یک نمونه ی دیگر SRS به حجم n_2 از یک جامعه $N - n_1$ تایی، ،خود یک نمونه SRS است.

با جایگزاری

ثابت کنید که

$$E(\bar{y_n}) = y_n$$

تعریف می کنیم:
Z_i :تعداد دفعات ظاهر شدن عضو i ام در نمونه

می توان گفت:

$$Z_i \sim B(n, \frac{1}{N})$$
$$E(Z_i) = \frac{n}{N}$$
$$Var(Z_i) = \frac{n}{N}(1 - \frac{1}{N})$$
$$Cov(Z_i, Z_j) = -\frac{n}{N^2}$$

در نتیجه:

$$E(\bar{y_n}) = \frac{1}{n}E(\sum_{i=1}^{n} y_i) = \frac{1}{n}E(\sum_{i=1}^{n} y_i) = \frac{1}{n}E[\sum_{i=1}^{N} Z_i y_i] = \frac{1}{\cancel{n}}\frac{\cancel{n}}{N}\sum_{i=1}^{N} y_i =$$
$$= \frac{\sum_{i=1}^{N} y_i}{N} = \bar{y_N}$$

حکم ثابت شد.

اثبات کنید که:

$$Var(\bar{y_n}) = \frac{\sigma^2}{n} = (1 - \frac{1}{N})\frac{S^2}{n}$$

اثبات:

$$Var(\bar{y_n}) = \frac{1}{n^2}Var(\sum_{i=1}^{n} y_i) = \frac{1}{n^2}Var(\sum_{i=1}^{N} Z_i y_i) =$$

$$= \frac{1}{n^2}\{\sum_{i=1}^{N} y_i^2 Var(Z_i) + \sum\sum_{i\neq j} y_i y_j Cov(Z_i, Z_j)\} = \frac{1}{n^2}\{\frac{n(N-1)}{N^2}\sum_{i=1}^{N} y_i^2 + \sum\sum_{i\neq j} y_i y_j Cov(Z_i, Z_j)\} =$$

$$= \frac{1}{n^2}\{\frac{n(N-1)}{N^2}\sum_{i=1}^{N} y_i^2 - \frac{n}{N^2}\sum\sum_{i\neq j} y_i y_j\} = \frac{1}{n^2}.\frac{n}{N^2}\{(N-1)\sum_{i=1}^{N} y_i^2 - \sum\sum_{i\neq j} y_i y_j\} =$$

$$= \frac{1}{n^2}.\frac{n}{N^2}.(N-1)\{\sum_{i=1}^{N} y_i^2 - \frac{1}{N-1}\sum\sum_{i\neq j} y_i y_j\} =$$

$$= \frac{1}{n^2}.\frac{n}{N^2}.(N-1)\{\sum_{i=1}^{N} y_i^2 - \frac{1}{N-1}(\sum_{i=1}^{N} y_i)^2 + \frac{1}{N-1}\sum_{i=1}^{N} y_i^2\} =$$

$$= \frac{1}{n^2}.\frac{n}{N^2}\{(N-1)\sum_{i=1}^{N} y_i^2 - (\sum_{i=1}^{N} y_i)^2 + \sum_{i=1}^{N} y_i^2\} = \frac{1}{n^2}.\frac{n}{N^2}\{N\sum_{i=1}^{N} y_i^2 - (\sum_{i=1}^{N} y_i)^2\} =$$

$$= \frac{1}{n^2}.\frac{n}{N^2}.N\{\sum_{i=1}^{N} y_i^2 - \frac{1}{N}(\sum_{i=1}^{N} y_i)^2\} = \frac{1}{n^2}.\frac{n}{N^2}.N\{\sum_{i=1}^{N} y_i^2 - N\bar{y_N}^2\} =$$

$$= \frac{1}{n^2}.\frac{n}{N^2}.N.N\sigma^2 = \frac{\sigma^2}{n}$$

حکم ثابت شد.

ثابت کنید که برآوردگر زیر نااریب است:

$$E(\frac{s^2}{n}) = \frac{\sigma^2}{n}$$

اثبات:

$$E(s^2)=\frac{1}{n-1}E(\sum_{i=1}^{n}y_i^2-n\bar{y_n}^2)=\frac{1}{n}\{E(\sum_{i=1}^{n}y_i^2)-nE(\bar{y_n}^2)\}=$$
$$=\frac{1}{n-1}\{E(\sum_{i=1}^{n}y_i^2)-n(Var(\bar{y_n})+[E(\bar{y_n})]^2)\}=\frac{1}{n-1}\{E(\sum_{i=1}^{n}y_i^2)-n(\frac{\sigma^2}{n})-n\bar{y_N}^2\}=$$
$$=\frac{1}{n-1}\{\frac{n}{N}E(\sum_{i=1}^{N}y_i^2)-\sigma^2-n\bar{y_N}^2\}=\frac{1}{n-1}\{\frac{n}{N}(\sum_{i=1}^{N}y_i^2)-\sigma^2-n\bar{y_N}^2\}=$$
$$=\frac{1}{n-1}.\frac{n}{N}\{\sum_{i=1}^{N}y_i^2-N\bar{y_N}^2-\frac{N}{n}\sigma^2\}=\frac{1}{n-1}.\frac{n}{N}\{N\sigma^2-\frac{N}{n}\sigma^2\}=$$
$$=\frac{1}{\cancel{n-1}}.\frac{\cancel{n}}{\cancel{N}}\frac{\cancel{N(n-1)}}{\cancel{n}}\sigma^2=\sigma^2$$

نسبت در یک جامعه

$$\begin{pmatrix}1\\0\\0\\\vdots\\1\end{pmatrix}\xrightarrow{SRS:Without replacing}\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}$$
$$P=\frac{A}{N}\longrightarrow p=\frac{a}{n}$$

کافی است ثابت کنیم که p یک برآوردگر مناسب برای P است. می توان گفت می توان گفت جامعه ی ما مقادیر ۰ و ۱ دارد شامل اعضای دارای وِزگی مورد نظر و اعضای فاقد این ویژگی، یعنی:

$$P=\frac{\sum_{i=1}^{N}y_i}{N}=\bar{y_N}$$
$$p=\frac{\sum_{i=1}^{n}y_i}{n}=\bar{y_n}$$

در نتیجه می دانیم:

بدون جایگزاری

$$E(p) = P$$
$$Var(p) = (1 - \frac{n}{N})\frac{S^2}{n} = \frac{N-n}{N-1}.\frac{P(P-1)}{n}$$
$$S^2 = \frac{1}{N-1}\{\sum_{i=1}^{N} y_i^2 - N\bar{y_N}^2\} = \frac{1}{N-1}\{A - N\frac{A}{N^2}\} =$$
$$= \frac{1}{N-1}NP(1-P)$$

می خواهیم $\hat{Var}(p)$ را محاسبه کنیم.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y_n})^2 = \frac{1}{n-1}\{\sum_{i=1}^{n} y_i^2 - n\bar{y_n}^2\} = \frac{1}{n-1}\{a - np^2\} = \frac{nP}{n-1}(1-P)$$
$$\Rightarrow \hat{Var}(p) = \frac{N-n}{n-1}.\frac{p(1-p)}{N}$$

رابطه ی زیر نیز برای برآورد تعداد اعضای خاص یک جامعه مفید است.

مثال فرض کنید جامعه ای شامل ۵ دانش آموز دختر و پسر، بصورت زیر است:

$$0 \quad 1 \quad 1 \quad 0 \quad 1$$

۱. تمام نمونه های دوتایی ممکن با روش SRS را تشکیل دهید.

$$p = 0 : (0,0)$$
$$p = \frac{1}{2} : (0,1),(0,1),(0,1),(0,1),(0,1),(0,1)$$
$$p = 1 : (1,1),(1,1),(1,1)$$

۲. بر اساس قسمت قبل ثابت کنید:

$$E(p) = P \qquad\qquad Var(p) = \frac{N-n}{N-1}\frac{P(1-P)}{n}$$

داریم:

p	0	$\frac{1}{2}$	1
$\frac{1}{10}$	$\frac{6}{10}$	$\frac{3}{10}$	

در نتیجه:

$$E(p) = 0 \times \frac{1}{10} + \frac{1}{2} \times \frac{6}{10} + 1 \times \frac{3}{10} = \frac{6}{10} = P$$
$$Var(p) = E(p^2) - [E(p)]^2 = \frac{9}{20} - \frac{36}{100} = \frac{9}{100}$$

پیدا کردن حجم نمونه

می دانیم که:

$$\hat{\theta} \sim N(\theta, Var(\hat{\theta}))$$

بر اساس احتمال

در برآورد θ با استفاده از $\hat{\theta}$ می خواهیم حجم نمونه را به گونه ای تعیین کنیم که قدر مطلق خطای مطلق با احتمال $1-\alpha$ کمتر از e باشد. یعنی:

$$P(|\hat{\theta}-\theta| < e) = 1-\alpha$$

می نویسیم:

$$P(|\hat{\theta}-\theta| < e) = P(P(-e < \hat{\theta}-\theta < e)) ==$$
$$= P(-\frac{e}{\sqrt{Var(\hat{\theta})}} < \frac{\theta-\hat{\theta}}{Var(\hat{\theta})} < \frac{e}{\sqrt{Var(\hat{\theta})}})$$

به سادگی می توان دریافت که:

$$Z_{\frac{\alpha}{2}} = -Z_{1-\frac{\alpha}{2}} = -\frac{e}{\sqrt{Var(\hat{\theta})}} \qquad Z_{1-\frac{\alpha}{2}} = -\frac{e}{\sqrt{Var(\hat{\theta})}}$$

در نهایت:

$$Z_{1-\frac{\alpha}{2}} = -\frac{e}{\sqrt{Var(\hat{\theta})}} \xrightarrow{Power\ 2} Z^2_{1-\frac{\alpha}{2}} = \frac{e^2}{Var(\hat{\theta})} \Rightarrow Var(\hat{\theta}) = \frac{e^2}{Z^2_{1-\frac{\alpha}{2}}}$$

مثال

برآورد $\bar{y_n}$ در روش SRS بدون جایگزاری:

$$Var(\bar{y_n}) = (1-\frac{n}{N})\frac{S^2}{n} = \frac{e^2}{Z^2_{1-\frac{\alpha}{2}}} = V_0$$
$$\frac{S^2}{n} - \frac{S^2}{N} = v_0 \Rightarrow \frac{s^2}{n} = V_0 + \frac{s^2}{N}$$
$$\frac{n}{S^2} = \frac{1}{V_0+\frac{S^2}{N}} \rightarrow n = \frac{S^2}{V_0+\frac{S^2}{N}}$$

از آنجایی که اطلاعتی نه از S^2 و نه از s^2، نیاز داریم تا یک نمونه ی مقدماتی به حجم m بگیریم و s_0^2 را استفاده کنیم.

$$s_0^2 = \frac{\sum_{i=1}^{m}(y_i-\bar{y_m})}{m-1}$$

برآورد P بدون جایگزاری:

$$Var(p) = \frac{e^2}{Z^2_{1-\frac{\alpha}{2}}} = V_0$$
$$\frac{N-n}{N_1}.\frac{P(1-P)}{n} = V_0$$
$$\frac{N-n}{n} = V_0\frac{N-1}{P(1-P)}$$
$$n = \frac{P(1-P)N}{(N-1)V_0 + P(1-P)}$$

می خواهیم برآورگری برای $P(1-P)$ پیدا کنیم.

$$E[p_0(1-p_0)] = E(p_0) - E(p_0^2) = P - \frac{N-m}{N-1}.\frac{P(1-P)}{m} - P^2 =$$
$$= P(1-P)[1 - \frac{N-m}{m(N-1)}] = P(1-P)\frac{N(m-1)}{m(N-1)}$$

در نتیجه برآوردگر زیر ناریب است.

$$\frac{m(N-1)}{N(m-1)}p_0(1-p_0)$$

مثال:

در یک دانشکده به کمک نمونه ی مقدماتی ۱۰۰ تایی درصد دانش جویان شاغل ۲۵ درصد برآورد شده است. قصد داریم با نمونه گیری SRS تعداد دانشجویان شاغل را به گونه ای برآورد کنیم که قدر مطلق خطای مطلق به احتمال ۹۵ درصد کمتر از ۱۰ درصد باشد.تعداد کل دانشجویان ۱۲۰۰ است. چه تعداد نمونه لازم است؟

۱.۰.۰ قدر مطلق خطای نسبی

در برآورد θ با استفاده از $\hat{\theta}$ می خواهیم حجم نمونه را به گونه ای تعیین کنیم که قدر مطلق خطای نسبی با احتمال $1-\alpha$ کمتر از e باشد.

$$P(|\frac{\hat{\theta}-\theta}{\theta}|<e)=P(-e<\frac{\hat{\theta}-\theta}{\theta}<e)$$
$$P(-e<\frac{\hat{\theta}-\theta}{\theta}<e)=P(-e\theta<\hat{\theta}-\theta<e\theta)$$
$$P(-e\theta<\hat{\theta}-\theta<e\theta)=P(-\frac{e\theta}{\sqrt{Var(\hat{\theta})}}<\frac{\hat{\theta}-\theta}{\sqrt{Var(\hat{\theta})}}<\frac{e\theta}{\sqrt{Var(\hat{\theta})}})$$
$$\Rightarrow \frac{e\theta}{\sqrt{Var(\hat{\theta})}}=Z_{1-\frac{\alpha}{2}}$$
$$\frac{e^2\theta^2}{Var(\hat{\theta})}=Z^2_{1-\frac{\alpha}{2}}$$
$$Var(\hat{\theta})=\frac{e^2\theta^2}{Z_{1-\frac{\alpha}{2}}}$$

از آنجایی که θ یک مقدار مجهول است از $\hat{\theta}$ استفاده می کنیم.

بر اساس طول بازه ی اطمینان

دربرآورد θ با استفاده از $\hat{\theta}$ می خواهیم حجم نمونه را به گونه ای تعیین کنیم که طول فاصله اطمینان $1-\alpha$ درصدی برابر با $2L$ شود.

$$Z = \frac{\hat{\theta}-\theta}{\sqrt{Var(\hat{\theta})}}$$

$$P(a < Z < b) = 1-\alpha$$

$$P(-Z_{1-\frac{\alpha}{2}} < \frac{\hat{\theta}-\theta}{Var(\hat{\theta})} < Z_{1-\frac{\alpha}{2}}) = P(Z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\theta})} < \hat{\theta}-\theta < Z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\theta})} =$$

$$= P[\hat{\theta} - Z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\theta})} < \theta < \hat{\theta} + Z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\theta})}]$$

$$2Z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\theta})} = 2L$$

$$L = Z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\theta})}$$

$$L^2 = Z^2_{1-\frac{\alpha}{2}}Var(\hat{\theta})$$

$$Var(\hat{\theta}) = \frac{L^2}{Z^2_{1-\frac{\alpha}{2}}}$$

مباحث میان ترم دوم:

نمونه گیری با طبقه بندی

تعریف

در نمونه گیری طبقه بندی جامعه می بایست به طبقات همگن طبقه بندی شود، به گونه ای که بین طبقات اختلاف وجود داشته باشد.

روش نمونه گیری

به گونه ای است که از هر طبقه نمونه ای به روش SRS انتخاب می شود و انتخاب نمونه از هر طبقه مستقل از انتخاب نمونه از طبقه دیگر است.
جامعه:

$$\begin{pmatrix} y_{1\,1} \\ \vdots \\ y_{1\,N_1} \\ \hline y_{2\,1} \\ \vdots \\ y_{2\,N_2} \\ \hline \vdots \\ \hline y_{H\,1} \\ \vdots \\ y_{H\,N_H} \end{pmatrix} \longrightarrow \begin{cases} N_1 : y_{1\,1}, \ldots, y_{1\,N_1} : \bar{y}_{1.} \\ N_2 : y_{2\,1}, \ldots, y_{2\,N_2} : \bar{y}_{2.} \\ \vdots \\ N_H : y_{H\,1}, \ldots, y_{H\,N_H} : \bar{y}_{H.} \end{cases}$$

نمونه:

$$\begin{pmatrix} y_{1\,1} \\ \vdots \\ y_{1\,n_1} \\ \hline y_{2\,1} \\ \vdots \\ y_{2\,n_2} \\ \hline \vdots \\ \hline y_{H\,1} \\ \vdots \\ y_{H\,n_H} \end{pmatrix} \longrightarrow \begin{cases} n_1 : y_{1\,1}, \ldots, y_{1\,n_1} : \bar{y}_1 \\ n_2 : y_{2\,1}, \ldots, y_{2\,n_2} : \bar{y}_2 \\ \vdots \\ n_H : y_{H\,1}, \ldots, y_{H\,n_H} : \bar{y}_H \end{cases}$$

چند تعریف:

۱. میانگین جامعه:

$$\bar{y}_N{}^{[1]} = \frac{\sum_{i=1}^{H}\sum_{j=1}^{N_i} y_{ij}}{N}$$

۲. انحراف از معیار جامعه:

$$S^2 = \frac{\sum_{i=1}^{H}\sum_{j=1}^{N_i}(y_{ij} - \bar{y}_N)^2}{N-1}$$

۳. میاگین طبقه i ام:

$$\bar{y}_{i.} = \frac{\sum_{j=1}^{N_i} y_{ij}}{N_i}$$

۴. انحراف از معیار طبقه i ام:

$$S_i^2 = \frac{\sum_{j=1}^{N_i}(y_{ij} - \bar{y}_{i.})^2}{N_i - 1}$$

۵. میانگین نمونه ی طبقه i ام:

$$\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

۶. انجراف از معیار نمونه i ام:

$$s_i^2 = \frac{\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2}{n_i - 1}$$

اگر قرار دهیم که $W_i = \frac{N_i}{N}$ آنگاه می توان گفت که $\bar{y}_{str}$ یک برآوردگر خوب برای $\bar{y_N}$ است.

$$\bar{y}_{str} = \sum_{i=1}^{H} W_i\,\bar{y}_i$$

ثابت کنید که:

$$E(\bar{y}_{str}) = \bar{y}_N$$

برهان:

$$E(\bar{y}_{str}) = E[\sum_{i=1}^{H} W_i\,\bar{y}_i] = \sum_{i=1}^{H} W_i\,E(\bar{y}_i) = \sum_{i=1}^{H}\frac{N_i}{N}\bar{y}_{i.} = \frac{1}{N}\sum_{i=1}^{H} N_i\,\bar{y}_{i.} =$$

$$= \frac{\sum_{i=1}^{H}\cancel{N_i}\frac{1}{\cancel{N_i}}\sum_{j=1}^{N_i} y_{ij}}{N} = \frac{\sum_{i=1}^{H}\sum_{j=1}^{N_i} y_{ij}}{N} = \bar{y}_N$$

---

[1] $N = \sum_{i=1}^{H} N_i$

مقدار خطای برآورد را پیدا کنید.
از آنجایی که برآوردگر ما اریب است، مقدار خطای آن با واریانس آن برابر است:

$$Var(\bar{y}_{str}) = Var[\sum_{i=1}^{H} W_i\,\bar{y}_i] = \sum_{i=1}^{H} W_i^2\,Var\bar{y}_i + \underbrace{\sum^{H}\sum^{H}_{i\neq j} Cov(\bar{y}_i,\bar{y}_j)}_{indpendent \Rightarrow 0} =$$

$$= \sum_{i=1}^{H} W_i^2\,(1-\frac{n_i}{N_i})\frac{s_i^2}{n_i}\text{[۲]} = \sum_{i=1}^{H}\frac{N_i^2}{N^2}(1-\frac{n_i}{N_i})\frac{s_i^2}{n_i} = \frac{1}{N^2}\sum_{i=1}^{H} N_i(N_i-n_i)\frac{s_i^2}{n_i}\text{[۴]} =$$

$$= \sum_{i=1}^{H} W_i^2\,\frac{S_i^2}{n_i} - \frac{1}{N}\sum_{i=1}^{H} W_i\,S_i^2\text{[۳]}$$

تعیین حجم نمونه

نمونه گیری با تخصیص متناسب

$$n_i \propto N_i$$
$$n_i = kN_i$$
$$\sum_{i=1}^{H} n_i = k\sum_{i=1}^{H} N_i$$
$$n = kN \Rightarrow k = \frac{n}{N}$$

اگر در روش نمونه گیری طبقه بندی، حجم نمونه ای که از طبقات می گیریم متناسب با حجم طبقات باشد، داریم:

$$\frac{n_i}{N_i} = \frac{N}{N} \qquad n_i = n.W_i\text{[۷]}$$

[۲] فرمول ۱
[۳] فرمول ۲
[۴] فرمول ۳

به این روش نمونه گیری طبق بندی با تخصیص متناسب می گویند.

$$Var(\bar{y}_str) = \sum_{i=1}^{H} W_i^2(1-\frac{n_i}{N_i})\frac{S_i^2}{n_i} = \sum_{i=1}^{H} W_i.\frac{N_i}{N}(1-\frac{n_i}{N_i})\frac{S_i^2}{n_i} =$$

$$\overset{\frac{N_i}{N}=\frac{n_i}{n}}{=} \sum_{i=1}^{H} W_i.\frac{\cancel{n_i}}{n}(1-\frac{n_i}{N_i})\frac{S_i^2}{\cancel{n_i}} = (1-\frac{n}{N})\frac{\sum_{i=1}^{H} W_i S_i^2}{n}$$

فرض کنید که $S_i^2 = S_0^2$:

$$(1-\frac{n}{N})\frac{\sum_{i=1}^{H} W_i S_i^2}{n} = (1-\frac{n}{N})\frac{\sum_{i=1}^{H} W_i S_0^2}{n} = (1-\frac{n}{N})\frac{S_0^2\sum_{i=1}^{H} W_i}{n} =$$

$$= (1-\frac{n}{N})\frac{S_0^2}{n}$$

در حقیقت با فرض $S_i^2 = S_0^2$،طبقه ها را یکسان فرض کردیم و نمونه گیری ما همان SRS بدون جایگزاری است.

تمرین:

فرض کنید منطقه ای دارای ۶۴ شهر است. منطقه را به دو طبقه تقسیم کرده ایم. طبقه اول ۱۶ شهر و طبقه ی دوم ۴۸ شهردارد. هدف برآورد جمعیت منطقه بر اساس نمونه ای به حجم ۲۴ شهر است. به منظور از سه روش زیر استفاده شده است.

۱. SRS

۲. طبقه بندی با تخصیص متناسب

۳. طبقه بندی با انتخاب ۱۲ شهر از هر منطقه دقت سه روش را محاسبه کنید.

i	$\sum_{j=1}^{N_i} y_{ij}$	$\sum_{j=1}^{N_i} y_{ij}^2$
1	10070	7145450
2	9408	2141720

$$T_N = 19568 \qquad S_1^2 = 53843$$

$$S^2 = 52445 \qquad S_2^2 = 5581$$

(آ) روش اول

$$n_1 = \frac{24}{64}\times 16 = 6 \qquad n_2 = \frac{24}{64}\times 48 = 18$$

۷ $W_i = \frac{N_i}{N}$

$$Var(\hat{T}_N) = Var(N\bar{y}_s tr) =$$
$$= N^2 Var(\bar{y}_s tr) = N^2(1-\frac{n}{N})\frac{S^2}{n} = (64)^2 \times (1-\frac{24}{64})\frac{52445}{24} = (2365)^2$$

(ب) روش دوم:

$$Var(\hat{T}_N) = N^2(1-\frac{n}{N})\frac{\sum_{i=1}^{2} W_i S_i^2}{n} = (64)^2(1-\frac{24}{64})\frac{\frac{1}{4}(53843) + \frac{3}{4}(5581)}{24} = (1372)^2$$

(ج) روش سوم:

$$Var(\hat{T}_N) = N^2 Var(\bar{y}_s tr) = N^2 \sum_{i=1}^{2} W_i^2 (1 - \frac{n_i}{N_i})\frac{S_i^2}{n_i} =$$
$$= (64)^2\{(\frac{16}{64})^2(1-\frac{12}{16})\frac{53843}{12} + (\frac{48}{64})^2(1-\frac{12}{48})\frac{52445}{12}\} = (1024)^2$$

دقت ها:
روش سوم > روش دوم > روش اول

طبقه بندی با تخصیص بهینه

تابع هزینه:

$$C = C_0 + \sum_{i=1}^{H} C_i n_i$$

- C : هزینه کل
- C_0 : هزینه ی ثابت
- C_i : هزینه برای هر نمونه از طبقه i ام

می خواهیم n_i را به گونه ای تعیین کنیم، که با هزینه ی ثابت C ، واریانس برآورد کمینه شود. با استفاده از روش لاگرانژ داریم که:

$$L=\sum_{i=1}^{H}W_i^2\frac{S_i^2}{n_i}-\sum_{i=1}^{H}W_i\frac{S_i^2}{N}+\lambda(C_0+\sum_{i=1}^{H}C_in_i-C)$$
$$\frac{\partial L}{\partial n_1}=-W_1^2\frac{S_1^2}{n_1^2}+\lambda C_1=0\Rightarrow n_1=\frac{W_1S_1}{\sqrt{\lambda C_1}}$$
$$\frac{\partial L}{\partial n_2}=-W_2^2\frac{S_2^2}{n_2^2}+\lambda C_2=0\Rightarrow n_2=\frac{W_2S_2}{\sqrt{\lambda C_2}}$$
$$\vdots$$
$$\frac{\partial L}{\partial n_H}=-W_H^2\frac{S_H^2}{n_H^2}+\lambda C_H=0\Rightarrow n_H=\frac{W_HS_H}{\sqrt{\lambda C_H}}$$
$$\frac{\partial L}{\partial \lambda}=C_0+\sum_{i=1}^{H}C_in_i-C=0$$

پیدا کردن λ :

$$\sum_{i=1}^{H}n_i=\sum_{i=1}^{H}\frac{W_iS_i}{\sqrt{\lambda}\sqrt{C_i}}$$
$$n=\frac{1}{\sqrt{\lambda}\sum_{i=1}^{H}\frac{W_iS_i}{\sqrt{C_i}}}$$
$$\sqrt{\lambda}=\frac{1}{n}\sum_{i=1}^{H}\frac{W_iS_i}{\sqrt{C_i}}$$
$$\Rightarrow n_i=\frac{W_iS_i}{\frac{1}{n}\sum_{i=1}^{H}\frac{W_iS_i}{\sqrt{C_i}}\sqrt{C_i}}=\frac{\frac{W_iS_i}{\sqrt{C_i}}}{\sum_{i=1}^{H}\frac{W_iS_i}{\sqrt{C_i}}}\times n$$

مقدار کمینه واریانس بر اساس روش طبقه بندی با تخصیص بهینه:

$$Var(\bar{y}_str)=\sum_{i=1}^{H}W_i^2\frac{S_i^2}{n_i}-\frac{\sum_{i=1}^{H}W_i^2S_i^2}{N}=\sum_{i=1}^{N}\frac{W_i^{\cancel{2}}S_i^{\cancel{2}}}{\frac{\frac{\cancel{W_iS_i}}{\sqrt{C_i}}}{\sum_{i=1}^{H}\frac{W_iS_i}{\sqrt{C_i}}}\times n}-\frac{\sum_{i=1}^{H}W_iS_i^2}{N}=$$
$$=\frac{1}{n}(\sum_{i=1}^{H}\frac{W_iS_i}{\sqrt{C_i}})\sum_{i=1}^{H}W_iS_i\sqrt{C_i}$$

پیدا کردن مقدار n مناسب:

$$C = C_0+\sum_{i=1}^{H} \frac{\frac{W_iS_i}{\sqrt{C_i}}}{\sum_{i=1}^{H}\frac{W_iS_i}{\sqrt{C_i}}}\times n\times C_i \Rightarrow C-C_0 = n\times\frac{1}{\sum_{i=1}^{H}\frac{W_iS_i}{\sqrt{C-i}}}\sum_{i=1}^{H} W_iS_i\sqrt{C_i} =$$
$$= n\times\frac{\sum_{i=1}^{H} W_iS_i\sqrt{C_i}}{\sum_{i=1}^{H}\frac{W_iS_i}{\sqrt{C-i}}} \Rightarrow n = (C-C_0)\frac{\sum_{i=1}^{H}\frac{W_iS_i}{\sqrt{C_i}}}{\sum_{i=1}^{H} W_iS_i\sqrt{C_i}}$$

تمرین:
n_i ها را به گونه ای تعیین کنید که با واریانس ثابت، میانگین هزینه ثابت شود.

تمرین

در یک نمونه گیری با طبقه بندی، تابع هزینه بصورت زیر است.

$$C = \sum_{i=1}^{2} C_i\, n_i$$

- W_i: وزن طبقه ی i ام
- S_i: انحراف معیار طبقه i ام

H	W_i	S_i	C_i
۱	۰.۴	۱۰	۴۰۰
۲	۰.۶	۲۰	۹۰۰

اگر n حجم نمونه باشد، برای $Var(\bar{y}_{str})$ مشخص وقتی مقدار C کمینه است، مقدار $\frac{n_1}{n}$ چقدر است؟

$$\frac{n_1}{n} = \frac{\frac{W_1S_1}{\sqrt{C_1}}}{\frac{W_1S_1}{\sqrt{C_1}}+\frac{W_2S_2}{\sqrt{C_2}}} = \frac{\frac{0.4\times10}{20}}{\frac{0.4\times10}{20}+\frac{0.6\times26}{30}} = \frac{1}{3}$$

تخصیص نیمن:

$$\forall i\in\{1,\dots,H\}:\ C_i = C$$
$$C^* = C_0+\sum_{i=1}^{H} n_iC = C_0+C\sum_{i=1}^{H} n_i = C_0+Cn$$
$$n_i = \frac{\frac{W_iS_i}{\cancel{\sqrt{C}}}}{\frac{\sum_{i=1}^{} W_iS_i}{\cancel{\sqrt{C}}}}\times n = \frac{W_iS_i}{\sum_{i=1}^{H} W_iS_i}\times n$$

$$Var(\bar{y}_{str}) = \sum_{i=1}^{H} W_i^2(1-\frac{n_i}{N_i})\frac{S_i^2}{n_i} = \sum_{i=1}^{H}\frac{W_i^2S_i^2}{n_i} - \sum_{i=1}^{H} W_i.\frac{\cancel{N_i}}{N}.\frac{\cancel{n_i}}{\cancel{N_i}}.\frac{S_i^2}{\cancel{n_i}} =$$

$$= \sum_{i=1}^{H}\frac{W_i^2S_i^2}{n_i} - \frac{1}{N}\sum_{i=1}^{H} W_iS_i^2 = \sum_{i=1}^{H} W_i^{\cancel{2}}\frac{S_i^2}{\frac{W_iS_i}{\sum_{i=1}^{H} W_IS_i}\times n} - \frac{1}{N}\sum_{i=1}^{H} W_iS_i^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{H} W_iS_i(\sum_{i=1}^{H} W_iS_i) - \frac{1}{N}\sum_{i=1}^{H} W_iS_i^2 = \frac{1}{n}(\sum_{i=1}^{H} W_iS_i)^2 - \frac{1}{N}\sum_{i=1}^{H} W_iS_i^2$$

فرض کنید حجم طبقات بزرگ باشد.

$$N_i \gg$$

آنگاه:

$$V_{ran} \geq V_{prop} \geq V_{opt}$$