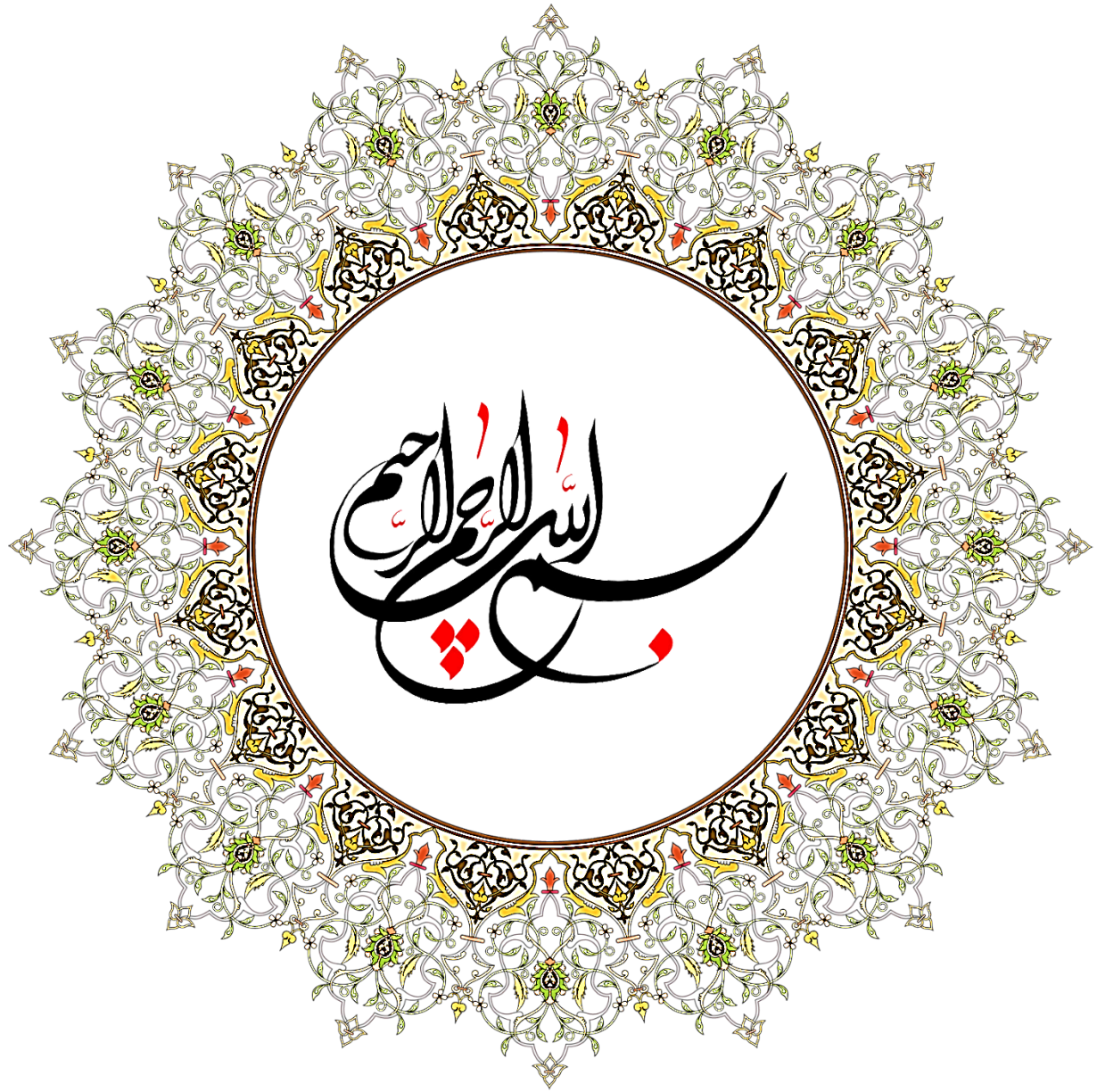

بسم الله الرحمن الرحیم

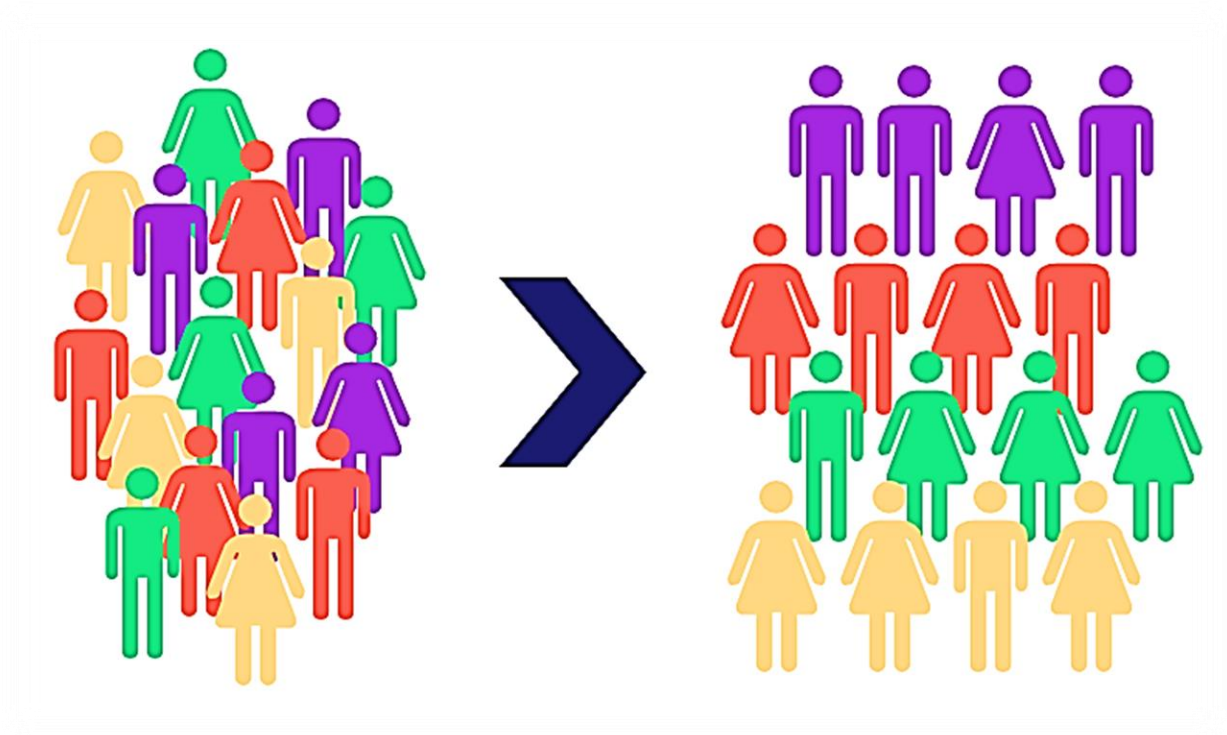
دانشگاه شهید چمران اهواز

روشهای نمونه گیری

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

درس پنجم

نمونه گیری تصادفی با طبقه بندی
Stratified random sampling



- هرگاه جامعه خیلی نا همگن باشد، دیگر با افزایش حجم نمونه و با وجود محدودیت های اقتصادی نمی توانیم با استفاده از نمونه گیری تصادفی ساده به برآوردهای دقیقی برسیم.
- نوسانات شدید یک صفت در جامعه موجب می شود تا به طبقه ها و گروههای مختلف جامعه نیز توجه شود.
- برای حجم نمونه های برابر خطای برآورد در طبقه بندی از تصادفی ساده کمتر است.
- هزینه یک پژوهش با طبقه بندی به مناطق جغرافیایی و.... کاهش پیدا می کند.
- نمونه گیری طبقه ای باعث شناسایی بهتر زیر جامعه ها می شود.
- طبقه بندی در جامعه بر حسب ویژگی های متفاوت آن جامعه است.
- در برخی از موارد هم طبقه ها به صورت طبیعی در جامعه تشکیل شده اند.

- واریانس درون طبقه ها باید کوچک باشد.(طبقه ها از درون باید همگن باشند).
- طبقه ها در بیرون باید نا همگن باشند و واریانس بین طبقه ها زیاد باشد.
- واریانس کل از مجموع واریانس مستقل طبقه ها محاسبه می شود پس بهتر است واریانس درون طبقه ها کم باشد.
- جنسیت، نژاد، شغل، میزان درآمد، وضعیت تاهل و..... مواردی هستند که جامعه بر حسب آنها طبقه بندی می شود.

جامعه به حجم N را به L طبقه به حجم های $N_1, N_2, \ldots, N_L$ تقسیم می کنیم. این طبقه ها متداخل نیستند و هر واحد تنها متعلق به یک طبقه است به گونه ای که :

$$N = \sum_{h=1}^{L} N_h$$

از این طبقه ها با روشهای نمونه گیری ساده و یا سیستماتیک و.... نمونه هایی به حجم $n_1, n_2, \ldots, n_L$ به گونه ای انتخاب می شوند که :

$$n = \sum_{h=1}^{L} n_h$$

حال نمونه ای به حجم n داریم، که چون به صورت تصادفی از طبقه ها انتخاب شده اند، به آن نمونه تصادفی با طبقه بندی گفته می شود.

معرفی نمادها و برخی از تعریف ها

- میانگین جامعه $\bar{Y}_N$
- تعداد کل واحدها در طبقه h ام N_h
- تعداد واحدهای نمونه از طبقه h ام n_h
- مقدار صفت واحد iام در طبقه hام Y_{hi}
- وزن طبقه h ام $W_h = \frac{N_h}{N}$
- کسر نمونه گیری برای طبقه h ام $f_h = \frac{n_h}{N_h}$
- میانگین طبقه h ام $\bar{y}_h = \frac{1}{N_h}\sum_{1}^{N_h} Y_{hi}$
- میانگین نمونه طبقه h ام $\bar{Y}_h = \frac{1}{n_h}\sum_{1}^{n_h} Y_{hi}$

تغییرات طبقه h ام $S^2 = \frac{1}{N_h - 1}\sum_{1}^{N_h}(Y_{hi} - \bar{y}_h)^2$

تغییرات نمونه h ام $s^2 = \frac{1}{n_h - 1}\sum_{1}^{n_h}(Y_{hi} - \bar{Y}_h)^2$

- میانگین موزون میانگین های نمونه ای طبقه ها را میانگین نمونه با طبقه بندی می نامیم و آن را به صورت زیر نمایش می دهیم:

$$\overline{Y}_{st} = \frac{1}{N}\sum_{h=1}^{L} N_h \overline{Y}_h = \sum_{h=1}^{L} W_h \overline{Y}_h$$

- این نمونه با نمونه تصادفی به حجم n از کل جامعه متفاوت است. $\overline{Y}_{st} \neq \overline{Y}_n$

قضیه ۱ : اگر در طبقه h ، میانگین نمونه برآورد کننده ناریب میانگین طبقه باشد، آنگاه میانگین موزون طبقه ها نیز برآورد کننده ناریب میانگین کل جامعه خواهد بود.

$$E(\overline{Y}_{st}) = \overline{Y}_N$$

$$\overline{Y}_N = \overline{Y}_{st}$$

- اگر T_N مجموع واحدهای جامعه باشد، آنگاه $N\overline{Y}_{st}$ برآورد کننده ناریب برای T_N است.

- قضیه ۲ : در نمونه گیری تصادفی با طبقه بندی واریانس میانگین موزون طبقه ها به صورت زیر است:

$$V(\bar{Y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \left(\frac{N_h - n_h}{N_h}\right) \frac{S_h^2}{n_h}$$

- در صورتی که نسبت حجم نمونه به حجم طبقه ها کوچک باشد واریانس به صورت زیر تعریف می شود:

$$V(\bar{Y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \frac{S_h^2}{n_h}$$

- واریانس مجموع واحدها نیز از رابطه زیر بدست می آید:

$$V(T_N) = \sum_{h=1}^{L} N_h^2 \left(\frac{N_h - n_h}{N_h}\right) \frac{S_h^2}{n_h}$$

- در همه روابط بالا برای برآورد واریانس از برآوردگر تغییرات جامعه استفاده می شود:

$$s_h^2 = \frac{1}{n_h - 1} \sum_{1}^{n_h} (Y_{hi} - \bar{Y}_h)^2$$

حدود اطمینان میانگین و مجموع واحدها در نمونه گیری با طبقه بندی

- اگر چهار طبقه به حجم های ۱۰، ۸، ۹ و ۱۲ داشته باشیم و از آنها به ترتیب ۳،۲،٤ و ۵ واحد به طور تصادفی استخراج کنیم، تعداد نمونه های ممکن چنین محاسبه می شود:

$$\binom{۱۰}{۴}\binom{۸}{۲}\binom{۹}{۳}\binom{۱۲}{۵} = ۳۹۱۱۸۴۶۴$$

- پس می توان امیدوار باشیم که توزیع میانگین موزون و مجموع واحدهای جامعه نرمال است و حدود آنها در سطح اطمینان $1-\alpha$ به صورت زیر تعریف می شود:

$$\bar{Y}_{st} - z\,\sigma(\bar{Y}_{st}) < \bar{Y}_{N} < \bar{Y}_{st} + z\,\sigma(\bar{Y}_{st})$$

$$N\bar{Y}_{st} - zN\,\sigma(\bar{Y}_{st}) < T_{N} < N\bar{Y}_{st} + zN\,\sigma(\bar{Y}_{st})$$

مثال ۱–۵ ، ۲–۵ و ۳–۵

حجم نمونه برای برآورد میانگین جامعه و مجموع واحدهای جامعه

- می توانیم برای حاشیه خطای دلخواه حجم نمونه را برای نمونه گیری طبقه ای برآورد کنیم. فرض کنیم که اندازه حاشیه کناری مدنظر پژوهشگر در سطح α برابر B باشد، آنگاه حجم نمونه برای برآورد میانگین جامعه به صورت زیر خواهد بود:

قابل برآورد

قابل محاسبه

$$n = \frac{\sum_{h=1}^{L} \frac{N_h^2 \sigma_h^2}{w_h}}{\frac{N^2 B^2}{z^2} - \sum_{h=1}^{L} N_h \sigma_h^2}$$

فرض کنیم که اندازه حاشیه کناری مدنظر پژوهشگر در سطح α برابر d باشد، آنگاه حجم نمونه برای برآورد مجموع واحدهای جامعه به صورت زیر خواهد بود:

$$n = \frac{\sum_{h=1}^{L} \frac{N_h^2 \sigma_h^2}{w_h}}{\frac{B^2}{z^2} - \sum_{h=1}^{L} N_h \sigma_h^2}$$

برای محاسبه واریانس طبقه های جامعه عموما دو راه حل پیشنهاد می شود:

- استفاده از پیشینه پژوهش و تجربه پژوهشگر مقادیر واریانس طبقه ها برآورد می شود.
- در برخی موارد نمونه گیری مقدماتی انجام می شود.
- دامنه تغییرات در هر طبقه را محاسبه کرده و آن را به ٦ تقسیم می کنند.
- امروزه نرم افزارهایی برای برآورد حجم نمونه ها در بازار موجود است.

مثال ٥– ٥

مثال ٦–٥

تعیین حجم نمونه (محاسبه وزن طبقه ها)

- در نمونه گیری هدف فراهم ساختن برآورد کننده هایی با واریانس کم و با کمترین هزینه ممکن است.
- در روش نمونه گیری با طبقه بندی هدف تعیین حجم نمونه در تک تک طبقه ها است.
- همواره سه عامل متفاوت در تعیین بهترین نمونه موثر هستند:

۱- تعداد کل عناصر در هر طبقه

۲- تغییر پذیری مشاهدات در هر طبقه

۳- هزینه هر مشاهده در هر طبقه

- فرض کنیم هزینه نمونه گیری برای هر واحد از هر طبقه برابر Ch باشد، آنگاه می توان نشان داد که حجم نمونه از هر طبقه (سهم هر طبقه در نمونه کل) از رابطه زیر محاسبه خواهد شد:

$$n_h = \frac{\frac{N_h \sigma_h}{\sqrt{C_h}}}{\sum_{h=1}^{L} \frac{N_h \sigma_h}{\sqrt{C_h}}} .n$$

- حجم طبقه و حجم نمونه رابطه مستقیم دارند. $n_h \propto N_h$
- حجم طبقه با واریانس جامعه رابطه مستقیم دارد. $n_h \propto \sigma_h$
- حجم طبقه با هزینه ها رابطه عکس دارد. $n_h \propto \frac{1}{\sqrt{C_h}}$

• برای محاسبه حجم طبقه ها چهار رویکرد رایج را معرفی می کنیم :

۱- برآورد حجم طبقه ها براساس حاشیه خطا(انتخاب بهینه برای حاشیه خطا):

ابتدا وزن طبقه را با استفاده از رابطه سهمیه هر طبقه محاسبه می کنیم:

$$n_h = \frac{\frac{N_h \sigma_h}{\sqrt{C_h}}}{\sum_{h=1}^{L} \frac{N_h \sigma_h}{\sqrt{C_h}}} . n \Rightarrow \frac{n_h}{n} = W_h = \frac{\frac{N_h \sigma_h}{\sqrt{C_h}}}{\sum_{h=1}^{L} \frac{N_h \sigma_h}{\sqrt{C_h}}}$$

مقدار وزن بدست آمده را در رابطه حجم نمونه بر حسب حاشیه خطا B قرار می دهیم:

$$n = \frac{(\frac{\sum_{k=1}^{L} N_k \sigma_k}{\sqrt{C_k}})(\sum_{h=1}^{L} N_h \sigma_h \sqrt{C_h})}{\frac{N^2 B^2}{z^2} + \sum_{k=1}^{L} N_h \sigma^2_h}$$

مثال ۷-۵

۲- انتساب بهینه نیمن : در این انتساب هزینه هر مشاهده برای همه طبقه ها یکسان است. بر اساس این انتساب رابطه n به صورت زیر خواهد بود:

$$n = \frac{(\sum_{h=1}^{L} N_h \sigma_h)^2}{\frac{N^2 B^2}{z^2} + \sum_{h=1}^{L} N_h \sigma^2_h}$$

و به همین ترتیب سهم هر طبقه نیز مشخص می شود:

$$n_h = \frac{N_h \sigma_h}{\sum_{h=1}^{L} N_h \sigma_h} . n$$

مثال ۸-۵ و ۹-۵

۳- هر گاه در روش نمونه گیری با طبقه بندی، حجم نمونه های طبقه ها با حجم طبقه ها متناسب باشد، به این نمونه گیری تخصیص متناسب می گوییم:

$$n_h = \frac{N_h}{N} n$$

اگر تخصیص به صورت متناسب باشد، برآورد حجم نمونه به صورت زیر است :

$$n = \frac{\sum_{h=1}^{L} N_h \sigma^2_h}{\frac{NB^2}{z^2} + \frac{1}{N} \sum_{h=1}^{L} N_h \sigma^2_h}$$

پس می توان سهم هر طبقه را به صورت زیر معین کرد:

$$n_h = \frac{N_h}{N} n$$

مثال ۱۰-۵

٤- پروژه دارای هزینه ثابت و مشخصی است، و حجم نمونه را چنان انتخاب می کنیم که واریانس نمونه گیری به کمترین مقدار برسد (رویکرد هزینه ثابت) :

$$n = \frac{C\left(\sum_{h=1}^{L} \frac{N_h \sigma_h}{\sqrt{C_h}}\right)}{\sum_{h=1}^{L} N_h \sigma_h \sqrt{C_h}}$$

C هزینه ثابت پروژه است.

$$n_h = \frac{\frac{N_h \sigma_h}{\sqrt{C_h}}}{\sum_{h=1}^{L} \frac{N_h \sigma_h}{\sqrt{C_h}}} . n$$

تکالیف

تمرین های منتخب فصل پنجم کتاب آمارگیری نمونه ای