УВОД У СТАТИСТИКУ час 9

26. април ’17.

ЛГЛАВАШ

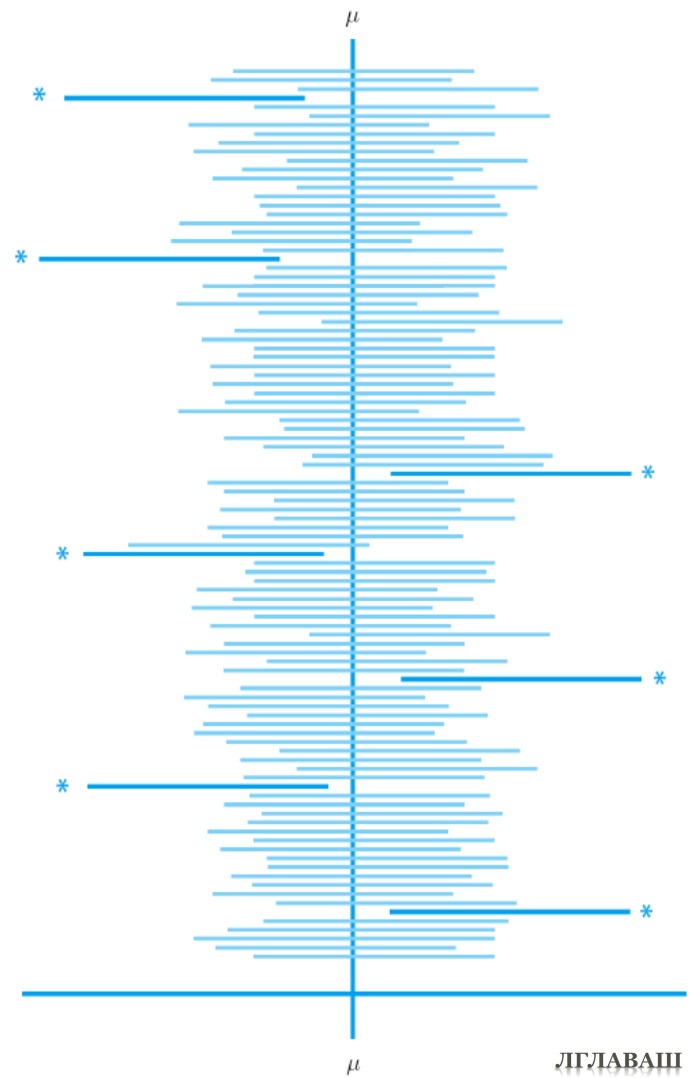
Интервално оцењивање

Интерпретација нивоа поверења

Нека је μ непознат параметар расподеле обележја X за који се, на основу простог случајног узорка обима n, формира 95% интервал поверења.

Исправна интерпретација нивоа поверења јесте: ако би се за посматрано обележје велики број пута узео узорак обима n и сваки пут одредио реализовани интервал поверења, тада би приближно 95% тих интервала садржало непознати параметар μ (the long-run relative frequency interpretation of probability).

На слици десно, вертикална права сече одговарајућу осу у тачној (али непознатој) вредности параметра μ. Хоризонталне дужи одговарају реализованим 95% интервалима поверења за 100 узорака. У приближно 95% случајева (тј. за тачно њих 93) реализовани интервал садржи тачну вредност μ, а у осталих приближно 5% случајева (тј. за тачно њих 7, који су означени звездицом) не садржи.



Пример 1

На националном нивоу испитиван је став јавног мњења о увођењу стандарда за аутомобиле са циљем повећања енергетске ефикасности горива. У те сврхе спроведена је анкета на простом случајном узорку обима 1012 одраслих становника који поседују возачку дозволу. Њих 810 изјаснило се у корист увођења оваквих стандарда.

Одредити 95% интервал поверења за удео (пропорцију) возача који подржавају увођење стандарда.

Решење:

Нека је p (непозната) вероватноћа да (произвољан) возач подржава увођење стандарда. Треба, уствари, одредити 95% интервал поверења за p.

Тражени интервал поверења је облика $[p_1, p_2]$, где су p_1 и p_2 решења квадратне једначине:

$$\left(n + z_\beta^2\right)p^2 - \left(2n\bar{x}_n + z_\beta^2\right)p + n\bar{x}_n^2 = 0,$$

а $z_\beta \in \mathbb{R}$, тако да је $P\left\{|Z| \leq z_\beta\right\} = \beta = 0.95$, $Z \in N(0,1)$.

Прво се добија: $z_\beta = 1.96$; затим, реализована вредност узорачке средине: $\bar{x}_n = \frac{405}{506}$.

Уврсте се конкретне вредности n, z_β, $\bar{x}_n$ у кв. једначину, која постаје:

$$1015.842p^2 - 1623.842p + 648.3202 = 0,$$

и чија су решења: $p_1 \approx 0.775$ и $p_2 \approx 0.824$.

Према томе (апроксимативни) 95% интервал поверења за непознати параметар p биномне расподеле је:

$$[0.775, 0.824].$$

△

Ако се изврши додатна апроксимација, која се састоји у замени параметра p у изразу за $D\bar{X}_n$ његовом тачкастом оценом $\hat{p} = \bar{X}_n$ добија се следећи 95% интервал поверења:

$$[0.776, 0.825].$$

Пример 2

Планирано је истраживање јавног здравља на широј територији једног великог града ради процене удела деце, узраста 0-14, на којима није спроведена имунизација (вакцинација) против дечје парализе. Организатори овог пројекта желели би да постигну да узорачки удео неадекватно имунизоване деце са вероватноћом бар 98% одступа, по апсолутној вредности, од тачног удела p за највише 0.05. Колики би требало да буде обим узорка?

Решење:

- Може да се искористи формула

$$\frac{z_\beta^2}{4d^2},$$

где је $\beta = 0.98$, $d = 0.05$, па да се за обим узорка n узме најмањи цео број већи (или једнак) од вредности добијене заменом конкретних β и d у формули.

Тако се добија: $n = 543$.

Пример 2 (наставак)

- Ако постоји оправдан разлог због кога се верује да је p обавезно мање од неког броја r_1, $r_1 < 0.5$, односно веће од неког броја r_2, $r_2 > 0.5$, фактор $\hat{p}(1-\hat{p})$ у формули
$$n = \hat{p}(1-\hat{p})\frac{z_\beta^2}{d^2}$$
за израчунавање обима узорка, може да се замени са $r_1(1-r_1)$, односно $r_2(1-r_2)$, што има утицај на (понекад и значајно) смањење обима узорка потребног за оцењивање p са задатом тачношћу.

  Ако нпр. претходна истраживања указују да је не више од 20% деце посматраног узраста неадекватно имунизовано добија се: $n = 348$. (редукција обима узорка за око 36%)
- Ако постоји (нека, може бити и „груба“) иницијална оцена $\hat{p}$ на основу прелиминарног узорка она се замењује у горњој формули за n. $\triangle$

Пример 3

Нека обележје X има $\mathcal{P}(\lambda)$, где је $\lambda > 0$.

Одредити $100 \cdot \beta\%$ интервал поверења за непознати параметар λ на основу простог случајног узорка $(X_1, X_2, \dots, X_n)$ обима n из обележја X.

Решење:

Познато је да је узорачка средина $\bar{X}_n$ непристрасна непознатог параметра λ. Такође, на основу важења адитивног закона, зна се да $n\bar{X}_n \in \mathcal{P}(n\lambda)$; стога се лако одређује $E(n\bar{X}_n) = n\lambda = D(n\bar{X}_n)$.

На основу важења централне граничне теореме:

$$\frac{\bar{X}_n - \lambda}{\sqrt{\lambda/n}} \xrightarrow{D} Z \in N(0,1),$$

и, одатле, коришћењем додатне апроксимације:

$$\frac{\bar{X}_n - \lambda}{\sqrt{\hat{\lambda}/n}} \xrightarrow{D} Z \in N(0,1),$$

при $n \to +\infty$, где је $\hat{\lambda} = \bar{X}_n$.

Тако се добија апроксимативни $100 \cdot \beta\%$ интервал поверења за непознати параметар λ:

$$\left[\bar{X}_n - z_\beta\sqrt{\bar{X}_n/n}, \bar{X}_n + z_\beta\sqrt{\bar{X}_n/n}\right]. \triangle$$

само за n велико!

ЛГЛАВАШ

Пример 4

Претпостави се да када се са тачке A трансмитује сигнал који има вредност μ, вредност сигнала примљена у тачки B има нормалну расподелу са мат. очекивањем μ и дисперзијом 4. Заправо, ако је послато μ примљено је $\mu+\epsilon$, где ϵ, које представља шум, има $N(0,4)$ расподелу. Примљен је следећи прост случајан узорак узастопних вредности сигнала:

$$7.45, 11.52, 7.94, 9.63, 11.41, 10.04, 11.31, 7.33, 5.72.$$

Одредити двострани 95% интервал поверења за μ.

Решење:

Тражени интервал поверења за непознати параметар μ је облика:

$$\left[\bar{X}_n - \frac{z_\beta}{\sqrt{n}}\sigma, \bar{X}_n + \frac{z_\beta}{\sqrt{n}}\sigma\right].$$

$z_\beta = 1.96$; реализована вредност узорачке средине: $\bar{x}_n = 9.15$.

Уврсте се конкретне вредности n, z_β, $\bar{x}_n$, σ у горњу формулу. Према томе 95% интервал поверења за μ је:

$$[7.84, 10.46]. \triangle$$

Пример 5

За артикле из исте серије, које производи одређена фабрика, претпоставља се да имају међусобно независне „дужине живота“ (у сатима) са истом густином расподеле датом са:

$$f(x;\,\theta) = \frac{1}{\theta} e^{-x/\theta}, 0 < x < \theta,$$

где је $\theta > 0$ непознат параметар.

Ако је укупна дужина живота 10 случајно одабраних артикала из исте серије једнака 1740 одредити једнострани доњи 90% интервал поверења за мат. очекивање θ дужине живота артикла.

Решење:

Обележје X од интереса има $\varepsilon\left(\frac{1}{\theta}\right)$ расподелу. Нека је $\lambda \coloneqq 1/\theta$. Познато је да је реципрочна вредност узорачке средине оцена параметра λ. Такође, зна се да $\sum_{j=1}^{n} X_j \in \gamma(n, \lambda)$. На основу везе између гама расподеле и χ^2 расподеле следи:

$$2\lambda \sum_{j=1}^{n} X_j \in \chi^2_{2n}.$$

Пример 5 (наставак)

Једнострани доњи интервал поверења за θ са задатим нивоом поверења одређује се из:

$$P\{2\lambda n\bar{X}_n > \chi^2_{2n;0.90}\} = 0.90,$$

па се добија да је тражени интервал поверења:

$$\left[0, \frac{2n\bar{X}_n}{\chi^2_{2n;0.90}}\right].$$

Уврсте се конкретне вредности n, $\chi^2_{2n;0.90}$, $\bar{x}_n$ у горњу формулу. Према томе 90% интервал поверења за θ је:

$$[0, 122.48].$$

На сличан начин би се одређивали једнострани горњи 90% интервал поверења:

$$[279.68, +\infty),$$

односно двострани 90% интервал поверења:

$$[110.79, 320.71]. \triangle$$