برآوردگرهای استوار مکان و مقیاس

سید مهدی امیر جهانشاهی[۱] افشین آشفته[۲] حسینعلی نیرومند[۳]

چکیده

در این مقاله برآوردگرهای استوار مکان یعنی میانگین وینزوری و میانگین پیراسته، برآوردگرهای نوع R, L, M و برآوردگر نوع پیتمن را معرفی می‌کنیم. میانگینهای استوار وینزوری ر پیراسته را برای میانگین‌های همساز و هندسی تعمیم می‌دهیم. همچنین برآوردگرهای استوار مقیاس یعنی برآوردگر وینزوری و پیراسته ، نوع M ، دامنه میان چارکی ، میانگین تفاضل جینی، MAD و S_n, Q_n را معرفی نموده و در ادامه آزمونهای تی تک نمونه‌ای استوار را معرفی می‌نماییم و با استفاده از شبیه سازی مونت کارلو و رویه‌های نرم‌افزار SAS به معرفی برنامه‌های لازم جهت محاسبه موارد فوق‌الذکر می‌پردازیم.

واژه های کلیدی : میانگینهای پیراسته و وینزوری، برآوردگرهای استوار.

۱. مقدمه

یک روش آماری را استوار می‌گوییم اگر به طور معقولی حتی زمانی که مفروضات الگوی آماری کاملا برقرار نباشد خوب عمل کند. معمولا مشاهدات پرت هستند که باعث تخطی از مفروضات می‌شوند و طبیعتا روی نتایج بسیار تأثیر گذار خواهند بود. بنابراین لزوم تشخیص و آشنایی با نحوه صحیح برخورد با چنین مشاهداتی کاملا احساس می‌شود. بر این اساس اخیرا آماره‌های استوار بسیار گسترش یافته و کاربرد فراوانی در تحقیقات علمی پیدا کرده‌اند.

۲. برآوردگرهای استوار مکان

الف) میانگینهای نامتقارن پیراسته و وینزوری

وقتی که مقادیر پرت در داده ها وجود دارد میانگیهای پیراسته و وینزوری برآوردگرهای استوار جامعه هستند که نسبتا غیر حساس به مقادیر پرت می‌باشند. بنابراین، این دو روش برای تعدیل کردن (حذف نمودن) اثر مقادیر پرت در نمونه بکار می‌روند. میانگین پیراسته مرتبه (r,s) به صورت زیر محاسبه می‌شود.

$$\bar{x}_{T_{(r,s)}} = \frac{1}{n-r-s}\sum_{i=r+1}^{n-s} x_{(i)}$$

[۱] گروه آمار، دانشگاه بیرجند

[۲] گروه آمار، دانشگاه اصفهان

[۳] گروه آمار، دانشگاه فردوسی مشهد

بعد از اینکه کوچکترین r مشاهده و بزرگترین s مشاهده از نمونه حذف شوند؛میانگین پیراسته محاسبه می‌شود. به بیانی دیگر مشاهدات در هر دو انتها پیراسته می‌شوند. میانگین وینزوری مرتبه (r,s) به صورت زیر محاسبه می‌شود.

$$\overline{x}_{w(r,s)} = \frac{۱}{n}\left\{(r+۱)x_{(r+۱)} + \sum_{i=r+۲}^{n-s-۱} x_{(i)} + s x_{(n-s)}\right\}$$

میانگین وینزوری با جایگزینی کوچکترین r مشاهده بوسیله $(r+۱)$امین مشاهده و بزرگترین s مشاهده با $(s+۱)$ امین مشاهده محاسبه می‌شود. به بیانی دیگر مشاهدات در هر دو طرف وینزوری می‌شوند.

میانگینهای متقارن پیراسته و وینزوری

در حالتیکه $r = s$ باشد این میانگینها را متقارن می‌نامیم. بنابراین در این حالت میانگینهای وینزوری و پیراسته عبارت خواهند بود.

$$\overline{x}_{T_{(k,k)}} = \frac{۱}{n-۲k}\sum_{i=k+۱}^{n-k} x_{(i)}$$

$$\overline{x}_{w(k,k)} = \frac{۱}{n}\left\{(k+۱)x_{(k+۱)} + \sum_{i=k+۲}^{n-k-۱} x_{(i)} + (k+۱)x_{(n-k)}\right\}$$

برای یک توزیع متقارن میانگینهای متقارن وینزوری و پیراسته بر آوردگرهای نااریب میانگین جامعه هستند. اما حتی وقتی داده‌ها دارای توزیع نرمال باشند توزیع این دو میانگین نرمال نیست. حالتی را می‌توان درنظر گرفت که مقادیر پرت در یکطرف مشاهدات قرار بگیرند. بنابراین لزومی ندارد ما از هر دو طرف مشاهدات تعدادی حذف کنیم. بر این اساس میانگینهای یکطرفه وینزوری و پیراسته به شکل زیر تعریف می‌شوند.

میانگین پیراسته یکطرفه مرتبهr=
$$\overline{x}_{Tr} = \frac{۱}{n-r}\sum_{i=r+۱}^{n} x_{(i)}$$

میانگین پیراسته یکطرفه مرتبهs=
$$\overline{x}_{Ts} = \frac{۱}{n-s-۱}\sum_{i=۱}^{n-s-۱} x_{(i)}$$

میانگین وینزوری یکطرفه مرتبهr=
$$\overline{x}_{wr} = \frac{۱}{n}\left\{(r+۱)x_{(r+۱)} + \sum_{i=r+۲}^{n} x_{(i)}\right\}$$

میانگین وینزوری یکطرفه مرتبهs=
$$\overline{x}_{ws} = \frac{۱}{n}\left\{\sum_{i=۱}^{s-۱} x_{(i)} + (n-s-۱)x_{(s)}\right\}$$

معمولا بیان درصد مشاهده‌ای که می‌بایست حذف شود راحتتر می‌باشد. بر این اساس می‌توانیم میانگین آلفا پیراسته $(\alpha - trimmed)$ یا $\overline{x}_{T(\alpha:\alpha)}$ را تعریف کنیم که در واقع همان میانگین پیراسته متقارن ولی از مرتبه α می‌باشد و منجر به حذف 2α درصد از مشاهدات می‌شود. با این تعریف تعداد αn مشاهده از هر طرف مشاهدات حذف می‌شود. فرض کنید قسمت صحیح αn ، r باشد یا به عبارتی $\alpha n = r + f$ که $(۰ < f < ۱)$ حال از هر طرف r مشاهده را حذف نماییم و برای $x_{(r+۱)}$ و $x_{(n-r)}$ وزن $(۱-f)$ را در نظر می‌گیریم. بنابراین میانگین آلفا پیراسته به شکل زیر تعریف می‌شود.

$$\overline{x}_{T(\alpha:\alpha)} = \frac{۱}{n(۱-۲\alpha)}\Big\{(۱-f)x_{(r+۱)} + x_{(r+۲)} + \ldots + x_{(n-r-۱)} + (۱-f)x_{(n-r)}\Big\}$$

به طور مشابه می‌توانیم میانگین آلفا وینزوری یا $\overline{x}_{w(\alpha:\alpha)}$ را به صورت زیر تعریف کنیم.

$$\overline{x}_{w(\alpha:\alpha)} = \frac{۱}{n}\left\{rx_{(r+۱)} + x_{(r+۲)} + \ldots + x_{(n-r)} + rx_{(n-r+۱)}\right\}$$

با توجه به اینکه از هر طرف $r + f + ۱ - f - ۱ = r$ مشاهده حذف می‌شوند. واضح است که میانگین‌های $۰-trimmed$ یا $۰-winsorized$ هر دو برابر میانگین نمونه هستند و میانگین $۱/۲-winsorized$ شبیه میانه نمونه و میانگین $۱/۴-trimmed$ را mid mean می‌نامیم.

ب) برآوردگر نوع M (برآوردگر نوع درستنمایی ماکزیمم)

فرض کنید مشاهدات $x_۱, \ldots, x_n$ از توزیعی پیوسته با تابع توزیع $F(x-\mu)$ و تابع چگالی احتمال $f(x-\mu)$ باشند. برآورد μ با ماکزیمم نمودن تابع درستنمایی بدست می‌آید، و از آنجا که $\rho(x-\mu) = -\ln f(x-\mu)$ ، برآورد مذکور با مینیمم نمودن $\sum_{j=۱}^{n}\rho(x_j - T_n)$ بدست می‌آید که در آن $\rho(\cdot)$ تابعی حقیقی

مقدار و غیر ثابت است. اگر $\rho(t)=t^{۲}$ باشد برآوردگر μ ، (T_n) میانگین نمونه خواهد بود و اگر $\rho(t)=|t|$ باشد برآوردگر μ ،(T_n) میانه نمونه خواهد بود. اگر $\rho(\cdot)$ پیوسته و مشتق آن $\psi(t)$ باشد می‌توانیم μ را بوسیله (T_n) برآورد کنیم بطوریکه

$$\sum_{j=۱}^{n}\psi(x_j-T_n)=\circ .$$

معمولا $\rho(\cdot)$ را محدب در نظر می‌گیریم. (T_n) یکتا، پایا، سازگار و بطور مجانبی نرمال (هوبر ۱۹٦٤و ۱۹۸۱) و [٥] و [٦] می‌باشد.

به عنوان مثال

$$\rho(t)=\begin{cases}\dfrac{t^{۲}}{۲} & ;\quad |t|\le k\\[2ex] k|t|-\dfrac{k^{۲}}{۲}; & |t|>k\end{cases}$$

که یک انتخاب $k=۱/۵$ می‌تواند باشد. در حالتی که مدل اولیه شامل یک پارامتر مقیاس باشد (تابع توزیع به شکل $F((x-\mu)/\sigma)$ باشد) شکل تصحیح شده برآوردگر M برای μ از حل معادله زیر بدست می‌آید.

$$\sum_{j=۱}^{n}\psi\left(\frac{x_j-T_n}{S}\right)=\circ$$

که در آن S برآوردگر استوار σ است و مستقلا برآورد می‌شود و با حل همزمان آن با μ بدست می‌آید.

هامپل [۱] در سال ۱۹۷۲ شکل دیگر تابع $\rho(t)$ هوبر را مورد استفاده قرار داد. هوبر نشان داد این برآوردگر در بین برآوردگرهای پایا برای یک توزیع متقارن مینیماکس می‌باشد.

$$\psi(t)=\begin{cases}t & |t|\le k\\ k\,\mathrm{sgn}(t) & |t|>k\end{cases}$$

و S نیز به صورت زیر در نظر گرفته می‌شود.

$$S=Median\{|x_j-\tilde{x}|\}/۰/۶۷۴۵$$

که $\tilde{x}$ میانه نمونه است. هامپل [۸] در سال ۱۹۷٤ برآوردگر استوار نوع M دیگری را با تابع نزولی سه قسمتی زیر پیشنهاد می‌کند.

$$\psi(t)=\begin{cases}t & ,|t|\le a\\ a\,\mathrm{sgn}\,t & ,a<|t|\le b\\ a(c\,\mathrm{sgn}\,t-t)/(c-b) & ,b<|t|\le c\\ \circ & ,|t|>c\end{cases}$$

که در آن c= ۸/۵ و b= ۳/٤ و a= ۱/۷ می‌باشند.

تابع دیگری نیز توسط اندروز [۱] در سال ۱۹۷۲ به شکل زیر ارائه شد

$$\psi(t)=\begin{cases}\sin(\dfrac{t}{k}) & |t|<k\pi\\[2ex] \circ & |t|>k\pi\end{cases}$$

که دو انتخاب برای k, $k=۱/۳۳۹$ و $k=۲/۱$ می‌توانند باشند. تابع دیگری نیز توسط توکی در سال ۱۹۷٤ به شکل زیر ارائه شد.

$$\rho(t)=\begin{cases}(۱-(\dfrac{t}{c})^{۲})^{۲} & t<c\\[2ex] \circ & t\ge c\end{cases}$$

یک انتخاب مناسب برای c، $c=۴/۶۸۵$ می‌تواند باشد.

ج) برآوردگرهای نوع L

یک تعمیم بدیهی تقریبهای وینزوری و پیراسته برآوردگر نوعL است که ترکیب خطی از مقادیر آماره‌های ترتیبی نمونه می‌باشد.

$$\tilde{\mu}=\sum_i c_i x_{(i)}$$

مثالی از کاربرد آن توسط هوبر ۱۹۷۲ ارائه شد. وی $c_i=\circ$ برای تمام مشاهدات به غیر از تک مشاهده یا ۲ مشاهده میانی تعریف کرد که همان میانه نمونه می‌شود. در حالتی که $c_i=۱/n$,$(i=۱,\ldots,n)$ باشد $\tilde{\mu}$ با میانگین نمونه یکی خواهد شد.

دیوید و شو [۲] در سال ۱۹۷۸ برای حالتی که ۱۰ مشاهده داشته باشیم، برآوردگرهای نوع L زیر را پیشنهاد نمودند.

$$=(x_{(۲)}+۳x_{(۳)}+۵x_{(۴)}+۷x_{(۵)}+۷x_{(۶)}+۵x_{(۷)}+۳x_{(۸)}+x_{(۹)})/۳۲$$

$$=(x_{(۳)}+۳x_{(۴)}+۵x_{(۵)}+۵x_{(۶)}+۳x_{(۷)}+x_{(۸)})/۱۸$$

گست ویرت و کوهن [۹] در سال ۱۹۷۰ برآوردگر نوع L زیر را پیشنهاد کردند

$$T_n = \gamma(x_{([np]+1)} + x_{(n-[np])}) - (1-2\gamma)\tilde{x}$$

که در آن ($0 < \gamma < 1$, $0 < p < 1$) می‌باشند. T_n ترکیب موزون صدکهای p ام بالا و پایین نمونه با وزن γ و میانه نمونه با وزن $1-2\gamma$ می‌باشد. یک حالت خاص T_n توسط گست ویرت در سال ۱۹٦٦ به شکل زیر ارائه شد

$$T_n = 0.3x_{([\frac{n}{3}+1])} + 0.4\tilde{x} + 0.3x_{(n-[\frac{n}{3}])}.$$

شکل دیگری نیز به نام trimean به شکل زیر ارائه شده که در آن h_1, h_2 چارکهای نمونه می‌باشند.

$$T_n = \frac{1}{4}(h_1 + 2\tilde{x} + h_2)$$

د) برآوردگر نوع R

هاجز و لهمن [۷] در سال ۱۹٦۳ اشاره کردند که برآوردگر μ می‌تواند از آزمون رتبه‌ای ویلکاکسن بدست آید. اثبات می‌شود که این برآوردگر استوار است. بر این اساس هاجز و لهمن برآوردگر خود را ارائه نمودند. که این برآوردگر میانه مجموعه $n(n+1)/2$ میانگین‌های جفتی $(x_j + x_l)/2$, $(l = 1,\dots,n;\quad j = 1,\dots,n;\quad j \neq l)$. می‌باشد. شکل ساده‌تر آن نیز ارائه شد که بر اساس میانگینهای متقارن مکانی مقادیر آماره‌های مرتب شده ($(n+1)/2$ مشاهده می‌باشند) محاسبه می‌شود. بیکل و هاجز برآوردگر نوع R مورد نظر خویش را با جایگزینی مشاهدات نمونه ($x_1,\dots,x_n$) با $[x_{(1)} + x_{(n)}]/2, [x_{(2)} + x_{(n-1)}]/2, \dots$ و محاسبه میانه نمونه معرفی کردند.

هـ) برآوردگر نوع پیتمن

جان در سال ۱۹۷۹ برآوردگر نوع p را به شکل زیر پیشنهاد کرد[٤].

$$T_n = \frac{\int \mu \prod_{i=1}^{n} \gamma(x_i - \mu)d\mu}{\int \prod_{i=1}^{n} \gamma(x_i - \mu)d\mu}$$

$$\gamma(x) = \frac{1}{1+x^2} + \frac{a+bx^2}{(1+x^2)^3}.$$

۳. تعمیم میانگینهای استوار

با توجه به مطالبی که در مورد برآوردگرهای استوار اشاره شد به تعمیم آنها برای میانگین همساز (هارمونیک) و هندسی می‌پردازیم. میانگین همساز پیراسته متقارن را به شکل زیر تعریف می‌کنیم

$$\bar{x}_{TH(k,k)} = \frac{1}{\frac{1}{n-2k}\sum_{i=k+1}^{n-k}\frac{1}{x_{(i)}}}.$$

همچنین میانگین همساز وینزوری متقارن را به شکل

$$\bar{x}_{WH(k,k)} = \frac{1}{\frac{1}{n}\left[\frac{1}{k+1}\cdot\frac{1}{x_{(k+1)}} + \sum_{i=k+2}^{n-k-1}\frac{1}{x_{(i)}} + \frac{1}{k+1}\cdot\frac{1}{x_{(n-k)}}\right]}$$

بعلاوه میانگین هندسی پیراسته متقارن را به شکل

$$\bar{x}_{TG(k,k)} = \sqrt[n-2k]{\prod_{i=k+1}^{n-k} x_{(i)}}$$

و میانگین هندسی وینزوری متقارن را به شکل زیر تعریف می‌کنیم

$$\bar{x}_{WG(k,k)} = \sqrt[n]{x_{(k+1)}^{k+1}\cdot\prod_{i=k+2}^{n-k-1}x_{(i)}\cdot x_{(n-k)}^{k+1}}.$$

٤. برآوردگرهای استوار مقیاس

فرض کنید n مشاهده مستقل $x_1,\dots,x_n$ از متغیر تصادفی X را در اختیار داریم و m برآوردگر مکانی است. انحراف n مشاهده از برآوردگر مکانی را به صورت $d_1 = x_1 - m,\dots,d_n = x_n - m$ نشان داده و بوسیله انتخاب تابع مناسب g(d) و با میانگینی از n مقدار

برآوردگری برای پراکندگی بر اساس انتخاب مناسب n به شکل زیر بدست می‌آوریم.

$$S=\frac{1}{n^*}\sum_{i=1}^{n} g(d_i)$$

که در آن لازم نیست n^* برابر n باشد. بنابراین با تعریف فوق کلاس گسترده‌ای از برآوردگرهای پراکندگی را خواهیم داشت. برای مثال در واریانس و انحراف معیار نمونه $n^*=n-1, g(d)=d^2, m=\bar{x}$ می‌باشد. در میانگین قدرمطلق انحراف از میانگین (MAD)، $m=\bar{x}$، $n^*=n\,, g(d)=|d|$ می‌باشند. بنابراین برای برآوردگر استوار پراکندگی از برآوردگر استوار مکان استفاده می‌نماییم. همچنین انتخاب برآورد استوار پراکندگی بجای اینکه بر اساس n مشاهده باشد می‌تواند بر اساس قسمتی از آنها باشد.

الف) برآوردگرهای وینزوری و پیراسته

برآوردگرهای وینزوری و پیراسته برای واریانس عبارتند از

$$S^2_{w(k,k)}=\frac{1}{n-1}\Big\{(k+1)(x_{(k+1)}-\bar{x}_{w(k,k)})^2+\sum_{j=k+2}^{n-k-1}(x_{(j)}-\bar{x}_{w(k,k)})^2+(k+1)(x_{(n-k)}-\bar{x}_{w(k,k)})^2\Big\}.$$

$$S^2_{T(k,k)}=\frac{1}{n-2k-1}\sum_{j=k+1}^{n-k}(x_{(j)}-\bar{x}_{T(k,k)})^2.$$

ب) میانه قدر مطلق انحراف

برآوردگر دیگر میانه قدر مطلق انحراف (MAD) از میانه می‌باشد (هامپل ۱۹۷۴). با در نظر گرفتن $g(d)=|d|$ خواهیم داشت

$$S_m=Median\{|x_1-\tilde{x}|,\ldots,|x_n-\tilde{x}|\}$$

که برآوردگر مفیدی برای پراکندگی است. به شکل زیر نیز می‌توان S_m را نشان داد.

$$MAD=med_i(|x_i-med_j(x_j)|)$$

که در آن میانه داخلی ($med_j(x_j)$) میانه n مشاهده و میانه خارجی med_i میانه n انحراف از میانه می‌باشد. برای توزیع نرمال ۱/۴۸۲۶MAD می‌تواند برای برآورد انحراف معیار بکار رود. آماره MAD کارایی پایین برای توزیع نرمال دارد و برای توزیعهای متقارن مناسب نخواهد بود.

ج) یک برآورد نوع L، ساده و مفید برای پراکندگی دامنه میان چارکی است

$$Q=\frac{1}{2}\{x_{(q_2)}-x_{(q_1)}\}$$

که در آن q_2,q_1 نزدیکترین اعداد صحیح به $3n/4, n/4$ هستند. انحراف معیار نمونه برآوردگر متداول مقیاس جامعه است اما به مقادیر پرت حساس می‌باشد و ممکن است وقتی که یک داده با یک عدد دلخواه جایگزین شود کراندار نماند. ولی حتی وقتی که قسمتی از داده‌ها با اعدادی دلخواه جایگزین می‌شوند برآوردگر استوار مقیاس برآوردها کراندار باقی می‌مانند. برای یک جامعه نرمال انحراف معیار (σ) می‌تواند بوسیله تقسیم نمودن دامنه چارکی بر ۱/۳۴۸۹۸ برآورد شود.

د) برآوردگر نوع M

همانطور که در مورد برآوردگر نوع M مربوط به میانگین اشاره شد برآورد با می‌نیمم نمودن $\sum_{j=1}^{n}\rho(x_j-T_n)$ بدست می‌آید و اگر $\rho(\cdot)$ پیوسته و مشتق آن $\psi(t)$ می‌باشد، μ را بوسیله T_n برآورد می‌کنیم بطوریکه $\sum_{j=1}^{n}\psi(x_j-T_n)=0$ می‌باشد. حال اگر $\psi(u)=uf'(u)/f(u)+1$ در نظر بگیریم برآوردگر درستنمایی ماکزیمم σ یعنی s_n را بدست می‌آوریم. با انتخاب $\psi(u)=\mathrm{sgn}(|u|-1)$ برآوردگر انحراف میانه از صفر ($Median\{|x_j|\}$) را خواهیم داشت.

و) میانگین تفاضل جینی

میانگین تفاضل جینی نیز برآوردگر استوار انحراف معیار می‌باشد که به صورت زیر تعریف می‌شود.

$$G=\frac{1}{\binom{n}{2}}\sum_{i<j}|x_i-x_j|$$

اگر مشاهدات از توزیع نرمال باشند، $\frac{\sqrt{\pi}}{۲}\cdot G$ برآوردگر نااریب انحراف معیار می‌باشد.

ی) برآوردگرهای S_n و Q_n

روسیو و کراکس [۱۰] در سال ۱۹۹۳ دو آماره جدید که متفاوت با MAD هستند را به نامهای S_n و Q_n ارائه نمودند.

$$S_n = ۱/۹۲۶ med_i(med_j \mid x_i - x_j \mid)$$

که میانه خارجی (med_i) میانهٔ n میانه از $\{\mid x_i - x_j \mid ; j = ۱,۲,\dots,n\}$ می‌باشد. برای تصحیح اریبی در نمونه‌های کوچک $C_{sn}S_n$ برای برآورد انحراف معیار بکار می‌رود که C_{sn} عامل تصحیح است (کراکس و روسیو ۱۹۹۲). آماره دوم به صورت زیر تعریف می‌شود.

$$Q_n = ۲/۲۲۱۹\{ \mid x_i - x_j \mid ; i < j\}_{(k)}$$

که $k = \binom{h}{۲}$, $h = \left[\frac{n}{۲}\right] + ۱$ و Q_n تقریباً ۲/۲۲۱۹ برابر آماره مرتبه k ام از $\binom{n}{۲}$ فاصله میان نقاط می‌باشد. آماره تصحیح اریبی شده $C_{qn}Q_n$ برای برآورد انحراف میانه بکار می‌رود که C_{qn} عامل تصحیح است.

۵. آزمون تی استوار برای میانگین

مجموع مربعات انحراف وینزوری به صورت زیر تعریف می‌شود.

$$SS^{۲}_{w(k,k)} = (k+۱)(x_{(k+۱)} - \bar{x}_{(w(k,k))})^{۲} + \sum_{i=k+۲}^{n-k-۱}(x_{(i)} - \bar{x}_{w(k,k)})^{۲} + (k+۱)(x_{(n-k)} - \bar{x}_{w(k,k)})^{۲}$$

یک برآورد استوار واریانس برای میانگین پیراسته بر اساس مجموع مربعات انحرافات وینزوری توسط توکی و مک لاگلین [۱۱] در سال ۱۹۶۳ ارائه شده است. همانطور که می‌دانیم آزمون فرض در مورد میانگین یکی از رایجترین تحلیلهای آماری است. این آزمون برای میانگین وینزوری به صورت زیر تعریف می‌شود.

آزمون تی پیراسته برای آزمون فرضیه $H_0 : \mu = \mu_0$ در مقابل $H_۱ : \mu \neq \mu_۰$ بر اساس آماره زیر محاسبه می‌شود

$$T_t = \frac{\bar{x}_{T(k,k)} - \mu_۰}{Std\ Err(\bar{x}_{T(k,k)})}$$

که در آن $Std\ Err(\bar{x}_{T(k,k)})$ خطای استاندارد $\bar{x}_{T(k,k)}$ می‌باشد

$$Std\ Err(\bar{x}_{T(k,k)}) = \frac{SS_{w(k,k)}}{\sqrt{(n-۲k)(n-۲k-۱)}}$$

آزمون تی وینزوری برای آزمون فرضیه $H_0 : \mu = \mu_0$ در مقابل $H_۱ : \mu \neq \mu_۰$ عبارتست از

$$T_w = \frac{\bar{x}_{w(k,k)} - \mu_۰}{Std\ Err(\bar{x}_{w(k,k)})}$$

که در آن

$$Std\ Err(\bar{x}_{w(k,k)}) = \frac{n-۱}{n-۲k-۱}\cdot\frac{SS_{w(k,k)}}{\sqrt{n(n-۱)}}$$

وقتی که داده ها از توزیعی متقارن باشند، توزیع آماره t پیراسته (T_t) یا t وینزوری (T_w) می‌توانند با توزیع تی استودنت با $n-۲k-۱$ درجه آزادی تقریب زده شوند (توکی و مک لاگین ۱۹۶۳، دیکسن و توکی ۱۹۶۸) [۳].

٦. شبیه‌سازی

در این قسمت با استفاده از شبیه‌سازی مونت کارلو ۲۰۰ مشاهده نرمال با میانگین ۲ و انحراف معیار ۵ تولید می کنیم. بوسیله رویه‌های Capability و Univariate از نرم افزار SAS ۸ برای مشاهدات حاصل، برآوردهای استوار مکان یعنی میانگین پیراسته ۱۰٪ و میانگین وینزوری ۱۰٪، به شکل دو طرفه و یکطرفه را محاسبه می کنیم. همچنین فرض $H_0 : \mu = \mu_۰$ درمقابل $H_۱ : \mu \neq \mu_۰$ توسط آزمون تی استوار (وینزوری و پیراسته) برای مشاهدات تولید شده بررسی می‌شود. بعلاوه برآوردگرهای استوار مقیاس مانند دامنه چارکی، میانگین تفاضل جینی و MAD و S_n, Q_n را محاسبه می کنیم. در ادامه از قابلیتهای رویه Stdize برای محاسبه برآوردگرهای استوار مکان و مقیاس به روشهای (IQR,MAD,ABW(C),AHUBER(K), AWAVE(K), L(P)) استفاده می کنیم.

```
 Var x;
Run;
Proc Stdize Method = ABW (4.685) Pstat;
 Title 'Biweight Method';
 Var x;
Run;
Proc Stdize Method = AHuber (1.5) Pstat;
 Title 'Huber Method';
 Var x;
Run;
Proc Stdize Method = Awave (1.339) Pstat;
 Title 'Andrews Wave Method';
 Var x;
Run;
Proc Stdize Method = L(1) Pstat;
 Title 'L1 Method';
Var x;
Proc Stdize Method = L(2) Pstat;
 Title 'L2 Method';
 Var x;
Run;
```

```
Data test;
 Do n=1 to 200;
  X=2+5*Rannor (136456);
 Output;
End;
Proc Means data = test;
  Var x;
Run;
Proc Capability Trimmed = 0.1 ( type = twosided alpha
  = 0.05) winsorized = 0.1 (type = twosided alpha =0.05
  ) Robust Scale;
  Var x;
Run ;
Proc  Capability  Trimmed  =  0.1 ( type = upper alpha =
  0.1 ) Winsorized = 0.1 (type = lower alpha = 0.01);
  Var x;
Run ;
Proc Univarbility Trimmed = 0.1 (type = twosided alpha
  =0.05) Winsorized = 0.1 (type = twosided alpha = 0.05
  ) Robust Scale;
  Var x;
Run ;
Proc Stdize Method = Mad Pstat;
  Title 'MAD Method';
```

خروجی مربوط به رویه Means که میانگین، انحراف معیار، کمینه و بیشینه را محاسبه می‌نماید.

The MEANS Procedure

N	Means	Std Dev	Minimum	Maximum
200	2.1605148	4.9619601	-16.3805028	14.9141577

خروجی مربوط به رویهٔ Capability که میانگینهای استوار، فواصل اطمینان استوار و آزمونهای تی استوار (پیراسته و وینزوری) را محاسبه می‌نماید.

The CAPABILITY Procedure

*** Trimmed Means ***

Percent Trimmed in Tail	Number Trimmed in Tail	Trimmed Mean	Std Error Trimmed Mean	95% Confidece Limits		DF	t for H0: Mu0=0.00	Pr>|t|
10.50	21	2.165358	0.357186	1.459847	2.870869	157	6.062262	<.0001

ادامهٔ خروجی مربوط به رویهٔ Capability که برآوردهای استوار مقیاس را ارائه می‌نماید.

Robust Measures of Scale

Measure	value	Estimate of Sigma
Interquartile Range	6.796220	5.038043
Gini's Mean Difference	5.557296	4.925025
MAD	3.379222	5.010035
Sn	4.787394	4.787394
Qn	4.972321	4.879608

خروجی مربوط به رویهٔ Stdize که برآوردگرهای استوار مکان و مقیاس را محاسبه می‌نماید.

The STDIZE Procedure

Location and Scale Measures

Location = median   Scale = meadian abs dev from median

Name	Location	Scale	N
x	2.553060	3.379222	200

Location = biweight 1- step M- estiamte   Scale = biweight A – estimate

x	0.026321	1.567310	200

Location = Huber 1- step M- estiamte   Scale = Huber A – estimate

x	0.002321	0.953031	200

Location = Wave 1- step M- estiamte   Scale = Wave A – estimate

x	0.137833	7.104601	200

Location = L(p)   Scale = L(p)

x	0.022223	0.109231	200

مراجع

[1] Anderws, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J Rogers, W.H. and Tukey, J.W., 1972. *Robust Estimation of Location: Survey and Advances,* Princeton University Press, Princeton, pp. 54,78,80-90.

[2] David, H.A. and Shu, V.S., 1979. *Robustness of Location Estimators in the Presence of an outliers,* pp. 235-250.

[3] Dvaid, H.A., 1978. *Contributions to Survey Sampling and Applied Statistics,* in honour, Hartley. Academic Press , New York.

[4] Dixon, W.J. and Tukey, J.W., 1968. *Approximate Behavior of the Distribution of Winsorized t (Trimming/Winsorization)* , Technometrics, Vol.10, pp. 83-98.

[5] John, M.V., 1979. *Robust Pitman- like Estimators,* pp. 49-60.

[6] John, M.V., Launer, R.L. and Wilkinson, G.N., 1979. *Robustness in statistics,* Academic Press, New York.

[7] Huber, P.J., 1964. *Robust Estimation of Location Parameter,* Annal. Math. Stat., Vol.35, pp. 73-101.

[8] Huber, P.J., 1981. *Robust Statistics,* John Wiley & Sons.

[9] Hodges, J.L.Jr. and Lehmann, E.L., 1963. *Estimates of Location Based on Rank Tests,* Ann. Math. Stat. Vol.34, pp. 598-611.

[10] Hampel, F.R., 1974. *The Influence Curve and its Role in Robust Estimation,* J. Amer. Stat. Asso. Vol.69, pp. 383-393.

[11] Gastwirth, J.L. and Cohen, M.L., 1970. *Small Sample Behavior of Some Robust Linear Estimators of Location,* J. Amer. Stat. Asso. Vol.65, pp. 964-973.

[12] Rousseeuw, P.J. and Croux, C., 1993. *Alternatives to the Median Absolute Deviation,* Jour. Amer. Stat. Asso., Vol. 88, pp. 1273-1283.

[13] Tukey, J.W. and McLaughlin, D.H., 1963. *Less Vulnerable Confidence and Significance Procedures for Location Based on a Single Sample: Trimming/ Winsorization1,* Sankhya A, Vol. 25, pp. 331-352.