# رگرسـ ون ژرفا در حالت چند گانه

**حمـ د شر ف**

**بانک مرکز  جمهور  اسلام  ا ران**

**چکـ ده**   در ا ن مقاله  ه  ان ت  ر  ه ژرفا در رگرسـ ون می‌پرداز م  رگرسـ ون ژرفا[1] که عدد صحـ حی  ن  و $n$ است  ه عنوان خاصـ تی از  رازش رگرسـ ونی در ن  ر گرفته می‌شود که رتبه  رازش رگرسـ ونی تعب ر می‌شود و مع اری  رای مقا سه  رازشهای مختلف رگرسـ ونی می  اشد  در ا ن مقاله  ا  ان رگرسـ ون ژرفا در حالت چندگانه  ه  ررسی خواص  ن پرداخته و روش ژرفتر ن رگرسـ ون را که نسبت  ه تبد لات  کنوا  ر متغ ر پاسخ هم‌ورداست  ارا ه می‌ده م  ا معرفی مع ار مقدار فرو ر زش نشان می‌ده م که ا ن روش  روشی استوار می  اشد  در پا ان  روردکننده ژرفتر ن رگرسـ ون را  ا د گر  روردکننده‌های رگرسـ ونی  ا محاسبه رگرسـ ون ژرفای  نها  رای داده‌های نمونه‌ای موجود مقا سه می‌کن م

**واژه‌ها  کلـ د**   ژرفا  نا رازا[2]  ژرفتر ن رگرسـ ون  مقدار فرو ر زش  رگرسـ ون ژرفای ماکسـ مال  هم‌وردا ی

## مقدمه

وقتی رگرسـ ون کمتر ن توانهای دوم را  ا استفاده از $n$ مشاهده  رای  ک مدل پارامتری $Y = X\beta + \varepsilon$  ه کار می  ر م فر  ها ی را در مورد  ردار خ  اها $\varepsilon$ در ن  ر می‌گ ر م   کی از ا ن فر  ها نرمال  ودن توز   خ  اهاست  در عمل انحرافاتی از ا ن فر  ها رخ می‌دهد  مثلا ممکن است توز  ز ر  نا ی خ  اها متقارن اما غ ر نرمال  اشد  ا ت زتر از نرمال  وده و دمهای کوتاهتری داشته  اشد  ا کش دگی کمتر از نرمال  ا دمهای پهن‌تر داشته  اشد و  ا ممکن است توز    ه صورت نرمال  اشد ولی دارای دور افتاده‌ها ی  اشد

در ا ن موارد  ه جای روش کمتر ن توانهای دوم از روشهای رگرسـ ون استوار استفاده می‌شود که در مقا سه  ا روش کمتر ن توانهای دوم نسبت  ه ا ن انحرافات حساسـ ت کمتری دارند   ک روش رگرسـ ونی استوارکه اخ را توسـ  روسـ ف و هو رت[3]   مـ رح شده است روش ژرفتر ن رگرسـ ون نام دارد که  ر پا ه ت  ر  ه رگرسـ ون ژرفا  نا شده است  رگرسـ ون ژرفا ک ف ت  رازش را اندازه‌گ ری کرده و مـ زان دوری  ن را از هر نا رازا اندازه می‌گ رد و   ان می‌کند که ا ن  رصفحه  رازشی در توصـ ف داده‌ها  ه چه اندازه خو   عمل می‌کند   نا را ن  رازشی  ا ژرفای  زرگ نسبت  ه

1) regression depth   2) nonfit   3) Rousseeuw and Hubert

داده‌ها متعادلتر است و لذا ک رازش خو ژرفای زرگتری نسبت ه ک رازش د دارد در ا ن مقاله ا مرور رگرس ون ژرفا در حالت چندگانه ه ررسی خواص ن پرداخته و روش ژرفتر ن رگرس ون را ان می‌کن م

در خش دوم ا ن مقاله ور مختصر رگرس ون ژرفا را در حالت ساده مرور می‌کن م در خش سوم رگرس ون ژرفا را در حالت چندگانه معرفی می‌کن م رگرس ون ژرفای ماکس مال و ژرفتر ن رگرس ون ه ترت در خشهای چهارم و پنجم ان می‌شود و در خش ششم خواص ژرفتر ن رگرس ون در حالت چندگانه ررسی می‌گردد ا ن خواص شامل هم‌وردا ی و استواری ژرفتر ن رگرس ون است و در انتها ا ارا ه مثال روردکننده ژرفتر ن رگرس ون را ا د گر روردکننده‌های رگرس ونی مقا سه می‌کن م

## مرور ر رگرس ون ژرفا در حالت ساده

هدف در رگرس ون ساده رازش ک خ راست $y = \theta_{۱}x + \theta_{۲}$ ه ک مجموعه داده $Z_n = \{(x_i, y_i); i = \quad , ..., n\} \subseteq R^{۲}$ است ک رازش را ه صورت $\theta = (\theta_{۱}, \theta_{۲})$ نشان می‌ده م که مولفه اول ن رورد ش و مولفه دوم ن جمله عرض از مبدا است مانده‌های مجموعه داده $Z_n$ متناس ا رازش $\theta$ را ه صورت $r_i(\theta) = r_i = y_i - \theta_{۱}x_i - \theta_{۲}$ نشان می‌ده م رای معرفی ژرفای ک رازش ا تدا ک نا رازا را تعر ف می‌کن م

**تعر ف ۱** رازش $\theta = (\theta_{۱}, \theta_{۲})$ رای مجموعه داده $Z_n$ نا رازا نام ده می‌شود اگر و فق اگر ک عدد حق قی $v_\theta = v$ مخالف همه $x_i$ ها وجود داشته اشد ه وری که

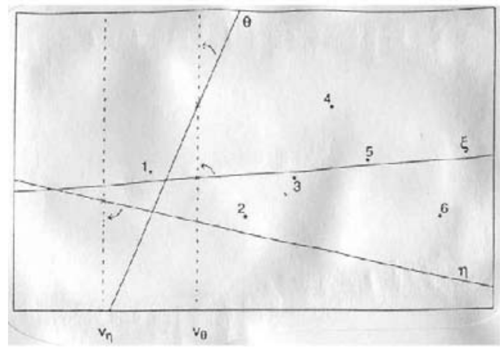
$$\forall\ x_i < v \ , \quad r_i(\theta) < \qquad , \qquad \forall\ x_i > v \ , \quad r_i(\theta) >$$

ا

$$\forall\ x_i < v \ , \quad r_i(\theta) > \qquad , \qquad \forall\ x_i > v \ , \quad r_i(\theta) <$$

خ ی که الا ا پا ن همه مشاهدات واق می‌شود هم شه ک نا رازا است عدا خواه م د د که نا رازاها در واق رازش‌ها ی ا ژرفای صفر هستند

شکل ک مجموعه داده را ا مشاهده و $\theta$ و $\eta$ نا رازا نشان می‌دهد مقاد ر مر و ه $v_\theta$ و $v_\eta$ نشان داده شده‌اند ا توجه ه شکل وجود $v$ متنا ر ا وجود ک ر د در روی خ عمودی که از نق ه $v$ رسم شده است چون $\forall\ x_i < v_\theta \quad < r_i(\theta)$ و همچن ن $\forall\ x_i > v_\theta \quad > r_i(\theta)$ لذا ا توجه ه تعر ف فوق $\theta$ نا رازا است همچن ن چون خ $\eta$ در پا ن همه مشاهدات قرار دارد ا ن خ ن ز نا رازا است از ن جا که تعر ف فوق رای خ $\xi$ صادق ن ست ا ن خ نا رازا ن ست همان ور که ذکر شد ا مشخص شدن $v$ نق ه متنا ر ا ن که محل تلاقی خ عمودی ترس م شده از نق ه $v$ ا خ رگرس ونی است و ا ر در نشان داده شده است مشخص می‌گردد ه عبارت د گر محل تلاقی خ رازش داده شده ا خ عمودی که از نق ه $v$ متنا ر ا ن خ ترس م می‌گردد ا ر د ر مشخص شده است از ا ن




شکل      مجموعه داده دو متغ ره  ا نا رازاهای $\theta$ و $\eta$ و  ک  رازش $\xi$  ا رگرس ون ژرفای

رو می‌توان خ  و   نا رازا را از روی شکل   مشخص کرد   ه ا ن ترت.  که خ   رازش داده شده را حول نق ه‌ای که  ا   ر در  ر روی  ن خ   مشخص شده است می‌گردا نم تا عمودی شود   اگر در ح ن چرخاندن  خ   از ه چ مشاهده‌ای عبور نکند نا رازا است   نا را ن خ  و $\theta$ و $\eta$  نا رازا هستند ولی خ   $\xi$  نا رازا ن ست چون وقتی  ن را حول نق  ه مذکور  مشخص  شده  ا   ر در  می‌گردا نم تا عمودی شود از مشاهدات    و    عبور می‌کند
 ه  ورکلی ژرفای  ک  رازش $\theta$ رای  ک مجموعه داده $Z_n$  ه حجم $n$  ه صورت ز ر   ان می‌شود

**تعر ف**      رگرس ون ژرفای  $rdepth$)  ک رازش $\theta$ متناس    ا مجموعه داده $Z_n$ عبارت است ازکمتر ن تعداد مشاهداتی که  ا د  رداشته شود تا $\theta$ نا رازا شود     ه   ور معادل  $rdepth(\theta, Z_n)$ عبارت است ازکوچکتر ن تعداد مانده‌ها ی که می  ا ست علامتشان تغ  ر کند تا $\theta$ نا رازا شود

 رای مثال خ   $\xi$ را در شکل    در ن ر می‌گ ر م    ا ن خ     ا حذف کردن مشاهدات    و    نا رازا می‌شود  ز را  ا قرار دادن $v_\xi$ مساوی  ا $v_\theta$ و حذف مشاهدات    و    می‌توان خ   $\xi$ را  ه صورت عمودی درورد  دون  ن که از مشاهده‌ای عبور کند    چون خ   $\xi$  ا حذف حداقل    مشاهده نا رازا می‌شود لذا   $rdepth(\xi, Z_n) =$

**توجه ۱**   تعار ف    و    در مواقعی که در $x_i$ ها تکرار وجود داشته  اشد  $x_i$ ها  ا هم مساوی  اشند  ن ز  رقرار است    و $x_i$ ها   ه ه چ فرض توز عی  ن از ندارند

**مرتبه زمانی**   گو  م $f(n)$ از مرتبه زمانی $g(n)$ است و  ن را  ا نماد $f(n) = O(g(n))$ نشان می‌ده م اگر و تنها اگر اعداد صح ح و مثبتی مانند $n_۰$ و $c$ وجود داشته  اشد    ور که  $f(n) \leq c\, g(n)$  رای تمام $n$های $n \geq n_۰$  رقرار  اشد

 رای محاسبه $rdepth(\theta, Z_n)$  ا تدا مشاهدات را  ه صورت $x_۱ \leq x_۲ \leq \cdots \leq x_n$ در

مرتبه زمانی $O(nlogn)$ مرتة می‌کنءم تعرءف می‌کنءم

$$L^{+}(v) = \#\{j; x_j \leq v, r_j \geq \ \}$$

و

$$R^{-}(v) = \#\{j; x_j > v, r_j \leq \ \}$$

$L^{-}$ و $R^{+}$ ه صورت مشاه تعرءف می‌شوند سپس $L^{+}(x_i)$ $L^{-}(x_i)$ $R^{-}(x_i)$ و $R^{+}(x_i)$ رای هر $i = \ , \cdots, n$ مشاه تعارءف فوق محاسبه می‌شوند $rdepth(\theta, Z_n)$ در $O(n)$ عمل ه صورت زءر محاسبه می‌شود

$$rdepth(\theta, Z_n) = \min_{۱\leq i\leq n} (\min\{L^{+}(x_i)+R^{-}(x_i)\ ,\ R^{+}(x_i)+L^{-}(x_i)\}) \quad (\ )$$

مثال زءر نحوه محاسبه رگرسءون ژرفای ءک خ را ا استفاده از را ه نشان می‌دهد
**مثال ۱** شکل مشاهده ا خ $\xi$ را نشان می‌دهد همانءورکه دءده می‌شود اءن خ نا‌رازا ءست ا استفاده از را ه رگرسءون ژرفای اءن خ ه صورت زءر محاسبه می‌شود

$$i = ۱ : \min\{L^{+}(x_۱) + R^{-}(x_۱), R^{+}(x_۱) + L^{-}(x_۱)\} = \min\{۱ + ۳, ۳ + ۱\} = ۴$$
$$i = ۲ : \min\{L^{+}(x_۲) + R^{-}(x_۲), R^{+}(x_۲) + L^{-}(x_۲)\} = \min\{۱ + ۲, ۳ + ۲\} = ۵$$
$$i = ۳ : \min\{L^{+}(x_۳) + R^{-}(x_۳), R^{+}(x_۳) + L^{-}(x_۳)\} = \min\{۲ + ۲, ۲ + ۲\} = ۴$$
$$i = ۴ : \min\{L^{+}(x_۴) + R^{-}(x_۴), R^{+}(x_۴) + L^{-}(x_۴)\} = \min\{۲ + ۱, ۲ + ۳\} = ۳$$
$$i = ۵ : \min\{L^{+}(x_۵) + R^{-}(x_۵), R^{+}(x_۵) + L^{-}(x_۵)\} = \min\{۳ + ۱, ۱ + ۳\} = ۴$$
$$i = ۶ : \min\{L^{+}(x_۶) + R^{-}(x_۶), R^{+}(x_۶) + L^{-}(x_۶)\} = \min\{۳ + ۰, ۱ + ۴\} = ۳$$
$$i = ۷ : \min\{L^{+}(x_۷) + R^{-}(x_۷), R^{+}(x_۷) + L^{-}(x_۷)\} = \min\{۴ + ۰, ۰ + ۴\} = ۴$$

بنابراءن

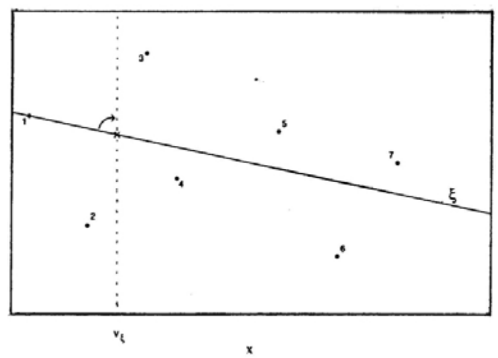
$$rdepth(\xi, Z_n) = \min(۴, ۵, ۴, ۳, ۴, ۳, ۴) = ۳$$

از روی شکل نءز دءده می‌شود که ا رداشتن مشاهدات و و می‌توان خ $\xi$ را دون اءن‌که از مشاهدات دءگر عبورکند حول نق ه متنا ر ا $v_\xi$ که ا ر در روی ن خ مشخص شده است دوران داد تا ه صورت عمودی در ءد ه عبارت دءگر ا رداشتن مشاهدات و و خ $\xi$ نا‌رازا می شود رگرسءون ژرفای خ $\xi$ را ر ا کران پاءءن رگرسءون ژرفای ماکسءمال عنی $= \lceil\frac{۷}{۳}\rceil = \lceil\frac{n}{۳}\rceil$ است که در خش‌های عد اءان می‌شود
**توجه** $\lceil\lambda\rceil$ کوچکترءن عدد صحءح زرگتر ا مساوی $\lambda$ است

## رگرسءون ژرفا در حالت چندگانه

در رگرسءون چندگانه مجموعه داده $Z_n$ ه صورت

$$Z_n = \{(x_{i۱}, \ldots, x_{i,p-۱}, y_i); i = \ , \ldots, n\} \subset \mathbb{R}^p$$




شکل       مجموعه داده دو متغ ره و رازش $\xi$ ا رگرس ون ژرفای

است $X$ را ه عنوان قسمتی از مختصات هر نق ه ه صورت
$X_i = (x_{i۱}, \ldots, x_{i,p-۱}) \in \mathbb{R}^{p-۱}$

در ن ر می‌گ ر م  اکنون می‌خواه م $y_i$ را ه وس له

$$\theta_۱ x_{i۱} + \cdots + \theta_{p-۱} x_{i,p-۱} + \theta_p = (X_i,\ \ )\theta'$$

عنی ه وس له ک ا رصفحه ف ن در $\mathbb{R}^p$ که $\theta = (\theta_۱, \ldots, \theta_p) \in \mathbb{R}^p$ است رازش ده م
در ا نجا ه چ فرض توز عی وجود ندارد

قبل از ان تعار ف و م ال ان خش ا تدا ا رصفحه و ا رصفحه ف ن را تعر ف می‌کن م

**تعر ف ۳** فرض کن د $[f = \alpha]$ نشان دهنده مجموعه س ح $\{x; f(x) = \alpha\}$ اشد ک ا رصفحه مجموعه‌ای ه شکل $[f = \alpha]$ است که $f$ تا عک خ ی غ رصفر روی ف ای رداری $X$ و $\alpha$ عددی حق قی است ه عبارت د گر مجموعه $H \subseteq \mathbb{R}^n$ ک ا رصفحه است اگر اعداد حق قی $a_۱, \ldots, a_n$ و $C$ ا $a_i \neq$ رای حداقل ک $i$ وجود داشته اشند ه وری که $H$ متشکل از همه نقا $x = (x_۱, \ldots, x_n)$ ی اشد که در را ه $\sum_i \alpha_i x_i = C$ صدق می‌کنند

**تعر ف** ک ا رصفحه ف ن در ف ای رداری $X$ مجموعه‌ای مانند $\mathcal{M}$ است ه وری که رای هر $x_۰$ در $\mathcal{M}$ $\mathcal{M} - x_۰$ ک ا رصفحه اشد ه عبارت د گر $\mathcal{M} \subseteq X$ ک ا رصفحه ف ن است اگر ک تا خ ی غ رصفر $f : X \longrightarrow \mathbb{F}$ و ک $\alpha$ در $\mathbb{F}$ وجود داشته اشد ه وری که $\mathcal{M} = \{x \in X \,;\, f(x) = \alpha\}$ $\mathbb{F}$ م دان حق قی $\mathbb{R}$ ا م دان مختل $\mathbb{C}$ است

**مثال** ف ای رداری $X = \mathbb{R}^۲$ و تا $f(x,y) = y$ را در ن ر می‌گ ر م در ا ن صورت مجموعه $\{(x,\ \ ); x \in \mathbb{R}\} = \{(x,y); f(x,y) = \ \ \} = [f = \ \ ]$ ک ا رصفحه است که

ف ای $\mathbb{R}^{۲}$ را به دو نیم‌ف ای بسته $\{(x,y); f(x,y) \geq \ \}$ و $\{(x,y); f(x,y) \leq \ \}$ افراز می‌کند مجموعه

$$C = \{(x,y); f(x,y) = \ \} = \{(x,\ ); x \in \mathbb{R}\}$$

یک ابرصفحه فین است زیرا بازای هر $(x_\circ,\ ) \in C$

$$C - (x_\circ,\ ) = \{(x - x_\circ,\ ); x \in R\} = \{(x - x_\circ,\ ); f(x - x_\circ,\ ) = \ \}$$

یک ابر صفحه است

**تعریف** برازش $\theta = (\theta_۱, \ldots, \theta_p)$ یک نابرازا نامیده می‌شود اگر وفقط اگر یک ابرصفحه فین $V$ در ف ای x وجود داشته باشد به وری‌که هیچ یک از $\mathrm{x}_i$ ها متعلق به $V$ نباشند و برای همه $\mathrm{x}_i$ ها در یکی از نیم‌ف اهای ن داشته باشیم $r_i(\theta) >$ و همچنین برای همه $\mathrm{x}_i$ ها در نیم‌ف ای باز دیگر $r_i(\theta) <$ باشد

شکل مثالی از یک صفحه نابرازاست در این مثال $p =$ است بنابراین $\theta$ به یک صفحه متعلق است ف ای $X$ به صورت صفحه افقی با $y \equiv$ که شامل خ $V$ است دیده می شود با این تعریف هر $\eta$ به وری‌که همه مانده‌ها مثبت یا همه مانده‌ها منفی باشند یک نابرازاست زیرا کافی است $V$ به صورتی انتخاب شود که همه $\mathrm{x}_i$ ها در یک رف ن واقع شوند

رگرسیون ژرفای هر $\theta \in \mathbb{R}^p$ متناسب با $Z_n \subset \mathbb{R}^p$ همانند تعریف به صورت زیر تعریف می‌شود

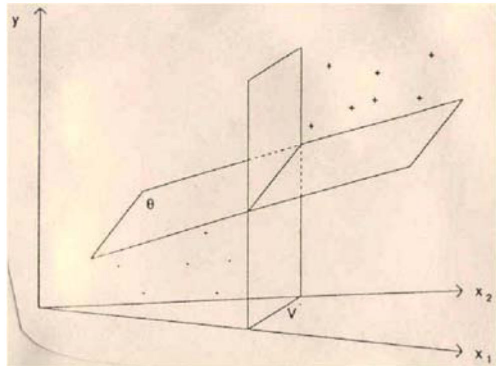
**تعریف** رگرسیون ژرفای یک برازش $\theta \in \mathbb{R}^p$ متناسب با مجموعه داده $Z_n \subset \mathbb{R}^p$ عبارت است از کوچکترین تعداد مشاهداتی که وقتی از مجموعه داده‌ها خارج شود $\theta$ نابرازا گردد به عبارت دیگر $rdepth(\theta, Z_n)$ عبارت است از کمترین تعداد مانده‌هایی که باید تغییر علامت داده شوند تا $\theta$ نابرازا شود

$rdepth(\theta, Z_n)$ همانند را به ان می‌شود بدین صورت که هر ابرصفحه فین $V$ در ف ای x مشاهدات را به دو مجموعه تقسیم می‌کند که نها را با $L(V)$ و $R(V)$ نشان می‌دهیم مجموعه‌های $L^{+}(V)$ و $L^{-}(V)$ به ترتیب تعداد مشاهدات در $L(V)$ با مانده‌های مثبت و تعداد مشاهدات در $L(V)$ با مانده‌های منفی است $R(V)$ نیز به ور مشابه به دو مجموعه $R^{+}(V)$ و $R^{-}(V)$ تقسیم می‌گردد بنابراین

$$rdepth(\theta, Z_n) = \min_{V}(\min\{L^{+}(V) + R^{-}(V)\ ,\ R^{+}(V) + L^{-}(V)\})$$

که $V$ شامل همه ابرصفحه‌های فین در ف ای x است برای محاسبه رگرسیون ژرفای یک برازش جستجو را به مجموعه‌ای متناهی از ابرصفحه‌های $V$ محدود می‌کنیم الگوریتم‌های مناسب برای محاسبه رگرسیون ژرفا در فصل بعد ارائه می‌شود

لازم به ذکر است که حالت تعمیم یافته ق ..به یعنی در حالت چندگانه $p$ بعدی زمانی که $Z_n = \{(X_i, y_i); i = \ , \ldots, n\}$ که $X_i = (x_{i۱}, \ldots, x_{i,p-۱}) \in \mathbb{R}^{p-۱}$ و




شکل مثالی از نا رازا $\theta \in \mathbb{R}^{۳}$ ا رصفحه ف ن $V$ در ف ای $X$ $(\mathbb{R}^{۲})$ مشاهدات ا مانده‌های مثبت را از مشاهدات ا مانده‌های منفی جدا می‌کند

$\theta = (\theta_{۱}, \ldots, \theta_p) \in \mathbb{R}^p$ اشد ن ز رقرار است

ق ه ۱ **خاص ت دق ق برازش** اگر تعداد مشاهداتی که روی $\theta$ قرار گ رند $k$ اشد که $\leq k \leq n$ نگاه

$$k \leq rdepth(\theta\ ,\ Z_n) \leq \left[\frac{n+k}{\ }\right]$$

رای $k = n$ دار م $rdepth(\theta, Z_n) = n$

□

**تو ح ۱** از نجاکه رورد‌کننده $L^{۱}$ هم شه از حداقل $p$ مشاهده می‌گذرد لذا رگرس ون ژرفای ن حداقل $p$ است لومف لد و است گر[۴] را ب ن د همچن ن ا ن خاص ت رای رورد‌کننده کمتر ن توان‌های دوم پ راسته $(LTS)$ که ه وس له روس ف م رح شده رقرار است روس ف

**تو ح** رورد کننده کمتر ن توان‌های دوم $(LS)$ هرگز نا رازا ن ست در حق قت اگر $Z_n$ ه صورتی اشد که $X_n = \{x_{۱}, x_{۲}, ..., x_n\}$ رتبه کامل داشته اشد و $n \geq \ p$ سپس $rdepth(\theta_{LS}, Z_n) \geq$
□

## رگرس ون ژرفا ماکس مال

کران‌های الا رای رگرس ون ژرفای ماکس مال در ق ه ز ر ارا ه می‌شود ز ر مجموعه‌ای از $\mathbb{R}^p$ در موقع ت عمومی قرار دارد هرگاه شتر از $p$ مشاهده در هر ز رف ای ف ن $(p - \ )$ عدی قرار نگ رد

4) Bloomfield and Steiger

**ق ه** اگر $(\mathrm{x}_i, y_i)$ در موقع ت عمومی اشند نگاه

$$\max_{\theta} rdepth(\theta, Z_n) \leq \left[\frac{n+p}{}\right]$$

□

**حدس ۱** رای هر مجموعه داده $Z_n \subset \mathbb{R}^p$ دار م

$$\max_{\theta} rdepth(\theta, Z_n) \geq \frac{n}{p+}$$

## ژرفتر ن رگرس ون

ژرفتر ن رگرس ون $T_r^*$ عنوان $\theta$ ا ی که $rdepth(\theta, Z_n)$ را . شه نه می‌کند تعر ف شده است ژرفتر ن رگرس ون $T_r^*$ "متعادل‌تر ن" رازش است $T_r^*$ را می‌توان ه وس له محاسبه رگرس ون ژرفای همه رازش‌ها ی که از م ان $p$ مشاهده می‌گذرد ه دست ورد

**تعر ف** روردکننده ژرفتر ن رگرس ون $T_r^*(Z_n)$ عبارت است از رازش $\theta$ ا زرگتر ن رگرس ون ژرفا عنی

$$T_r^*(Z_n) = arg\max_{\theta} rdepth(\theta, Z_n)$$

**توجه ۳** در رگرس ون چندگانه ژرفتر ن رگرس ون رخلاف $L^1$ و $LTS$ رای تبد لات کنوای $y$ هم‌وردا است

**مثال ۳** می‌خواه م را ه ن م زان خار مصرفی ماهانه توس ک ماش ن خار $Y$ ا متوس دمای ماهانه ه فارنهات $X_1$ و تعداد روزهای کار در ماه $X_2$ را مورد رسی قرار می‌ده م د ن من ور مشاهده جم وری شده است که در جدول نشان داده شده‌اند ه دراپر و اسم ت[5] مراجعه شود

ا استفاده از روش‌های رگرس ونی مختلف را رگرس ونی را رورد کرده و رگرس ون ژرفای هر ک از ا ن رازش‌ها را ه دست ورده‌ا م نتا ج در جدول مده است د ده می‌شود که روردکننده ژرفتر ن رگرس ون رگرس ون ژرفای الاتری نسبت ه سا ر روردکننده‌ها دارد و درنت جه متعادل‌تر ن رازش است

## خواص ه نگ ژرفتر ن رگرس ون

درا ن خش ع ی از خواص هم‌وردا ی ژرفتر ن رگرس ون را ان می‌کن م همچن ن ا معرفی کردن مقدار فرور زش ه‌عنوان مع ار پا داری روردکننده در را ر نقا پرت نشان می‌ده م که روش ژرفتر ن رگرس ون ا توجه ه ا ن مع ار روشی استوار است

5) Draper and Smith

جدول      داده‌های ماش ن خار

| $X_۱$ | $X_۲$ | $Y$ | $X_۱$ | $X_۲$ | $Y$ |
|---|---|---|---|---|---|
| | | | | | |

جدول      مقا سه رگرس ون ژرفای روردکننده‌های رگرس ونی ا روردکننده ژرفتر ن رگرس ون

| روردکننده ژرفتر ن رگرس ون | رورد هو ر | $LAD$ | $LS$ | روش |
|---|---|---|---|---|
| | | | | $\hat{\beta}_۰$ |
| | | | | $\hat{\beta}_۱$ |
| | | | | $\hat{\beta}_۲$ |
| | | | | رگرس ون ژرفا |

## ۱ هم‌وردا ی ژرفتر ن رگرس ون

ررسی خواص فوق در ق ه ز ر مده است
ق ه ۳ مجموعه داده $Z_n = \{(X_i, Y_i); i = \ , \dots, n\}$ که در ن

$$X_i = (x_{i,۱}, \dots, x_{i,p-۱}) \in \mathbb{R}^{p-۱}$$

و روردکننده ژرفتر ن رگرس ون $T_r^*(Z_n) = arg_{\theta}\max\ rdepth(\theta, Z_n)$ را در ن ر می‌گ ر م
**a** روردکننده ژرفتر ن رگرس ون رگرس ون هم‌وردا است عنی رای هر ردار ستونی $V$

$$T_r^*(Z_n') = T_r^*(Z_n) + V$$

که $Z_n' = \{(X_i, y_i + X_i V); i = \ , \dots, n\}$ و $T_r^*(Z_n') = arg_{\theta}\max\ rdepth(\theta, Z_n')$
**b** روردکننده ژرفتر ن رگرس ون مق اس هم‌وردا است عنی رای هر $c$ ثا ت

$$T_r^*(Z_n') = c\ T_r^*(Z_n)$$

که $Z_n' = \{(X_i, cy_i);\ i = \ , \dots, n\}$ و $T_r^*(Z_n') = arg_{\theta}\max\ rdepth(\theta, Z_n')$ **c**

روردکننده ژرفتر ن رگرس ون ف ن هم‌وردا است عنی رای هر ماتر س مر ناو ژه $A$

$$T_r^*(Z_n') = A^{-۱}\ T_r^*(Z_n)$$

که $Z_n' = \{(X_i A, y_i);\ i = \ , \dots, n\}$ و $T_r^*(Z_n') = arg_{\theta}\max\ rdepth(\theta, Z_n')$

## استوار ژرفتر ن رگرس ون

ک مع ار معروف رای پا داری روورد کننده در را ر نقا پرت مقدار فرور زش است نا ه تعر ف داناهو و هو ر[۶] مقدار فرور زش هر روردکننده $T_n$ ه صورت ز ر است

$$\varepsilon_n^*(T_n, Z_n) = min\left\{\frac{k}{n};\ \sup_{Z_n'} ||T_n(Z_n') - T_n(Z_n)|| = \infty\right\}$$

که $Z_n'$ شامل همه مجموعه داده‌های دست مده ا جا گز نی هر $k$ مشاهده دلخواه $Z_n$ ا مقاد ر اخت اری است نا را ن مقدار فرور زش کوچکتر ن کسری از مشاهدات تبد ل شده مغشوش شده ای است که می‌تواند روردکننده را ه دلخواه منحرف کند لازم ه ذکر است که

6) Donoho and Huber

ا ن اغتشاش فقط به نقا پرت در $y_i$ محدود نمی‌شود بلکه $Z'_n$ می‌تواند شامل نقا پرت در $x_i$ ها نیز باشد

**حدس** رای هر مجموعه داده $Z_n \subset \mathbb{R}^p$ با $\mathrm{x}_i$ ها در موقعیت عمومی باشند نگاه

$$\varepsilon_n^*(T_r^*, Z_n) \geq \frac{}{n}\left( \frac{n}{p+} - p + \right) \approx \frac{}{p+}$$

را به ان می‌کند که مقدار فرور زش ژرفتر ن رگرس ون همواره مثبت است ا ن مقدار زمانی برابر با کران پا ن $\frac{۱}{p+۱}$ می‌شود که داده‌های اصلی خود و ژه باشند مثلا وقتی نها روی منحنی گشتاور قرار گ رند
بنابرا ن ژرفتر ن رگرس ون در برابر نقا نافذ همانند نقا دورافتاده عمودی استوار است بعلاوه ژرفتر ن رگرس ون متفاوت از رگرس ون $L^۱$ تعر ف شده به صورت $L^۱(Z_n) = arg \min_\theta \sum_{i=۱}^{n} |r_i(\theta)|$ است لازم به ذکر است که مقدار فرور زش $L^۱$ صفر است که در نت جه س پذ ری ن در برابر نقا نافذ می باشد

## مثال‌ها

در ا ن بخش برآوردکننده ژرفتر ن رگرس ون را برای داده‌های موجود بدست ورده و نشان می‌ده م که در مقا سه با برآوردکننده‌های د گر دارای ب شتر ن رگرس ون ژرفاست ا ن مقا سه با محاسبه رگرس ون ژرفای برآوردکننده‌ها انجام می‌گ رد بد ن منظ ور از الگور تمها ی که برنامه مر و به نها با استفاده از نرم‌افزار Splus و ز ان برنامه‌نو سی فرترن نوشته شده است و از و سا ت http://win-www.uia.ac.be/u/statis/ قابل دسترسی است استفاده می‌گردد

### ۱ مثال ۱

در ک تحق ق در سال م زان لودگی رودخانه در ا الت ن و ورک بررسی شده است هدف بررسی سهم هر نو از زم ن‌های مورد استفاده در ا راف رودخانه بر روی لودگی رودخانه‌هاست ا ن لودگی براساس م انگ ن غل ت ن تروژن م لی‌گرم اندازه‌گ ری می‌شود نو زم ن‌های مورد استفاده در ا راف رودخانه‌ها و م زان لودگی نها در ز ر ب ان شده است داده‌ها در جدول نشان داده شده‌اند به چاترجی و همکاران[7] مراجعه شود

$Y$ م انگ ن غل ت ن تروژن م لی‌گرم بر ل تر براساس نمونه‌ها ی که در فواصل منظ م بهار تا ستان و ماههای پا ز گرفته شده است

7) Chatterjee and et al.

جدول   داده‌های رودخانه‌های ن و ورک

| $X_۱$ | $X_۲$ | $X_۳$ | $X_۴$ | $Y$ | $X_۱$ | $X_۲$ | $X_۳$ | $X_۴$ | $Y$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | |

جدول   رورد پارامترهای مدل و رگرسه ون ژرفای روردکننده‌های مختلف رگرسه ونی

| روردکننده ژرفتر ن رگرسه ون | رورد هو ر | $LAD$ | $LS$ | روش |
|---|---|---|---|---|
| | | | | $\hat{\beta}_۰$ |
| | | | | $\hat{\beta}_۱$ |
| | | | | $\hat{\beta}_۲$ |
| | | | | $\hat{\beta}_۳$ |
| | | | | $\hat{\beta}_۴$ |
| | | | | رگرسه ون ژرفا |
| | | | | $\sum e_i^۲$ |
| | | | | $\sum \lvert e_i \rvert$ |

$X_۱$   کشاورزی   درصد مساحت زم ن مورد استفاده رای کشاورزی

$X_۲$   جنگلی   درصد زم ن جنگلی

$X_۳$   مسکونی   درصد مساحت زم ن مورد استفاده رای سکونت

$X_۴$   تجاری صنعتی   درصد مساحت زم ن مورد استفاده تجاری ا صنعتی

ا ررسی داده‌های فوق د ده می‌شود که مشاهدات و دور افتاده و نافذ و در ع ن حال موثر هستند نا را ن انت ار می‌رود که روش‌های $LS$ $LAD$ و رورد هو ر تحت تاث ر ا ن نقا قرار گ رند و ور قا ل ملاح ه رگرسه ون ژرفای کوچکتری نسبت ه روردکننده ژرفتر ن رگرسه ون داشته اشند ا نگاه کردن ه جدول مشخص می‌شود که روردکننده رگرسه ونی $LAD$ و رورد هو ر نزد ک ه هم هستند و دارای رگرسه ون ژرفای هستند چون روردکننده ژرفتر ن رگرسه ون نسبت ه نقا دورافتاده و نافذ استوار است لذا تحت تاث ر ا ن نقا قرار نمی‌گ رد و ه ن ر می‌رسد که رازشی مناسه رای داده‌هاست ا توجه ه ا نکه دارای شتر ن رگرسه ون ژرفا عنی است

جدول    جدول    رورد پارامترهای مدل و رگرس ون ژرفای روردکننده‌های مختلف رگرس ونی

| روش | $LS$ | $LAD$ | رورد هو ر | روردکننده ژرفتر ن رگرس ون |
|---|---|---|---|---|
| $\hat{\beta}_{0}$ | | | | |
| $\hat{\beta}_{1}$ | | | | |
| $\hat{\beta}_{2}$ | | | | |
| $\hat{\beta}_{3}$ | | | | |
| $\hat{\beta}_{4}$ | | | | |
| رگرس ون ژرفا | | | | |
| $\sum e_i^2$ | | | | |
| $\sum \lvert e_i \rvert$ | | | | |

## مثال    شب ه‌ساز

رای ررسی مناس  ودن روردکننده ژرفتر ن رگرس ون     نق ه در   عد م ا ق ا مدل $y = - \quad + x_1 + \quad x_2 - \quad x_3 + \quad x_4 + e$ تول د می‌کن م که $x_1$ $x_2$ $x_3$ $x_4$ و $e$ از توز  نرمال استاندارد گرفته شده‌اند  پس از محاسبه $y$ ا استفاده از مقاد ر تول د شده $x_1$ $x_2$ $x_3$ $x_4$ و $e$ روردکننده‌های رگرس ونی  دست  مده از روش‌های مختلف و روردکننده ژرفتر ن رگرس ون را  رای ا ن مقاد ر  دست می ور م  نتا ج  دست  مده در جدول ز ر  مده‌اند

ا مقا سه  را   رورد شده  دست  مده از روش‌های مختلف د ده می‌شود که  را  دست  مده از روش ژرفتر ن رگرس ون نسبت  ه سا ر روش‌ها  ه  را  مدل مفروض نزد ک‌تر است و در ع ن حال دارای  شتر ن رگرس ون ژرفا است   ه عبارت د گر روردکننده ژرفتر ن رگرس ون

$$y = - \quad + \quad x_1 + \quad x_2 - \quad x_3 + \quad x_4$$

دارای رگرس ون ژرفای تقر بی    است

## نت جه‌گ ر

وقتی که استواری و تشخ ص نقا  دورافتاده اهم ت دارند روش $LMS$ روشی مناس  است اما زمانی که نقا  دورافتاده ز ادی وجود ندارد  می‌توان  ه وس له روش‌های $LMS$ ا $LTS$ ررسی کرد  و  کنوا ی اهم ت دارد روش ژرفتر ن رگرس ون $T_r^*$ انتخا ی مناس  است

## مراجع

[1] Bloomfield, P., and Steiger, W. (1983). Least Absolute Deviations: Theory, Applications, and Algorithms, Boston:Birkhauser.

[2] Chatterjee, S., and Hadi, A.S., and Price, B. (2000), Regression Analysis by Example, New York: John Wiley.

[3] Donoho, D.L., and Huber, P.J. (1983), "The notion of breakdown point", in A Festschrift for Erich Lehmann, eds. Bickel, P., Doksum, K., and Hodges, J.L., Belmont: Wadsworth.

[4] Draper, N., Smith, H. (1998) Applied Regression Analysis, third edition, John Wiley & Sons.

[5] Rousseeuw, P.J., (1984) "Least Median of Squares Regression", Journal of the American Statistical Association, **79**, 871-880.

[6] Rousseeuw, P.J., and Hubert, M. (1999), "Regression Depth", Journal of the American Statistical Association, **94**, 388-402.