

Scatter Plot

&

How to Draw & Use Scatter Plot in Papers ?

Session 2:

farvardin, 15, 1405

April 4, 2026

Presented by:

A Keshtkar MD, MPH, PhD

TUMS – SPH

Researchware Academy

Season 2

Statistical Reviews



مجموعه اول از سری
جلسات «مرور آماری»

نمودار پراکنش

و چگونه باید این نمودار را ترسیم
نموده و در مقالات بکار برد؟

جلسه دوم این مجموعه :
شنبه ۱۵ فروردین ۱۴۰۵
۴ آوریل ۲۰۲۶

طراحی، تدوین و تدریس:
دکتر عباس کشتکار
عضو هیات علمی دانشگاه تهران
مدیر آکادمی پژوهش افزار

فصل اول: ۲ جلسه

مرور آماری

مجموعه اول از سری
جلسات «مرور آماری»



برازش مدلها در «نمودار پراکنش»



نکات کاربردی مهم در نمودار پراکنش: (۲)

✓ از آنجا که «نمودار پراکنش»، ابزار مشترک در «تحلیل همبستگی» و «تحلیل رگرسیون (خطی)» است، فلذا در کار با این نمودار باید حتماً با تحلیل رگرسیون خطی و غیرخطی آشنایی داشته و حداقل اینکه علاوه بر رگرسیون خطی، با انواع پر کاربرد رگرسیونهای غیرخطی آشنایی داشته باشید:

■ رگرسیون خطی

■ رگرسیون درجه دو یا Quadratic

■ رگرسیون درجه سه یا Cubic

■ رگرسیون چند جمله‌ای کسری یا Fractional Polynomial

■ رگرسیون ناپارامتری یا LOESS / LOWESS: Local regression / LOcally WEighted Scatterplot Smoothing

■ ...

مثالهای کاربردی در نمودار پراکنش: (۱)

✓ مثال ۱) در نمودار روبرو که رابطه

بین سن با تراکم استخوانی در زنان

یک جامعه را نشان میدهد،

میخواهیم با هدف ارزیابی اکتشافی،

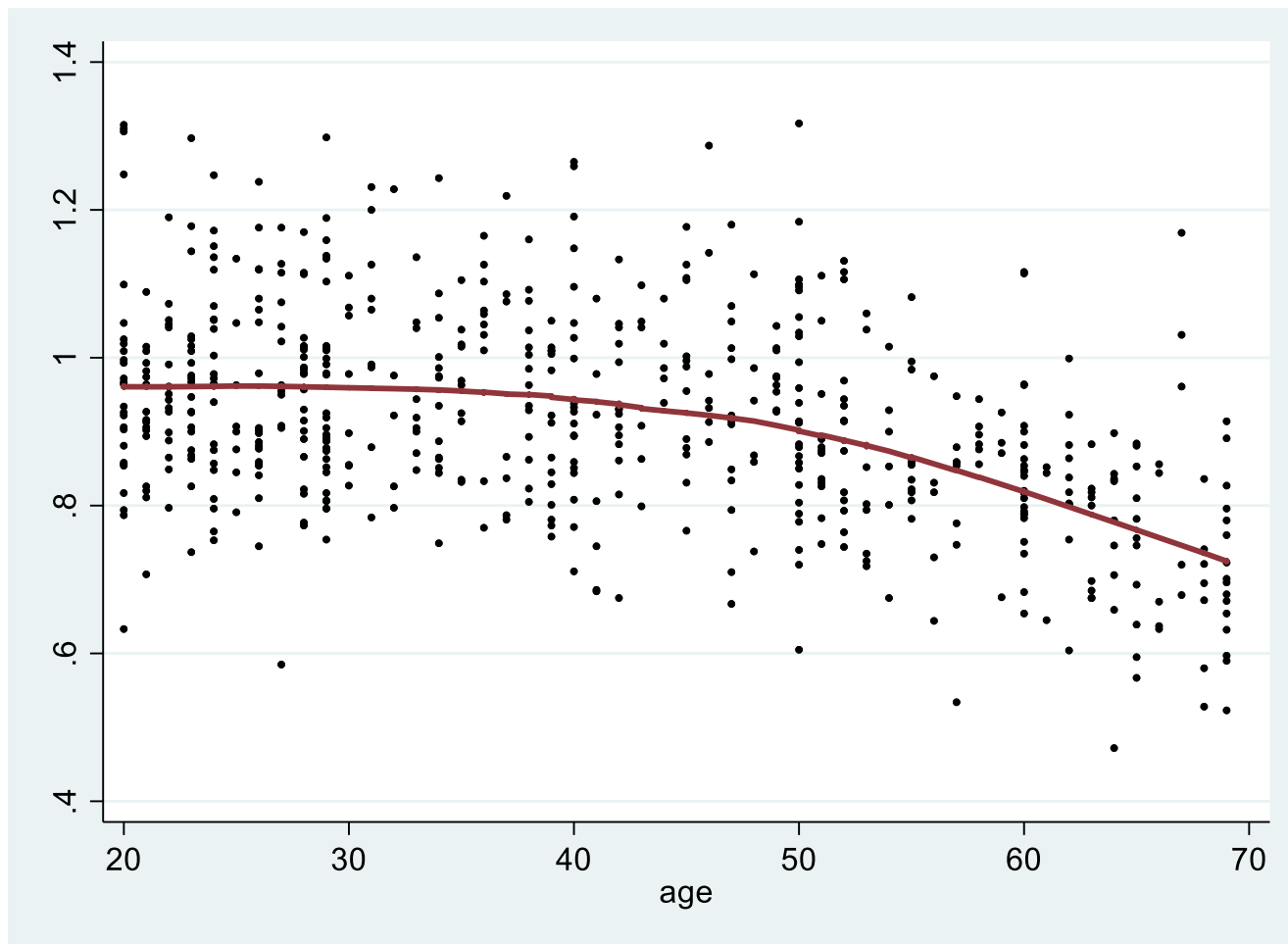
بینیم ماهیت این رابطه به کدامیک

از روابط رگرسیونی، نزدیک است؟

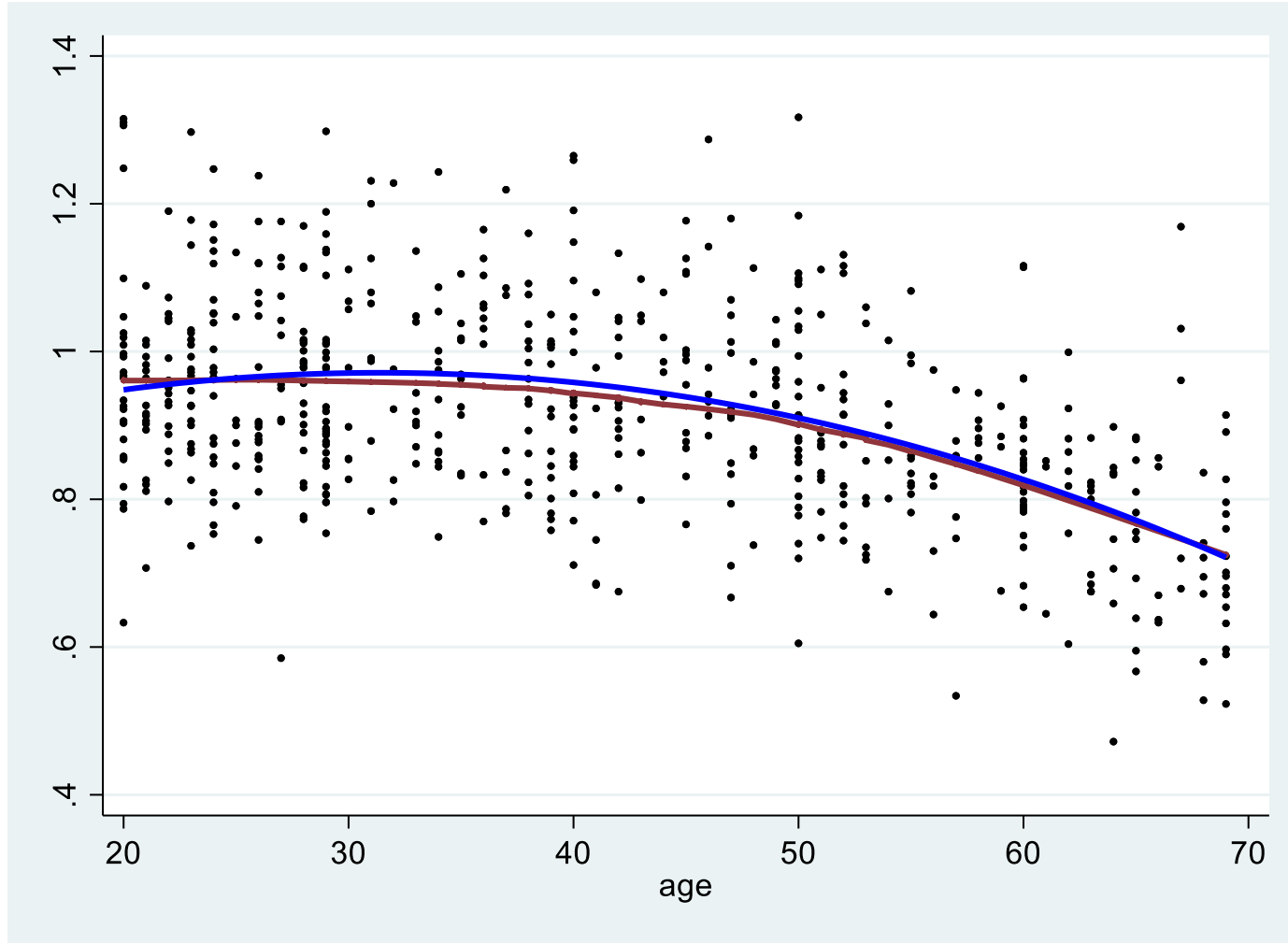
■ ابتدا از نمودار پراکنش LOWESS

میتوانیم استفاده کنیم که بتوانیم رفتار

دیتای مشاهده شده را بهتر بینیم!



مثالهای کاربردی در نمودار پراکنشی: (۲)



✓ مثال (۱) در نمودار روبرو که رابطه بین

سن با تراکم استخوانی در زنان یک

جامعه را نشان میدهد، میخواهیم با هدف

ارزیابی اکتشافی، بینیم ماهیت این

رابطه به کدامیک از روابط رگرسیونی،

نزدیک است؟

■ اگر به نمودار قبلی، برازش درجه ۲ را

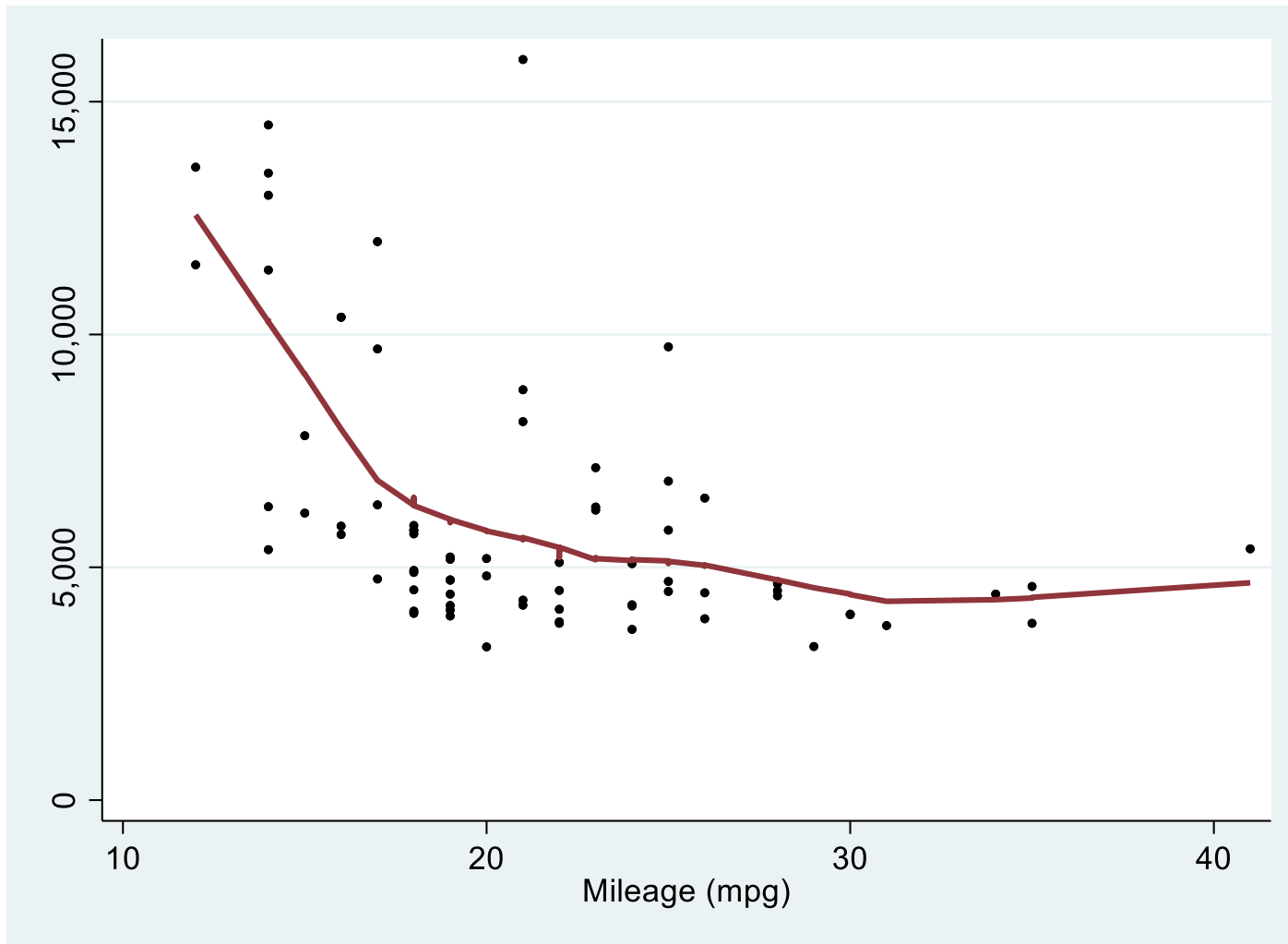
اضافه کنیم، مشاهده میکنیم که رفتار این

دو الگو بشدت بهم نزدیک است. پس

میتوانیم از رگرسیون درجه ۲ برای این

برآورد، استفاده نماییم.

مثالهای کاربردی در نمودار پراکنش: (۳)



✓ مثال ۲) در نمودار روبرو رابطه بین

شاخص مایل مسافت طی شده به ازای

هر گالن بنزین با قیمت خودروهای امریکا

در دهه ۷۰ میلادی را مشاهده میکنید،

میخواهیم با هدف ارزیابی اکتشافی،

بینیم ماهیت این رابطه به کدامیک از

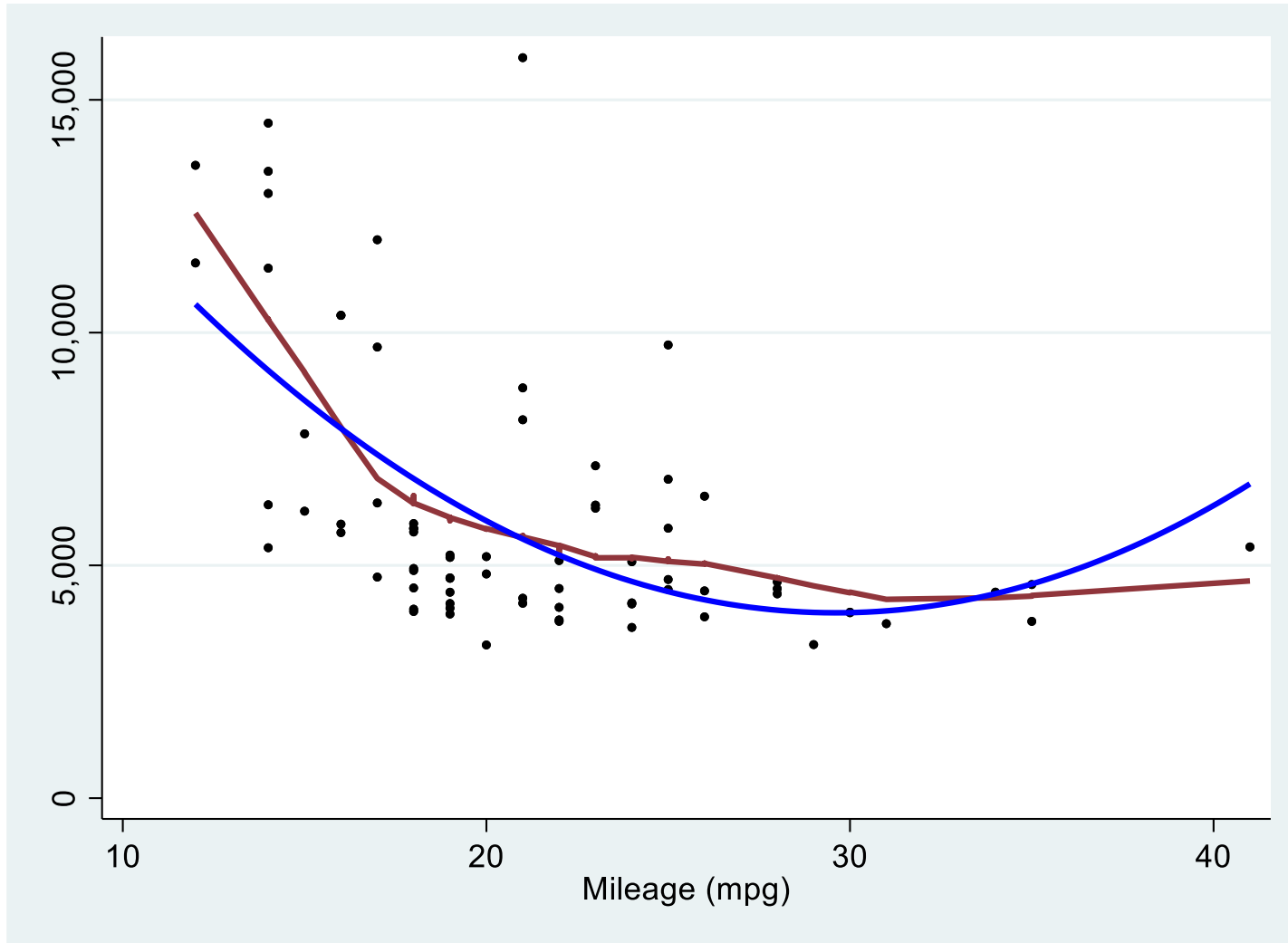
روابط رگرسیونی، نزدیک است؟

■ ابتدا از برازش LOWESS استفاده میکنیم.

شاید بنظر برسد که این الگو نیز مانند

مثال قبلی به الگوی درجه ۲ نزدیک باشد

مثالهای کاربردی در نمودار پراکنش: (۴)



✓ مثال ۲) در نمودار روبرو رابطه بین

شاخص مایل مسافت طی شده به ازای

هر گالن بنزین با قیمت خودروهای امریکا

در دهه ۷۰ میلادی را مشاهده میکنید،

میخواهیم با هدف ارزیابی اکتشافی،

بینیم ماهیت این رابطه به کدامیک از

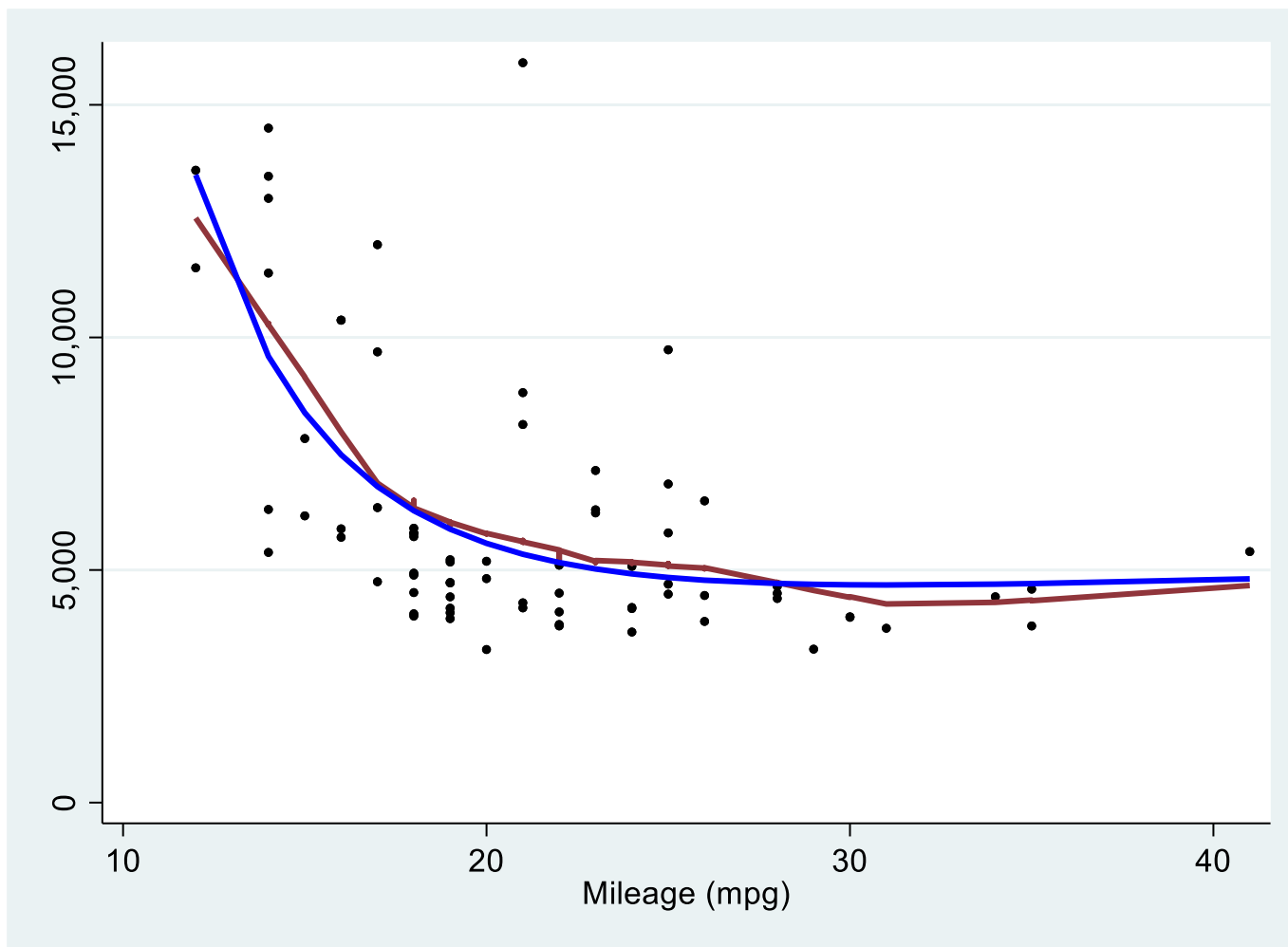
روابط رگرسیونی، نزدیک است؟

■ ابتدا از برازش LOWESS استفاده میکنیم.

شاید بنظر برسد که این الگو نیز مانند

مثال قبلی به الگوی درجه ۲ نزدیک باشد

مثالهای کاربردی در نمودار پراکنش: (۵)



✓ مثال ۲) در نمودار روبرو رابطه بین شاخص مایل مسافت طی شده به ازای هر گالن بنزین با قیمت خودروهای امریکا در دهه ۷۰ میلادی (دلار) را مشاهده میکنید، میخواهیم با هدف ارزیابی اکتشافی، ببینیم ماهیت این رابطه به کدامیک از روابط رگرسیونی، نزدیک است؟

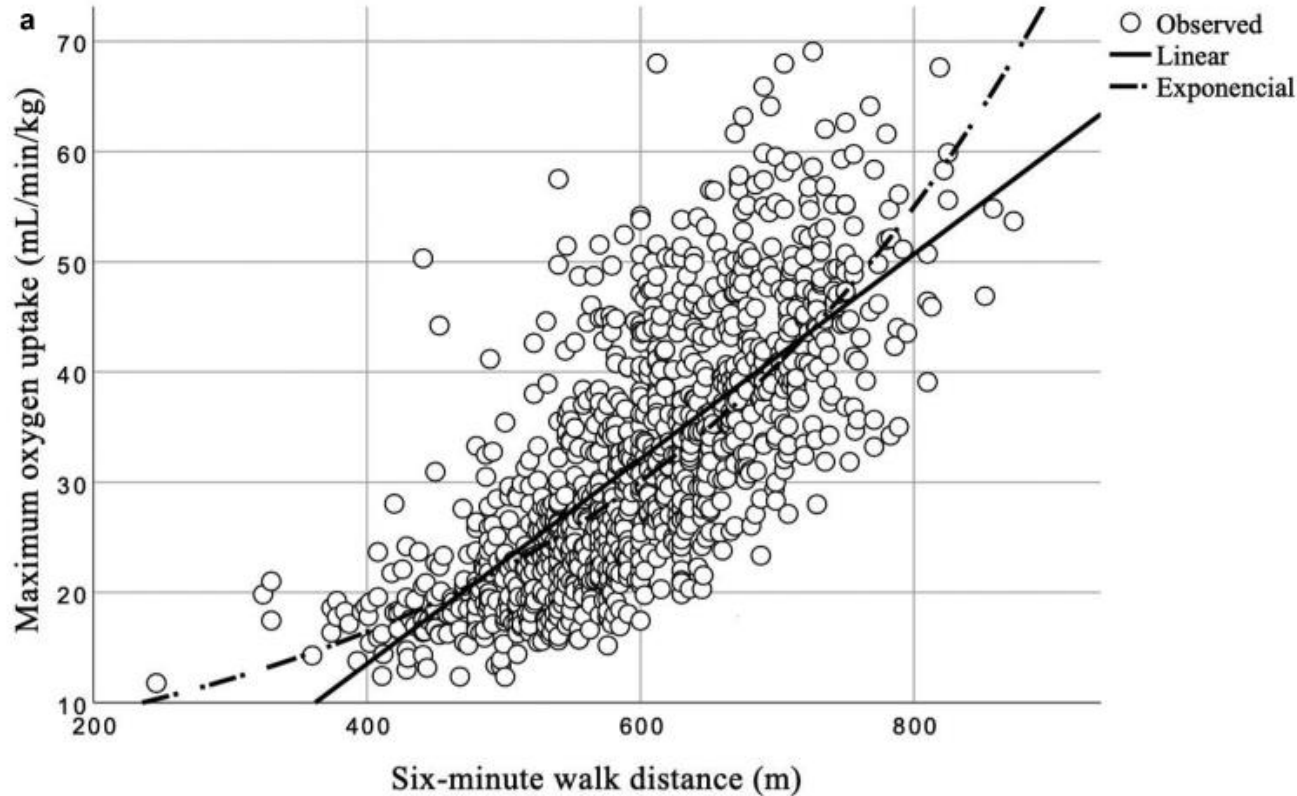
■ این بار برازش «رگرسیون چندجمله‌ای کسری» را امتحان میکنیم! شرایط برازشی، بهبود قابل توجهی نشان میدهد.

Predicting maximum oxygen uptake using the six-minute walk distance in adults: What is the best curve fit estimation?

The study involved 1,295 adult participants aged 18–80 years. To predict the $\dot{V}O_{2\max}$ directly assessed in the treadmill cardiopulmonary exercise testing (CPET) by 6MWD alone, we used the statistical package curve estimation tool to identify the best-fit equation (e.g., linear, exponential, quadratic). After identifying the best fit, a $\dot{V}O_{2\max}$ predictive equation was obtained based on the 6MWD. Further, we secondarily elaborated a 6MWD prediction using a multiple regression equation considering age, sex, weight, height, body mass index and classic cardiovascular risk factors as predictors. After a sensitivity analysis and a collinearity diagnosis using Variation Inflation Factor (VIF), the main predictors for the 6MWD were established.

We found that $\dot{V}O_{2\max}$ can be better predicted from 6MWD following an exponential equation (Figure 1A): $[\dot{V}O_{2\max} (\text{mLO}_2 \bullet \text{kg}^{-1} \bullet \text{min}^{-1}) = 4.910 \times \exp (0.003 \times \text{DTC6(m)})]; R^2 = 0.548]$. This exponential model demonstrated a higher accuracy than a linear model $[\dot{V}O_{2\max} (\text{mLO}_2 \bullet \text{kg}^{-1} \bullet \text{min}^{-1}) = -23.535 + (0.093 \times \text{DTC6(m)})]; R^2 = 0.500]$, especially for individuals with lower CRF. The intraclass coefficient of correlations (95% confidence interval) were 0.795 (0.729 to 0.844) and 0.801 (0.732 to 0.866) for linear and exponential models, respectively. Although the linear model's R^2 value is close, the standard error of the estimate was $8.12 \text{ mLO}_2 \bullet \text{kg}^{-1} \bullet \text{min}^{-1}$ in the linear model compared to only $0.24 \text{ mLO}_2 \bullet \text{kg}^{-1} \bullet \text{min}^{-1}$ in the exponential model. On the other hand, both models underestimated higher values of $\dot{V}O_{2\max}$ with a smaller bias related to the exponential model (Figures 1B–1E). Additionally, we provided an alternative prediction equation for 6MWD in which age, sex, and height accounted for 38.3% of the total variability in the 6MWD while cardiovascular risk factors explained the remaining 11.8% (Table 1).

مثال کاربردی در مقاله : (۲)



✓ همانطور که در نمودار پراکنش

مشاهده میکنید، رابطه شاخص مسافت

پیاده روی شش دقیقه‌ای با ماکزیمم

دریافت اکسیژن با دو الگوی خطی و

نمایی (Exponential) مورد ارزیابی

قرار گرفته است:

■ کاملاً بدیهی است که الگوی برازش

«نمایی» از الگوی برازش خطی، بهتر

میتواند این دیتا را پیش‌بینی نماید.

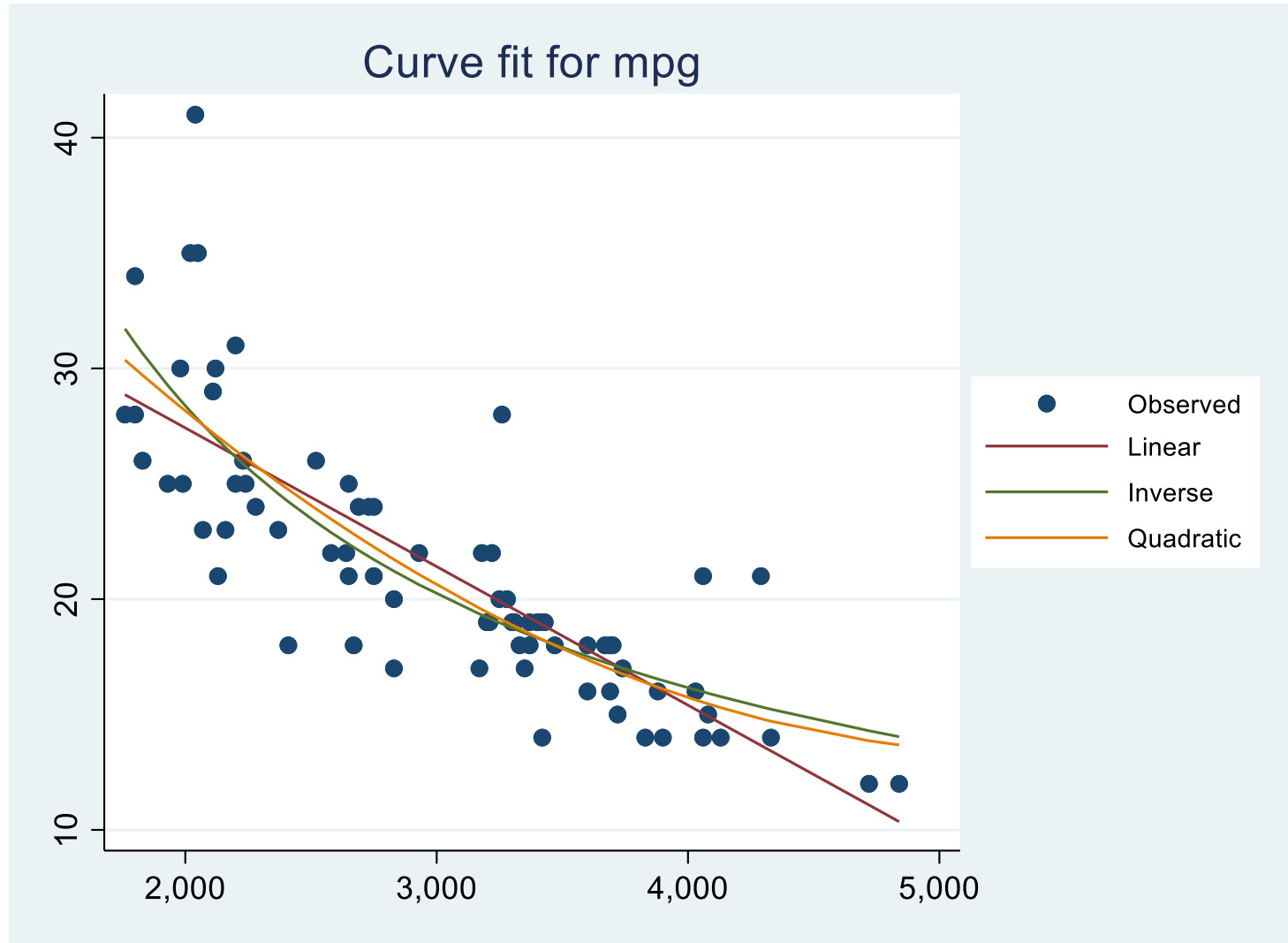
نکات کاربردی مهم در نمودار پراکنش: (۳)

با توجه به تعدد مدل‌های مختلف برازشی در رابطه بین دو متغیر کمی در شرایط مختلف و حیطه‌های علمی متفاوت، چگونه میتوانیم برای یک رابطه مشخص بین دو متغیر X و Y در یک نمودار پراکنش، مدل‌های مختلف برازشی را مورد ارزیابی قرار داد؟؟؟

✓ در نرم‌افزار Stata یک ماژول نان کُر به نام **“curvefit”** تدوین شده که میتواند تا حداکثر ۳۵ مدل آماری یا برازشی مختلف را مورد ارزیابی قرار دهد.

✓ در این ماژول، خوشبختانه هم **خروجی نموداری** (نمودار پراکنش به همراه مدل‌های برازشی) و هم **خروجی جدولی** (پارامترهای معادلات رگرسیونی مربوطه به همراه شاخص برازش R^2) ارائه میشود.

مثالهای کاربردی در نمودار پراکنش: ماژول curvefit



✓ در دیتاست خودروهای امریکایی،

میخواهیم رابطه بین **وزن خودروها** با

مصرف سوخت آنها را ارزیابی کنیم. از

بین مدل‌های زیر، کدام مناسبتر است؟

■ مدل خطی

■ مدل درجه ۲

■ مدل معکوس یا Inverse

مثالهای کاربردی در نمودار پراکنش: ماژول curvefit

✓ در دیتاست خودروهای امریکایی،

میخواهیم رابطه بین وزن خودروها با

مصرف سوخت آنها را ارزیابی کنیم. از

بین مدل‌های زیر، کدام مناسبتر است؟

■ مدل خطی

■ مدل درجه ۲

■ مدل معکوس یا Inverse

Variable	Linear	Inverse	Quadratic
b0 _cons	39.440284	3.9380592	51.183082
	24.44	2.65	8.87
	0.0000	0.0098	0.0000
b1 _cons	-.00600869	48893.647	-.01415806
	-11.60	12.11	-3.65
	0.0000	0.0000	0.0005
b2 _cons			1.324e-06
			2.12
			0.0378
Statistics			
N	74	74	74
r2_a	.64669141	.6661801	.6629782

Legend: b/t/p

پایان جلسه دوم وینار



عضویت در گروه تلگرامی
پرسی و پاسخ و رفع
اشکالات آکادمی

@amarafzar2011

اگر علاقمند هستید دانش و
مهارت‌های پژوهشی خود را ارتقاء
دهید، در برنامه اشتراک طلایی
آکادمی پژوهش‌افزار عضو شوید!

<https://researchware.org/gold-membership/>