

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



کارگاه آموزش SPSS

مدرس: سرکار خانم دکتر سیده سولماز طالبی

جلسه پنجم



رگرسیون خطی

- رگرسیون برای پیش بینی مورد استفاده قرار میگیرد فقط در دامنه مقادیر متغیر مستقل قابل قبول است.

مثال:

اگر داده های شما فشار خون زنان ۴۰ تا ۶۰ ساله ایرانی است و شما خط رگرسیونی برای پیش بینی فشار خون با این داده ها برآزش داده اید، باید دقت کنید این خط برای زنان خارج از این رنج سنی و یا غیر ایرانی کاربرد ندارد.



رگرسیون خطی

• اهداف رگرسیون:

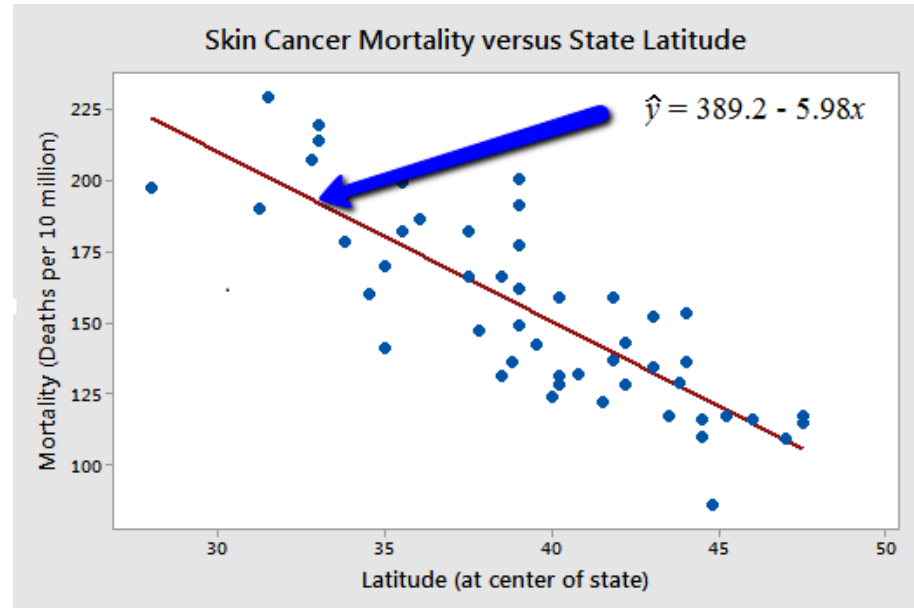
۱. آیا یک مجموعه متغیر برای پیش بینی یک پیشامد مناسبند؟
۲. هر متغیر چه میزان در پیش بینی سهم دارد.

مثال:

هدف یافتن یک خط برای پیش بینی وزن نوزاد است. عواملی که مهم به نظر میرسند: سن مادر، نوبت زایمان، شاغل بودن مادر، تغذیه مادر و جنس جنین است. در این مثال وزن نوزاد متغیر وابسته و یا پاسخ (Y) و سایر متغیرها، متغیر مستقل، توصیفی و یا پیش بینی کننده (X) نام دارند.

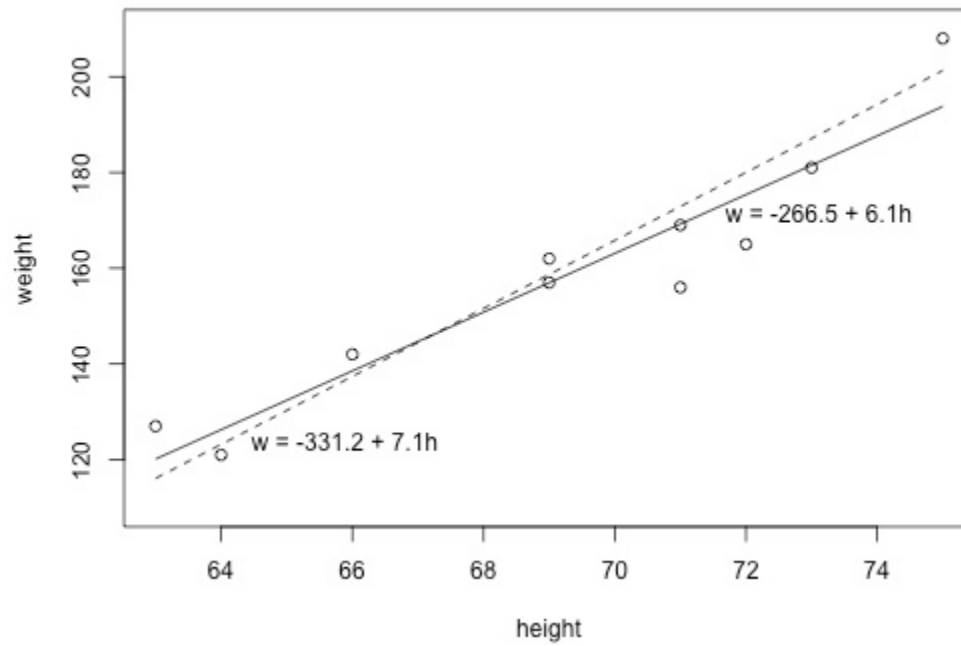


رگرسیون خطی





بهترین مدل رگرسیونی





فروض رگرسیون خطی

1. متغیر وابسته باید کمی و نرمال باشد.
2. متغیرهای مستقل میتوانند کمی و یا کیفی اسمی و رتبه ایی باشند.
3. استقلال داده ها
4. وجود رابطه خطی بین متغیر وابسته و هر متغیر مستقل.
5. ثبات واریانس
6. عدم وجود هم خطی شدید
7. عدم وجود داده پرت



رگرسیون خطی

هم خطی:

در رگرسیون چند متغیره (چندگانه یا چند متغیره) امکان دارد متغیرهای مستقلی که وارد مدل میکنیم به طریقی با هم ارتباط داشته باشند. مثل BMI با وزن و قد. در رگرسیون وارد کردن این ۳ متغیر باهم نه تنها اطلاع اضافی به ما نمیدهد بلکه برآوردها را نیز با مشکل مواجهه میکند.



برازش رگرسیون خطی

- اولین گام Data mining است. در این مرحله سنجیدن همبستگی بین متغیرها و اطمینان از تعریف درست متغیرها صورت میگیرد.

تعریف داده ها:

با توجه به هدف رگرسیون متغیرها وارد نرم افزار میشوند.

تعریف داده های اسمی:

برای تعریف اینگونه متغیرها، از Dummy variable کمک میگیریم.

مثال: برای متغیر گروه خونی که ۴ سطح دارد باید ۳ متغیر وارد مدل شود. متغیر اول گروه خونی A داشتن، متغیر دوم گروه خونی B داشتن و متغیر سوم گروه خونی AB داشتن است. گروه خونی O به عنوان گروه مرجع در نظر گرفته میشود.



رگرسیون خطی ساده (تک متغیره) و چندگانه (چند متغیره)

اگر یک متغیر مستقل وارد مدل کنیم رگرسیون تک متغیره و اگر چند متغیر مستقل وارد کنیم چند متغیر میشود. معمولا برای رگرسیون چند متغیر، ابتدا برای هر متغیر مستقل یک رگرسیون تک متغیره برازش داده میشود و اگر sig کمتر از ۲۰ درصد بود، آن متغیر را وارد رگرسیون چند متغیره میکنیم. این کار برای جلوگیری از هدر رفتن درجه آزادی صورت میگیرد.

به یاد داشته باشیم که برای هر متغیر باید حداقل ۱۰ مشاهده وجود داشته باشد. مثلا اگر میخواهید ۳ متغیر مستقل وارد مدل کنید باید حداقل ۳۰ مشاهده داشته باشید.



رگرسیون خطی

Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Reports
Descriptive Statistics
Custom Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear

	cancer	surgery	treatment
	0	1	1
	0	2	1
	0	1	1
	0	1	1

Automatic Linear Modeling...
Linear...

Linear Regression

Dependent:

Block 1 of 1
Previous Next

Independent(s):

Method: Enter

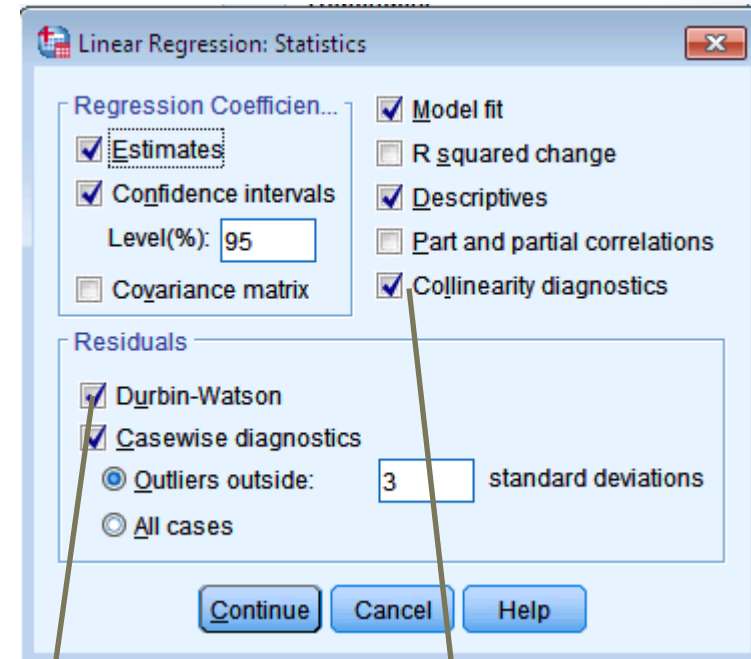
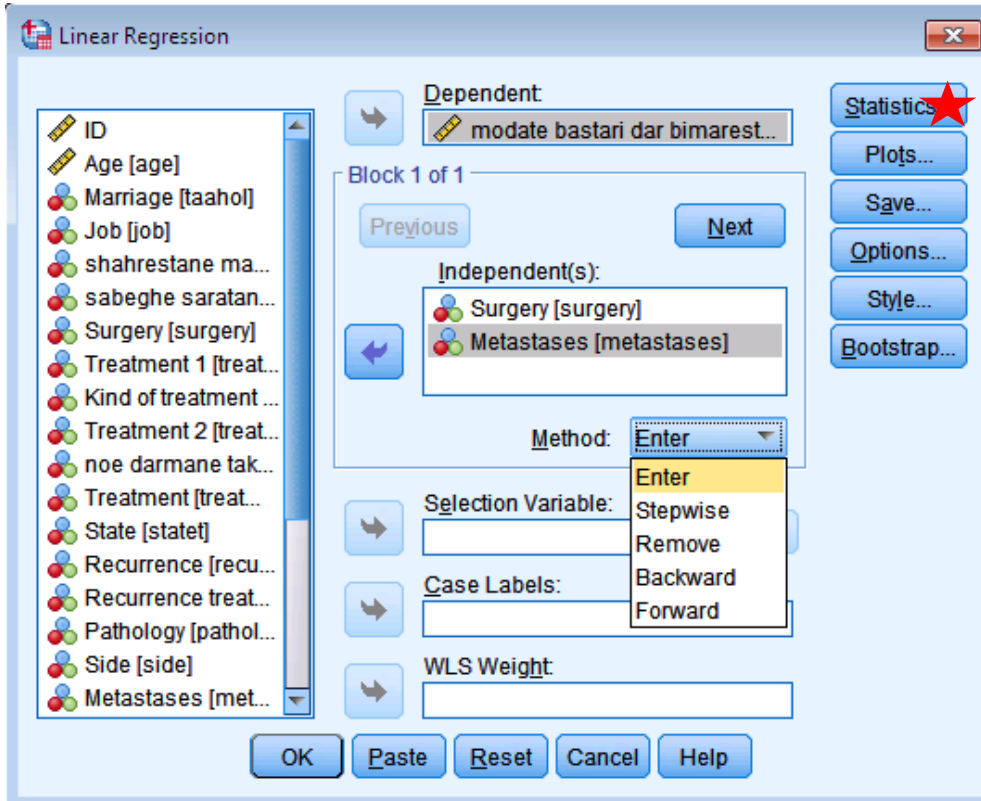
Selection Variable: Rule...

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

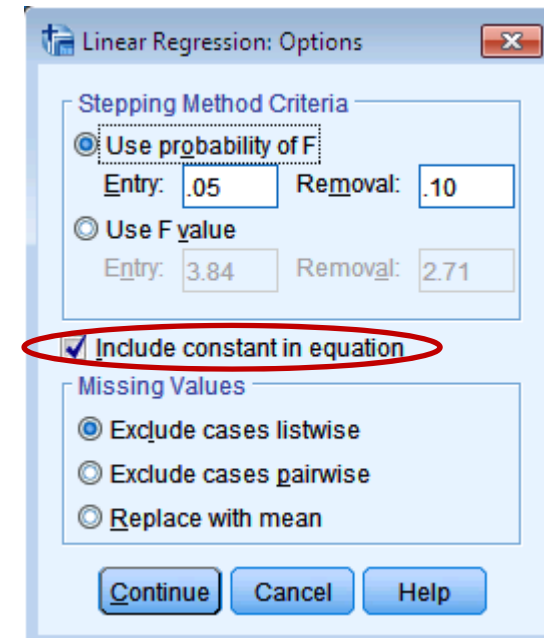
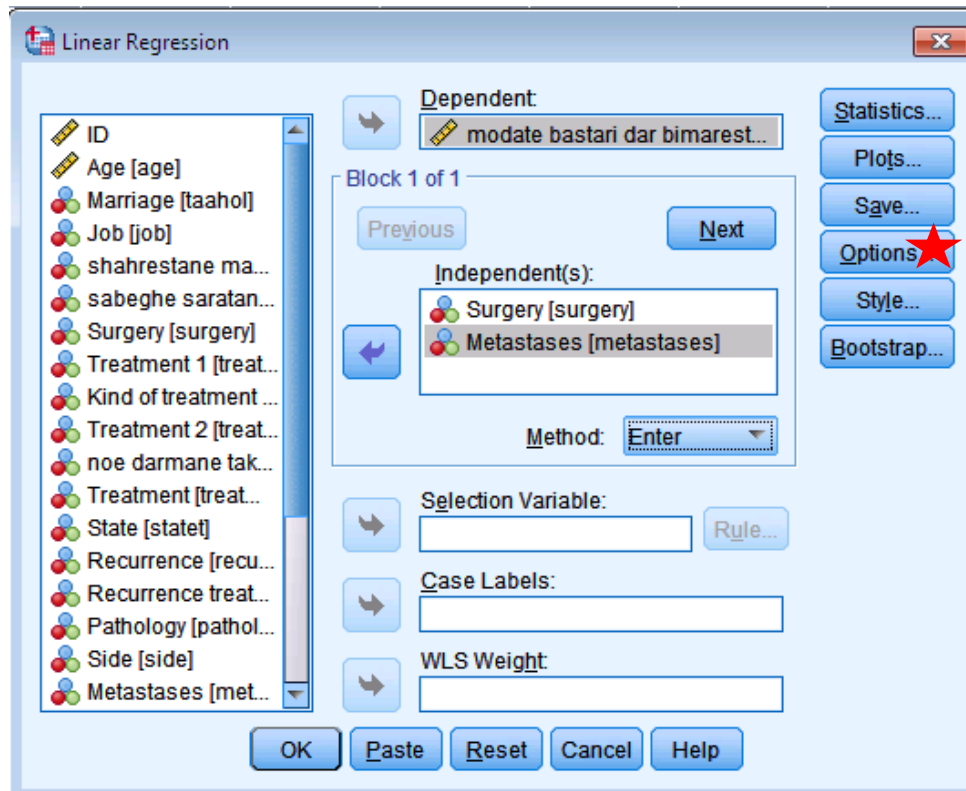
Statistics...
Plots...
Save...
Options...
Style...
Bootstrap...



استقلال باقیمانده ها را
میسنجد.

هم خطی بین متغیرها
را میسنجد.

رگرسیون خطی



Descriptive Statistics

	Mean	Std. Deviation	N
modate bastari dar bimarestan be roz	11.34	5.726	100
Surgery	1.80	1.101	100
Metastases	.38	.488	100

Correlations

		modate bastari dar bimarestan be roz	Surgery	Metastases
Pearson Correlation	modate bastari dar bimarestan be roz	1.000	-.124	-.086
	Surgery	-.124	1.000	-.083
	Metastases	-.086	-.083	1.000
Sig. (1-tailed)	modate bastari dar bimarestan be roz	.	.110	.196
	Surgery	.110	.	.207
	Metastases	.196	.207	.
N	modate bastari dar bimarestan be roz	100	100	100
	Surgery	100	100	100
	Metastases	100	100	100

همبستگی بین متغیرهای مستقل نباید زیاد باشد

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Metastases, Surgery ^b	.	Enter

a. Dependent Variable: modate bastari dar bimarestan be roz

b. All requested variables entered.

میزان همبستگی متغیر وابسته و مستقل را نشان میدهد

مقدار واریانس تبیین شده توسط مدل را نشان میدهد.

Model Summary^a

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.157 ^a	.025	.005	5.713	1.797

a. Predictors: (Constant), Metastases, Surgery

b. Dependent Variable: modate bastari dar bimarestan be roz

اگر به ۲ نزدیک باشد یعنی باقیمانده ها استقلال دارند. به ۰ نزدیک باشد یعنی همبستگی مثبت و به ۴ نزدیک باشد یعنی همبستگی منفی

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	80.248	2	40.124	1.229	.297 ^b
	Residual	3166.192	97	32.641		
	Total	3246.440	99			

a. Dependent Variable: modate bastari dar bimarestan be roz

b. Predictors: (Constant), Metastases, Surgery

فرضیه مورد آزمون صفر بودن تمام ضرایب است. اگر رد شود یعنی ضرایب متغیرها صفر نیستند و این یعنی متغیرها باید در مدل باشند.

هر متغیری که معنادار شد یعنی باید در مدل باقی بماند.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	13.008	1.219		10.674	.000	10.589	15.427		
	Surgery	-.685	.523	-.132	-1.309	.193	-1.724	.353	.993	1.007
	Metastases	-1.143	1.181	-.097	-.968	.335	-3.487	1.201	.993	1.007

a. Dependent Variable: modate bastari dar bimarestan be roz

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$



$$y = 13.008 - 0.68 \text{ surgery} - 1.14 \text{ Metastases} + \varepsilon$$

نزدیک ۱ عدم همخطی، نزدیک ۱۰ همخطی شدید



مثال

متغیر سمت درگیر یک متغیر اسمی است. ابتدا آن را به صورت درست وارد کرده و سپس میزان تاثیر متغیرهای سمت درگیر، گرید بیماری، استیج بیماری و متاستاز داشتن را بر روی تعداد غدد درگیر بدست آورید.