

مدل رگرسیون

(EXCEL)

تدوین: مرکز تحلیل آماری خوارزمی

www.kharazmi-statistics.ir

مرکز آماری خوارزمی

یکی از پرکاربردترین مباحث آماری رگرسیون است. در حقیقت تحلیل رگرسیونی فن و تکنیکی آماری برای بررسی و مدل سازی ارتباط بین متغیرها است. رگرسیون تقریباً در هر زمینه ای از جمله مهندسی، فیزیک، اقتصاد، مدیریت، علوم زیستی، بیولوژی و علوم اجتماعی برای برآورد و پیش بینی مورد نیاز است. می توان گفت تحلیل رگرسیونی، پرکاربردترین روش در بین تکنیک های آماری است. این که نرم افزار اکسل چقدر در این حیطه قابلیت دارد مشخص نیست ولی می توان از اکسل در حد ساده رگرسیون گیری انجام داد و گاهی مشکلات را از پیش رو برداشت. در ادامه روش انجام تحلیل رگرسیون به وسیله ی نرم افزار excel را بیان خواهیم کرد.

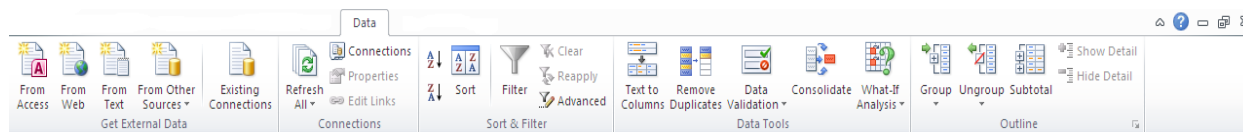
مطالبی که در این فایل بیان شده است عبارتند از:

چگونگی فعال کردن آیکون *Data Analysis*

تحلیل رگرسیون

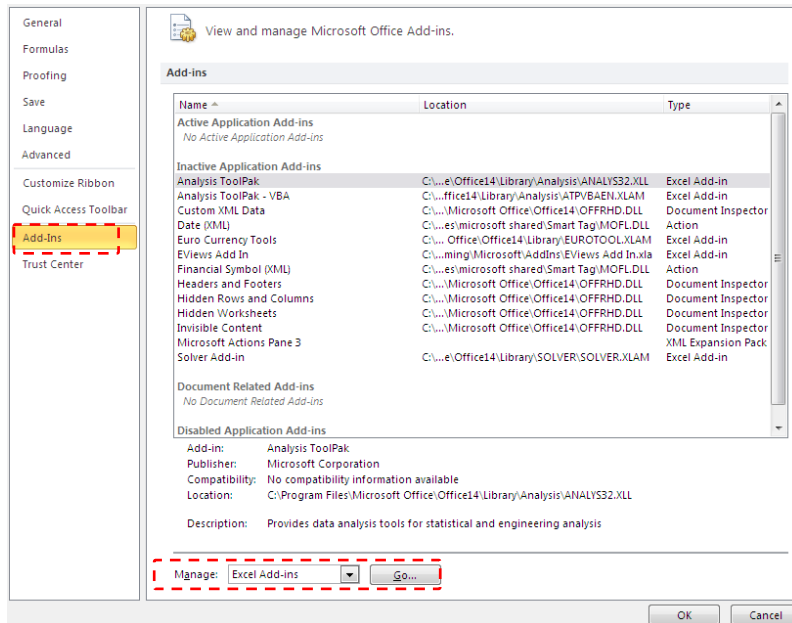
چگونگی فعال کردن آیکون *Data Analysis*

در حالت پیش فرض در نرم افزار، این آیکون روشن نیست. و نیاز است تا تنظیم خاصی انجام شود تا این امکان در صفحه ی اصلی نرم افزار فعال گردد. (لازم به ذکر است مراحل بیان شده در این متن متناسب با نرم افزار *excel2010* است). پس از نصب نرم افزار زبانه *Data* به شکل زیر است و آیکون *Data Analysis* وجود ندارد.

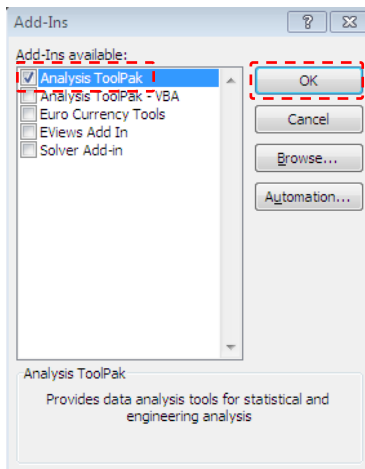


پس از باز کردن excel، برای فعال کردن *Data Analysis* مسیر زیر را طی نمایید.

File > options > Add-Ins

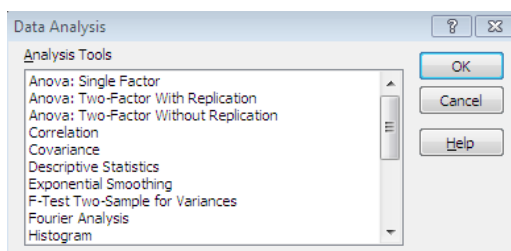
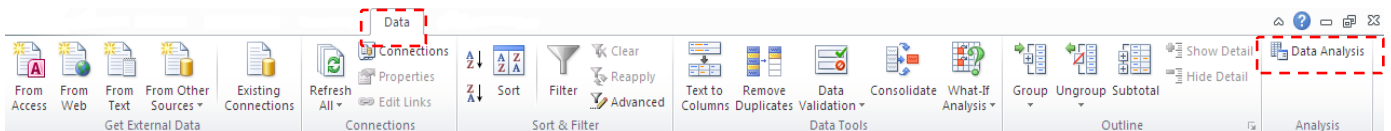


در پنجره ی باز شده بر روی گزینه ی GO در پایین پنجره کلیک کرده تا پنجره ای با عنوان Add-Ins باز گردد.



در این پنجره گزینه ی Analysis ToolPak را تیک دار می کنید. پس از آن با زدن دکمه ی OK آیکن مربوط به تحلیل داده در زبانه ی Data ظاهر می شود.

همانطور که مشاهده می کنید آیکن Data Analysis در انتهای زبانه ی Data قرار داده شده است.



با کلیک کردن بر روی آیکن Data Analysis پنجره ای مشابه تصویر روبرو نمایش داده می شود. که می توانید از تمامی امکانات آن بهره ببرید.

تحلیل رگرسیون

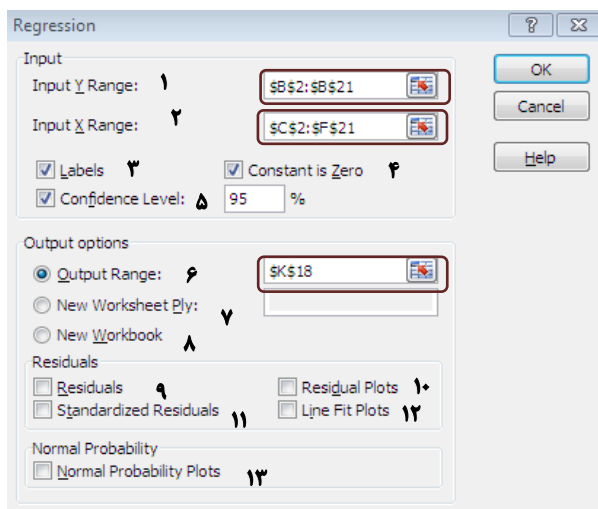
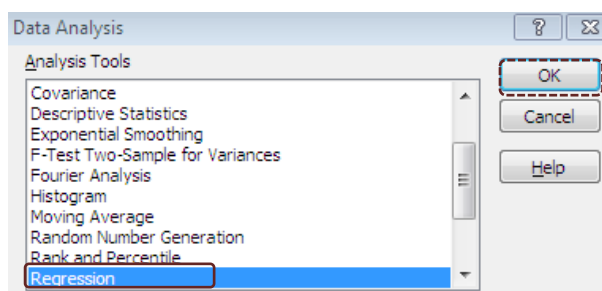
وقتی تعدادی متغیر مستقل برای یک متغیر وابسته داریم، می خواهیم ببینیم که آیا می توانیم با داشتن متغیرهای مستقل با مقادیر متفاوت متغیر وابسته را پیش بینی کنیم. این مسئله به کرات در تحلیل های علوم انسانی و سایر علوم کاربرد دارد.

Y	A	B	C	D
3	1	3	2	1
5	5	5	4	5
4	8	6	7	7
2	0	6	9	9
6	7	8	0	6
7	4	3	6	5
9	2	1	3	3
7	4	3	1	2
4	3	9	4	1
2	6	6	6	5
4	8	4	4	4
6	9	8	2	7
3	1	3	2	9
5	4	2	3	6
7	3	6	5	4
2	7	5	7	3
4	9	3	4	7
5	5	9	4	3
7	3	7	7	9

برای انجام تحلیل رگرسیون مثال زیر را در نظر بگیرید. ۵ متغیر داریم که هر یک از آنها ۱۹ مشاهده دارند تصور کنید متغیر اول (Y) متغیر وابسته و دیگر متغیرها، متغیرهای مستقل هستند. می خواهیم مدل رگرسیونی مناسب را به داده های زیر رگرس کنیم.

برای انجام رگرسیون مسیر زیر را طی نمایید.


Data > Data Analysis > Regression > OK



پس از کلیک بر روی دکمه ی OK پنجره ی روبرو باز می شود.

هر یک از فیلدها را با توجه به توضیحات ارائه شده پر نمایید. توضیحات هر یک از فیلدها در ادامه بیان شده است.

۱. این پنجره جهت ورود متغیر وابسته است. (محدوده ی مورد نظر در تصویر مربوط به داده ها با کادر قرمز مشخص شده است).
۲. این پنجره جهت ورود متغیر یا متغیرهای وابسته تا حداکثر ۱۶ متغیر می باشد. (محدوده ی مورد نظر در تصویر مربوط به داده ها با کادر سبز مشخص شده است).
۳. اگر این گزینه را علامت بزیند ردیف اول یعنی نامها به درستی جای خود قرار گرفته و اکسل خطا نمی دهد.
۴. مقدار ثابت را صفر در نظر می گیرد.
۵. سطح معنی داری با پیش بینی ۹۵٪ می باشد.
۶. آدرس مربوط به نتایج (نتایج در سلولی در همان صفحه چاپ شود).
۷. آدرس مربوط به نتایج (نتایج در برگ (sheet) دیگر در همان فایل چاپ شود).
۸. آدرس مربوط به نتایج (نتایج در فایل جدیدی چاپ شود).
۹. مقادیر باقیمانده را در خروجی نشان می دهد.
۱۰. نمودار مقادیر باقیمانده را رسم می کند.
۱۱. مقادیر باقیمانده استاندارد را در خروجی نشان می دهد.
۱۲. جهت رسم نموداری برای مقادیر پیش بینی در برابر مقادیر واقعی استفاده می شود.
۱۳. نمودار احتمال نرمال را رسم می کند.

با کلیک کردن بر روی  پنجره باریک شده و شما می توانید محدوده ی سلول هایی که می خواهید را انتخاب کنید. و با زدن دکمه ی enter محدوده انتخاب شده و دوباره پنجره به حالت قبل باز می گردد. در تصویر مربوط به داده محدوده ی داده ها با کادر سبز رنگ مشخص شده است. با انتخاب دکمه ی رادیویی columns به نرم افزار گفته می شود که گروه ها به همان صورت ستون های اولیه هستند و نه سطری. با توجه به توضیحات ارائه شده تمامی فیلدها را پر کنید. و سپس دکمه ی ok را بزیند تا نتایج چیزی مشابه تصاویر پایین نمایش داده شود.

Regression Statistics	
Multiple R	0.857249895
R Square	0.734877382
Adjusted R Square	0.615186192
Standard Error	3.011178098
Observations	19

ANOVA					
	df	SS	MS	F	Significance F
Regression	4	376.992097	94.24802424	10.3944	0.000401377
Residual	15	136.007903	9.067193535		
Total	19	513			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
1	0.208229598	0.242811248	0.857578056	0.40462	-0.309310326	0.725769523	-0.309310326	0.725769523
3	0.245620332	0.277906296	0.88382428	0.390737	-0.346722916	0.83796358	-0.346722916	0.83796358
2	0.079715892	0.295626468	0.269650726	0.791103	-0.550397009	0.709828792	-0.550397009	0.709828792
1	0.31414662	0.269275211	1.166637726	0.26157	-0.259799906	0.888093147	-0.259799906	0.888093147

جدول مربوط به محاسبه R Square، R تعدیل شده، جدول تحلیل واریانس، آماره F، درجه های آزادی، مقادیر p-value

ضرایب رگرسیونی متغیرهای مستقل ۱ و ۳

سطح معنی داری متغیرهای

خروجی بدست آمده از آزمون رگرسیون شامل سه جدول مهم است، جدول آماره های رگرسیونی، جدول آزمون آنووا و جدول ضرایب.

در جدول اول مقادیر شاخص های آماری ضریب همبستگی چندگانه (Multiple R) و ضریب تعیین (R Square) و تعداد مشاهدات گزارش شده است. تعداد مشاهدات یا تعداد نمونه برابر با ۱۹ است. ضریب همبستگی چندگانه میزان همبستگی متغیرهای مستقل با متغیر وابسته را نشان می دهد و ضریب تعیین مقدار واریانس متغیر وابسته که ناشی از متغیرهای مستقل است (یا این که متغیرهای مستقل توانسته اند چه میزان از واریانس یا تغییرات متغیر وابسته را تبیین می کنند) را نشان می دهد. بر اساس نتایج بدست آمده نتیجه می گیریم که ضریب تعیین برابر با ۷۳٪ بدست آمده است که نشان می دهد حدود ۷۳ درصد از واریانس یا تغییرات متغیر وابسته این مثال را متغیرهای مستقل تبیین می کنند. هرچه میزان ضریب تعیین بیشتر باشد نشان از تاثیر بیشتر متغیرهای مستقل بر وابسته است.

جدول آنووا (ANOVA) برازش مدل رگرسیون را نشان می دهد، بدین معنا که آیا مدل رگرسیونی مناسب است و متغیرهای مستقل به کار رفته می توانند متغیر وابسته را پیش بینی کنند یا خیر. چنانچه سطح معنی داری آزمون F (Significance F) کمتر از مقدار ۰.۰۵ باشد نشان از برازش مدل رگرسیونی دارد. در این مثال سطح معنی داری کمتر از ۰.۰۵ است (sig= P یا ۰.۰۰۴۰) است که نشان می دهد مدل رگرسیونی مناسب بوده و متغیرهای مستقل توانسته اند به طور معنی داری تغییرات متغیر وابسته را پیش بینی کنند و بر متغیر وابسته تاثیرگذار باشند.

جدول آخر، جدول ضرایب نام دارد که در آن نتیجه تاثیر هر متغیر مستقل بر وابسته ارائه شده است. دو شاخص مهم شاخص ضریب تاثیر یا ضریب رگرسیونی استاندارد شده (Coefficients) و سطح معنی داری (P-Value) است. ضریب تاثیر میزان یا بزرگی تاثیر هر متغیر مستقل بر وابسته را نشان می دهد و سطح معنی داری، معنی دار بودن آماری این تاثیر را نشان می دهد. نتایج نشان می دهد دو متغیر مستقل ۱ و ۳ با ضرایب ۰/۳۱۴ و ۰/۲۴۵، تاثیر بیشتری بر متغیر وابسته دارند و توان پیش‌بینی‌کنندگی بالاتری در مقایسه با دومتغیر مستقل دیگر دارند.

چنانچه سطح معنی داری هر متغیر مستقل کمتر از مقدار ۰/۰۵ باشد ($P < 0.05$)، نشان می دهد تاثیر آن متغیر مستقل بر متغیر وابسته از نظر آماری تایید می شود. نتایج بدست آمده نشان می دهد سطح معنی داری برای تمامی متغیرهای مستقل بیشتر از مقدار ۰/۰۵ است که بدین معناست که از نظر آماری تاثیر هیچکدام از متغیرهای مستقل بر وابسته تایید نمی شود.

نکته: علت اصلی این عدم معنی داری در تعداد مشاهدات یا نمونه هاست. دقت شود که آزمون رگرسیون نیازمند مشاهدات زیادی است که در کتب آماری و روش تحقیق، روش هایی برای محاسبه حداقل تعداد نمونه برای آزمون رگرسیون گزارش شده است. چنانچه تعداد مشاهدات بیشتر بود (مثلا بیشتر از ۱۰۰ مشاهده)، می توانستیم شاهد معنی دار بودن تاثیر برخی از متغیرهای مستقل بر وابسته باشیم.

منبع:

- اکسل (مقدمه، تحلیل واریانس، رگرسیون و آزمون میانگین)، تهیه و گرد آوری: حمید مسعودی، تابستان ۱۳۸۹، دانشگاه فردوسی مشهد

<http://www.iranresearches.ir>