

جدول تحلیل واریانس

(EXCEL)

تدوین: مرکز تحلیل آمار خوارزمی

www.kharazmi-statistics.ir

مرکز آماری خوارزمی

مقدمه:

نرم افزار EXCEL از سری نرم افزارهای اداری پرکاربرد است. یکی از امکانات موجود در نرم افزار امکان تحلیل های آماری است. تحلیل های ساده و ابتدایی و در در عین حال پرکاربرد آماری. از جمله آزمون های مختلف، مدل رگرسیون، جدول تحلیل واریانس و محاسبه ی آماره های توصیفی. در این متن بر آنیم تا چگونگی بدست آوردن جدول تحلیل واریانس در نرم افزار را بیان نماییم.

در متن پیش رو مطالب زیر بیان شده است.

چگونگی فعال کردن آیکون *Data Analysis*

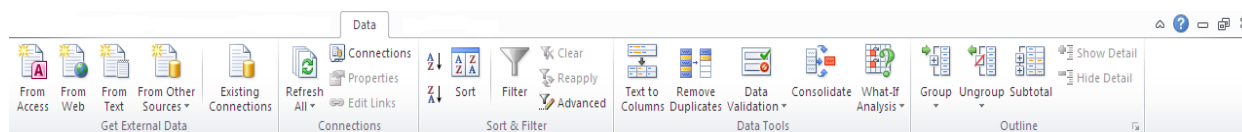
تحلیل واریانس تک فاکتوری (*ANOVA: Single Factor*)

تحلیل واریانس دو فاکتوری با تکرار (*ANOVA: Two-Factor with Replication*)

تحلیل واریانس دو فاکتوری بدون تکرار (*ANOVA: Two-Factor without Replication*)

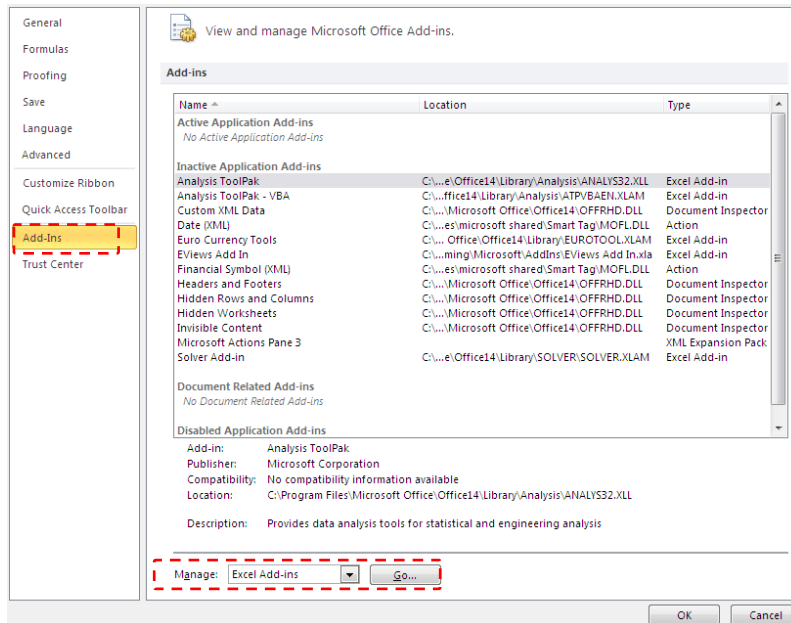
چگونگی فعال کردن آیکون *Data Analysis*

ر 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 *excel2011* است). پس از نصب نرم افزار زبانه *Data* به شکل زیر است و آیکون *Data Analysis* وجود ندارد.

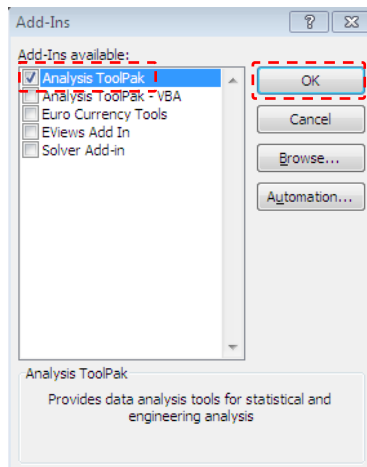


پس از باز کردن excel، برای فعال کردن *Data Analysis* مسیر زیر را طی نمایید.

File > options > Add-Ins

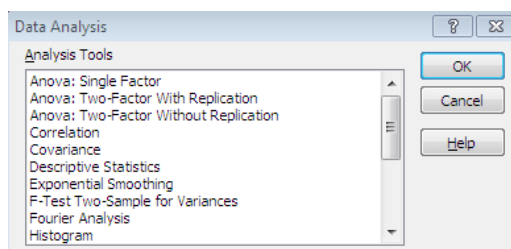
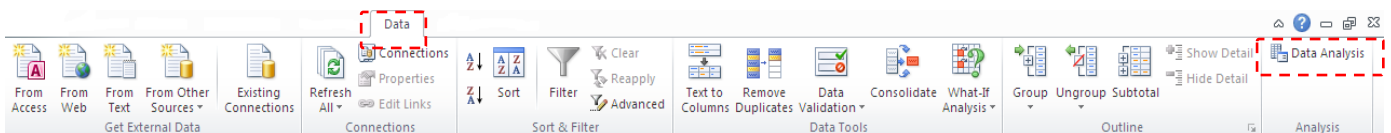


در پنجره ی باز شده بر روی گزینه ی GO در پایین پنجره کلیک کرده تا پنجره ای با عنوان Add-Ins باز گردد.



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 آماری قرار داده شده در این نرم افزار استفاده نمایید.

تحلیل واریانس تک فاکتوری (ANOVA: Single Factor)

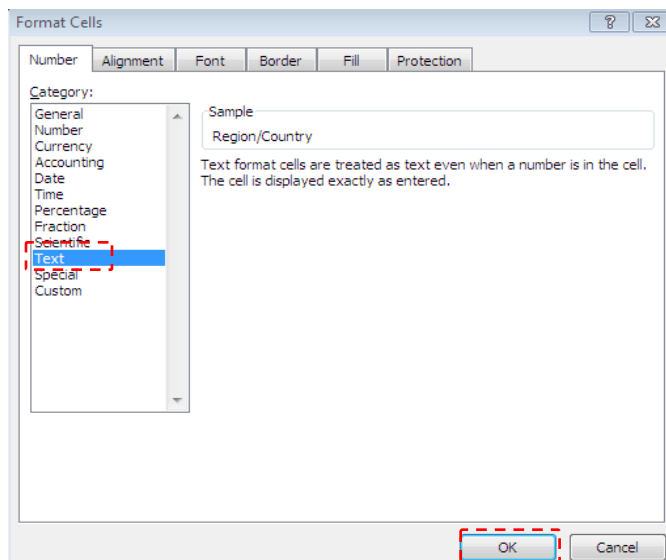
داده های انتخابی در این قسمت میزان انتشار گاز دی اکسید کربن در چند کشور منتخب است. ستون اول نام کشور و ستون های بعدی میزان انتشار با توجه به سال های انتشار است.

Region/Country	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
United States	1175.4	1157.7	1232.6	1243.5	1221.8	1265.6	1293.6	1376.4	1358.7	1342.2	1357.2
Canada	185.5	180.2	186.7	185.8	170.2	162.3	159.38	157.95	165.27	169.99	171.36
Mexico/Chile	107	135.8	118.4	120.1	119.1	135.2	135.93	137.13	144.98	150.24	157.74
Japan	174.3	192.5	210	204.7	204.5	215.4	238.95	280.74	261.89	241.62	241.68
South Korea	63.9	68.3	73.8	75.1	72.2	90.4	89.26	88.84	89.55	90.07	90.79
Australia/New Zealand	64.7	65.5	68.2	70.9	69.8	71.1	71.6	72.43	72.8	73.41	75.09
Russia	767.2	815	815.2	832.2	735.1	814.4	781.18	775.44	783.53	789.08	795.79
China	91.7	110.5	138.1	151.1	173.3	209	220.18	243.69	264.73	290.8	314.73
India	71.9	77.6	82.7	85.9	103.8	125.6	128.62	124.68	125.39	127.4	128.5
Brazil	36.3	37.7	38.7	46.1	36.5	49.2	57.39	54.3	58.86	61.27	63.19

برای انجام محاسبات ابتدا باید برای نرم افزار مشخص شود که کدام یک از ستون ها در قالب عددی و کدام یک در قالب حروفی هستند. نرم افزار excel به صورت پیش فرض قالب سلول ها را General در نظر می گیرد. برای این کار باید قالب مربوط به نام کشورها به صورت حروفی و سال ها را عددی تعریف کنیم. برای این منظور مسیر زیر را طی نمایید.

بر روی سلول Region/Country کلیک راست کنید و بر روی عبارت Format Cells... کلیک کنید. پنجره ی روبرو باز می شود. سپس در قسمت Category بر روی گزینه ی Text کلیک کنید. و دکمه ی Ok را بزنید.

این عمل برای آن انجام می شود که اگر در حالت عددی باشد و حروف وارد کنیم، هنگام محاسبات خطا می



دهد.

برای انجام تحلیل واریانس تک فاکتوری مسیر زیر را طی نمایید.

Data > Data Analysis > Anova: Single Factor > Ok

پس از انجام مراحل زیر پنجره ی پایین باز می شود.

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 **radiobutton** **columns** به نرم افزار گفته می شود که گروه ها به همان صورت ستون های ماه هستند و نه سطری.

با توجه به توضیحات ارائه شده در تصویر بالا تمامی فیلد ها را پر کنید. و سپس دکمه ی **OK** را بزنید تا نتایج چیزی مشابه تصویر پایین نمایش داده شود.

SUMMARY				
Groups	Count	Sum	Average	Variance
2005	10	2737.9	273.79	146661.8
2006	10	2840.8	284.08	146006.5
2007	10	2964.4	296.44	159507.6
2008	10	3015.4	301.54	162603.9
2009	10	2906.3	290.63	147300.8
2010	10	3138.2	313.82	160698.1
2011	10	3176.09	317.609	161919.4
2012	10	3311.6	331.16	179188.8
2013	10	3325.7	332.57	174829.4
2014	10	3336.08	333.608	171011.7
2015	10	3396.07	339.607	174226.8

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	50750.63	10	5075.063	0.031293	0.999999	1.927679
Within Groups	16055593	99	162177.7			
Total	16106344	109				

همان طور که مشاهده می کنید در جدول خلاصه (Summary)، مجموع، متوسط و واریانس هر یک از سال ها ارائه شده است و در جدول آنالیز واریانس مقدار آماره **F** و مقدار **p-value** محاسبه شده است.

تحلیل جدول:

با توجه به عدد $p\text{-value} = 0.9999$ که از

مقدار $\alpha = 0.05$ بزرگتر است پس دلیلی بر رد فرض صفر وجود ندارد و در نتیجه فرض صفر رد نمی شود. در واقع فرض صفر که معنی داری تفاوت بین میانگین سال های متفاوت بوده رد نمی شود و میانگین سال های مختلف تفاوت معنی داری با هم ندارند. پولی در صورتی که فرض صفر رد شود این آزمون توانایی آن را ندارد که مشخص کند کدام یک از سال ها میانگین متفاوتی دارند و کدام یک تفاوت معنی داری ندارند. به عبارت دیگر فرض صفر و فرض مقابل به صورت زیر است.

$$\begin{cases} H_0: \mu_1 = \mu_2 = \dots = \mu_n \\ H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_n \end{cases}$$

تحلیل واریانس دو فاکتوری با تکرار (ANOVA: Two-Factor with Replication)

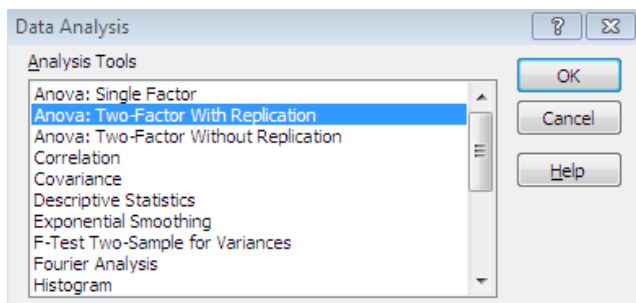
برای بدست آوردن جدول تحلیل واریانس دو فکتوری با تکرار نیاز به سری داده هایی است که چند دوره تکرار شده باشد. یعنی داروی اول برای خانوم ها و آقایان در چهار دوره تکرار شده است.

	A	B	C	D
1	pations	drug1	drug2	drug3
2	male	3	2	4
3		5	4	6
4		7	3	5
5		4	6	3
6	female	2	5	8
7		3	8	7
8		8	7	6
9		8	3	5

در این سری داده سه دارو در چهار مرحله برای گروه مذکر و مونث تست شده است. خروجی ها در این فایل به نمایش گذاشته شده است.

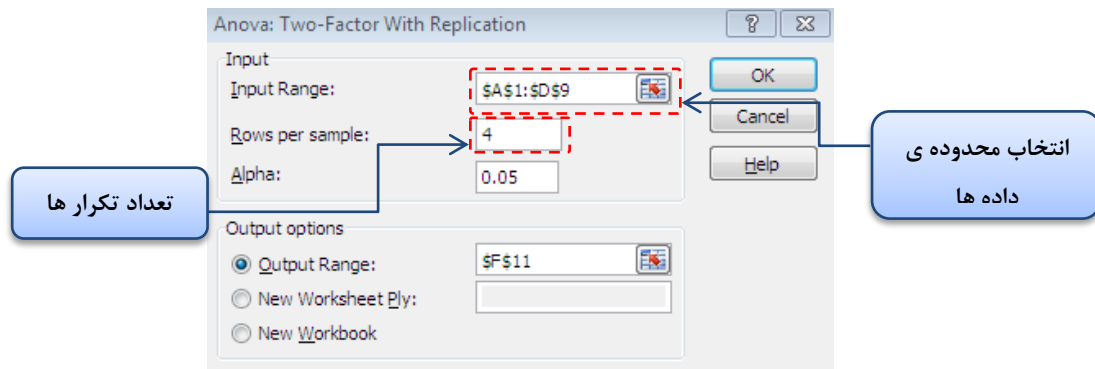
برای رسم جدول تحلیل واریانس دو فکتوری با تکرار مسیر زیر را طی نمایید.

Data > Data Analysis > Anova: Two-Factor With Replication > Ok



همانند تصویر روبرو.

پس از آن پنجره پایین ظاهری می شود.



توضیحات باقی گزینه ها (output range, new worksheet ply, new workbook, Alph) همانند حالت قبل است. محدوده ی داده های انتخاب شده با کادر بنفش رنگ مشخص شده است. لازم به ذکر است در این حالت نیازی نیست تا قالب سلول ها را تغییر دهید و آنها را به صورت Text تنظیم کنید.

تفسیر خروجی:

در خروجی نرم افزار مجموع، متوسط و واریانس هر یک از دو حالت مونث و مذکر برای تک تک داروها و به صورت کلی محاسبه شده است.

SUMMARY	drug1	drug2	drug3	Total		
<i>male</i>						
Count	4	4	4	12		
Sum	19	15	18	52		
Average	4.75	3.75	4.5	4.333333		
Variance	2.916667	2.916667	1.666667	2.242424		
<i>female</i>						
Count	4	4	4	12		
Sum	21	23	26	70		
Average	5.25	5.75	6.5	5.833333		
Variance	10.25	4.916667	1.666667	4.878788		
<i>Total</i>						
Count	8	8	8			
Sum	40	38	44			
Average	5	4.75	5.5			
Variance	5.714286	4.5	2.571429			
ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Sample	13.5	1	13.5	3.328767	0.084723	4.413873
Columns	2.333333	2	1.166667	0.287671	0.753392	3.554557
Interaction	3	2	1.5	0.369863	0.695959	3.554557
Within	73	18	4.055556			
Total	91.83333	23				

در جدول تحلیل واریانس نیز مقدار آماره F و مقدار p-value برای سه حالت بیان شده است.

۱. Sample: این آماره تفاوت میانگین میان دو گروه زنان و مردان را می سنجد. با توجه به مقدار p-value که بزرگتر از ۰.۰۵ است فرض صفر (تفاوت میان میانگین در دو گروه زن و مرد) رد نمی شود.

۲. Columns: این آماره تفاوت میان سه گروه از داروها را می سنجد. عدد p-value محاسبه شده در این

بخش نیز از مقدار $\alpha = 0.05$ بزرگتر است. در نتیجه در اینجا نیز فرض صفر رد نمی شود. Interaction: این آماره تفاوت میانگین تعاملی بین دو گروه (دارو و جنسیت) را می سنجد. در این بخش مقدار آماره برابر ۰.۶۹ است که از عدد ۰.۰۵ بزرگتر است. در نتیجه در این قسمت نیز فرض صفر رد نمی شود. در واقع می توان گفت تاثیر دارو ها برای دو گروه مردان و زنان به صورت یکسان است. و تفاوت معنی دارو در تاثیر این دو گروه جنسیت وجود ندارد.

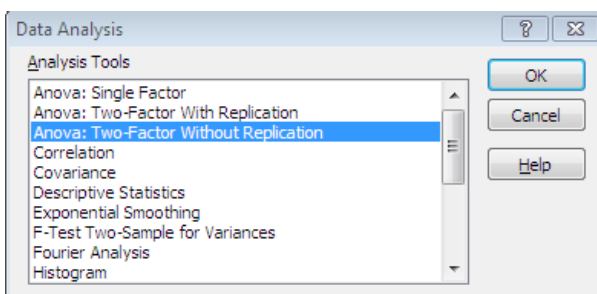
تحلیل واریانس دو فاکتوری بدون تکرار (ANOVA: Two-Factor without Replication)

برای این مرحله داده هایی همانند داده های تصویر پایین انتخاب شده است. داده ها در این مرحله مربوط به نمرات چند نفر از دانش آموزان در دوس ریاضی، شیمی و فیزیک است.

	A	B	C	D
1	student	ریاضی	شیمی	فیزیک
2	زهرا	16	16	18
3	سوسن	17	19	15
4	مریم	15	17	16
5	لاله	18	16	18
6	فاطمه	15	18	17
7	شیوا	19	17	14

همان طور که مشاهده میکنید نمرات از ۲۰ و ۶ دانش آموز آورده شده است.

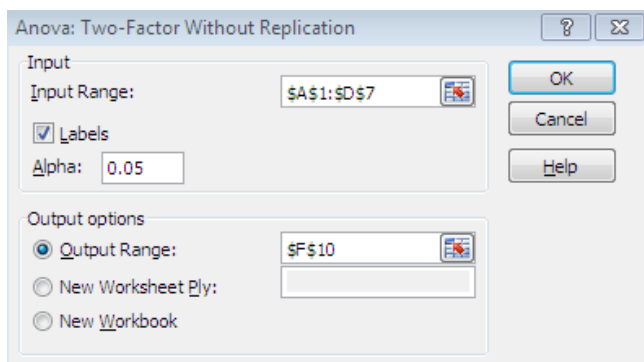
در اینجا میخواهیم اختلاف میانگین بین نمرات دوس و دانش آموزان بسنجیم.



همچون دو حالت قبلی مسیر زیر را طی نمایید تا

پنجره ی مربوط به جدول تحلیل واریانی دو فاکتوری بدون تکرار باز شود.

Data > Data Analysis > Anova: Two-Factor Without Replication > Ok



توضیحات کادر روبرو همانند حالت های قبل است. برای مطالعه دوباره [اینجا](#) را کلیک کنید.

محدوده ی داده هایی که در این مثال انتخاب کرده ایم با کادر قرمز مشخص شده است.

Anova: Two-Factor Without Replication						
SUMMARY	Count	Sum	Average	Variance		
زهرا	3	50	16.66667	1.333333		
سوسن	3	51	17	4		
مریم	3	48	16	1		
لاله	3	52	17.33333	1.333333		
فاطمه	3	50	16.66667	2.333333		
شیوا	3	50	16.66667	6.333333		
ریاضی	6	100	16.66667	2.666667		
شیمی	6	103	17.16667	1.366667		
فیزیک	6	98	16.33333	2.666667		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Rows	2.944444	5	0.588889	0.192727	0.958442	3.325835
Columns	2.111111	2	1.055556	0.345455	0.716023	4.102821
Error	30.55556	10	3.055556			
Total	35.61111	17				

پس از زدن دکمه ی ok جدول تحلیل واریانس رسم و نمایش داده می شود.

همانند دو حالت قبلی معدل هر یک از دانش آموزان و واریانس نمرات هر فرد و همچنین متوسط نمره در هر درس در جدول اول محاسبه شده است. در جدول تحلیل واریانس نیز مقدار آماره F و همچنین مقدار p-value محاسبه و نمایش داده شده است.

تفسیر خروجی: با توجه به مقدار p-value در قسمت row که بیشتر از ۰.۰۵ است پس فرض صفر یعنی برابری میانگین های نمرات دانش آموزان با یک دیگر رد نمیشود. در واقع معدل زهرا با معدل سوسن و دیگر دانش آموزان تفاوت فاحشی با یکدیگر ندارد. همچنین با توجه به عدد p-value برای columns این عدد نیز از مقدار ۰.۰۵ بزرگتر است. با توجه به این اعداد فرض صفر رد نمی شود. پس نتیجه می گیریم که میانگین نمره درس ریاضی با میانگین نمره درس فیزیک و میانگین نمره ی شیمی تفاوت فاحشی ندارد.

منبع:

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