

نوار منوی Data Analysis

Excel

اضافه کردن منوی Data Analysis

- برای نصب این افزونه که به نام Data Analysis معروف است، باید گام‌های زیر را مطابق با تصاویر مربوطه طی کنید. در این منو بسیاری از آزمون‌های و تحلیل‌های آماری قابل اجرا هستند.

- گام اول: از فهرست File دستور

Options را انتخاب کنید. پنجره‌ای

به شکل زیر ظاهر خواهد شد.

برای دسترسی به منوها و نصب

آنها در اکسل، کافی است از

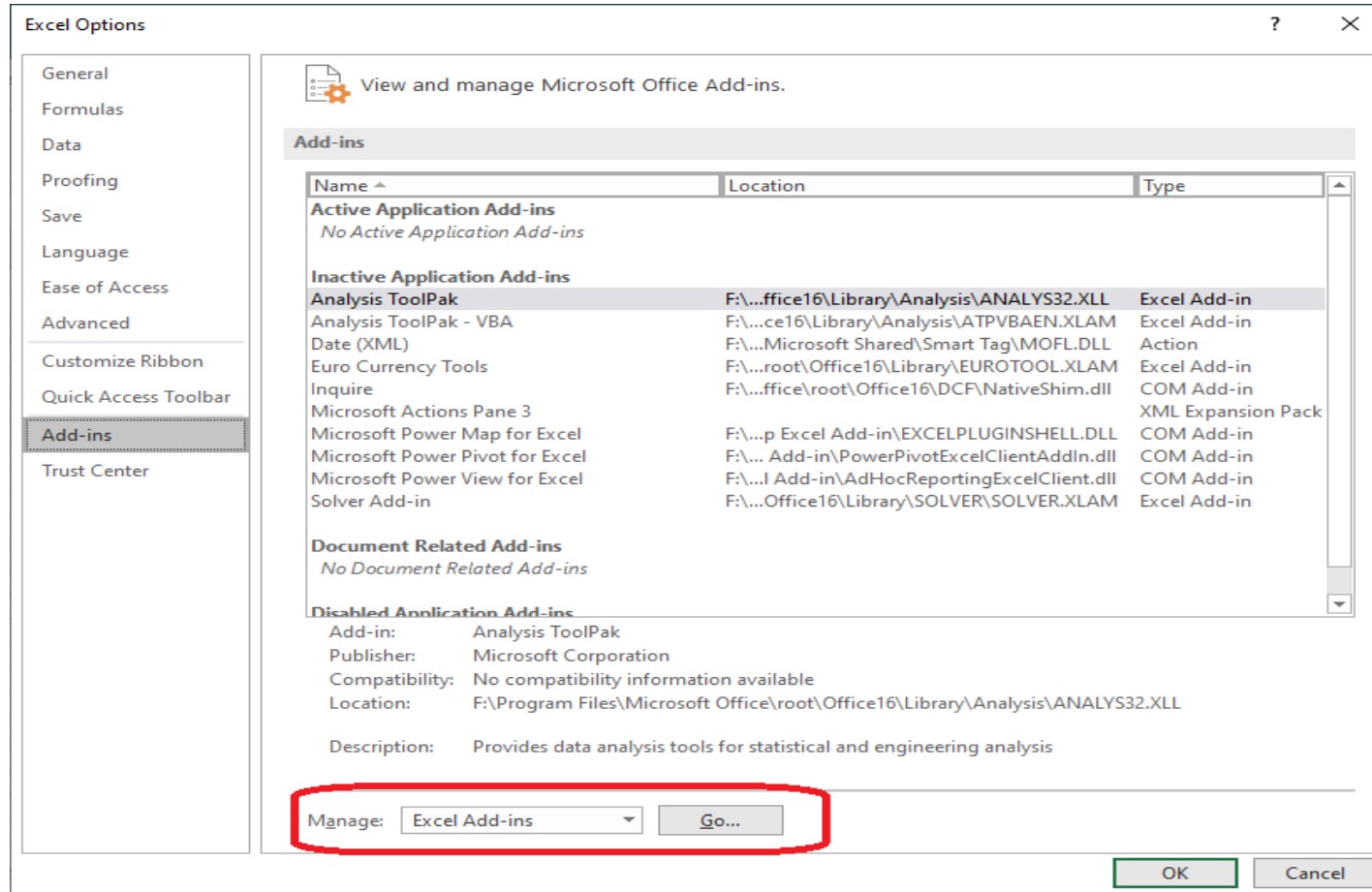
کادر سمت چپ تصویر ۱، بخش Add-ins

را انتخاب کنید. به این ترتیب در سمت

راست پنجره، لیستی از منوهای در

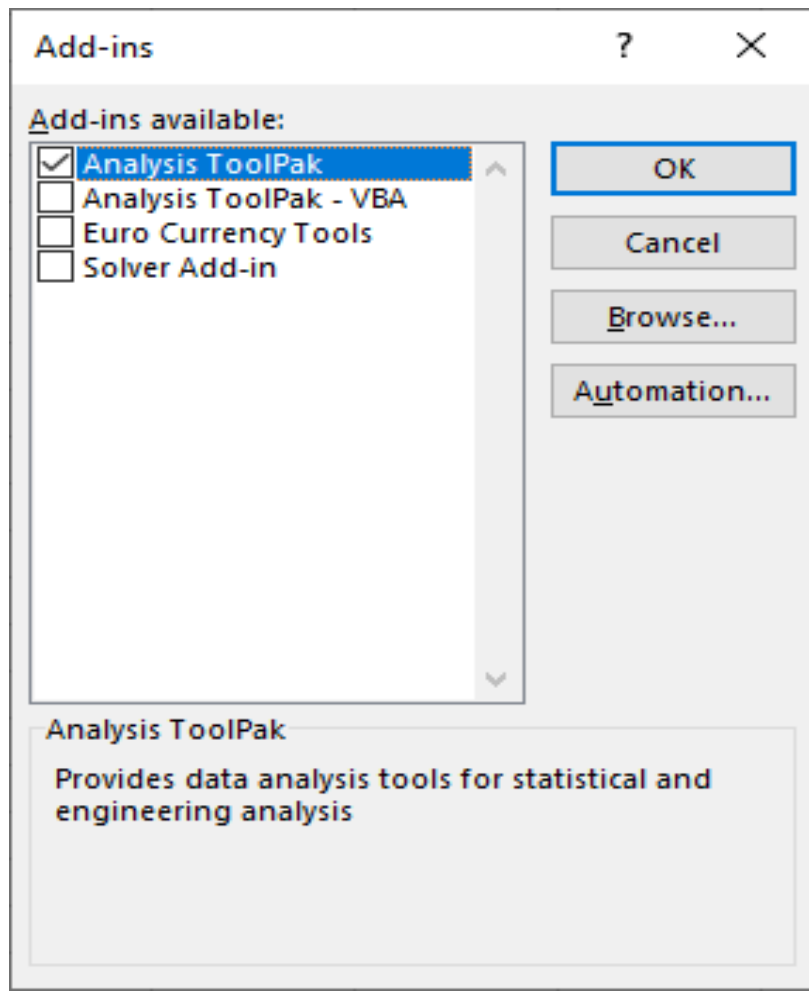
حال اجرا (Active) یا غیرفعال (Inactive)

ظاهر می‌شود.



اضافه کردن منوی Data Analysis

- اگر می‌خواهید منوهای غیرفعال را به صورت فعال در آورید، در بخش Manage گزینه Excel Add-ins را انتخاب و دکمه Go... را بزنید. به این ترتیب وارد گام دوم می‌شوید.



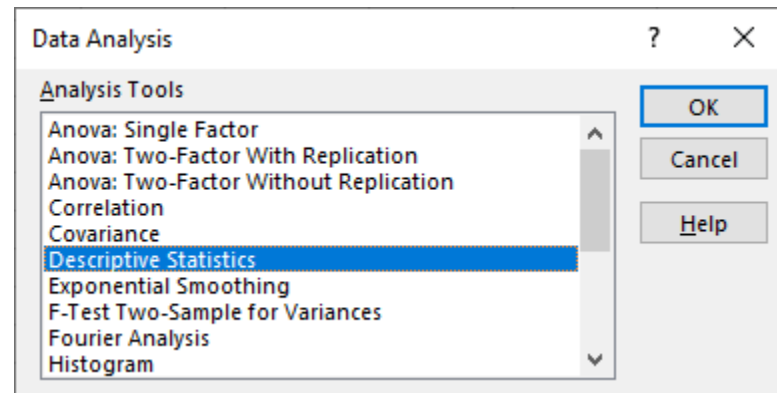
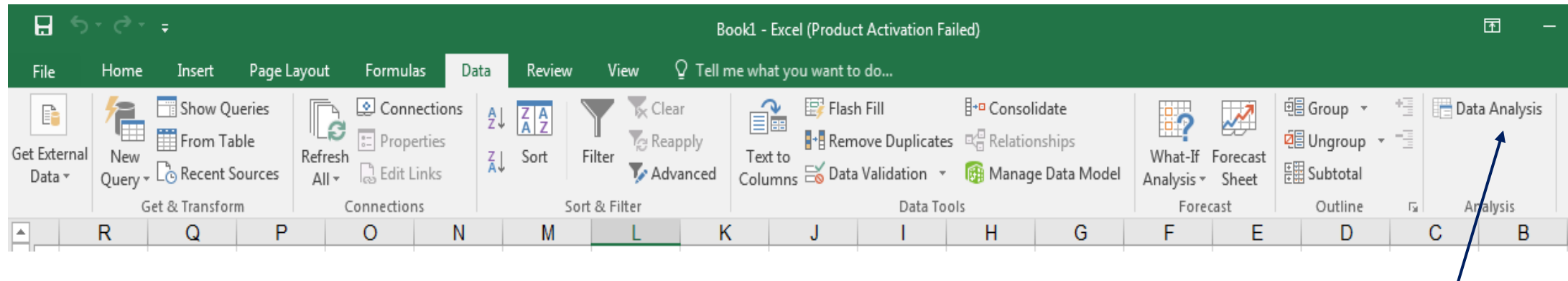
- گام دوم: در تصویر روبرو،

برای فعال سازی منوی محاسبات آماری

در اکسل گزینه اول یعنی Analysis ToolPak را انتخاب کرده‌ایم. با فشردن دکمه OK عملیات بارگذاری این منوی صورت خواهد گرفت.

اضافه کردن منوی Data Analysis

- با اجرای این دو گام، در منوی Data زیر منوی جدیدی به نام Data Analysis ظاهر می‌شود. به این ترتیب با انتخاب این دکمه، پنجره‌ای مطابق با تصویر زیر نمایان می‌شود که می‌توانید انواع محاسبات و تحلیل‌های آماری را از داخل آن انتخاب نمایید.



گزینه Descriptive Statistics

- برای استفاده از این گزینه فرض کنید در یک ستون نمرات 10 دانشجو را بصورت زیر وارد کرده ایم:

The screenshot shows an Excel spreadsheet with the following data in column A:

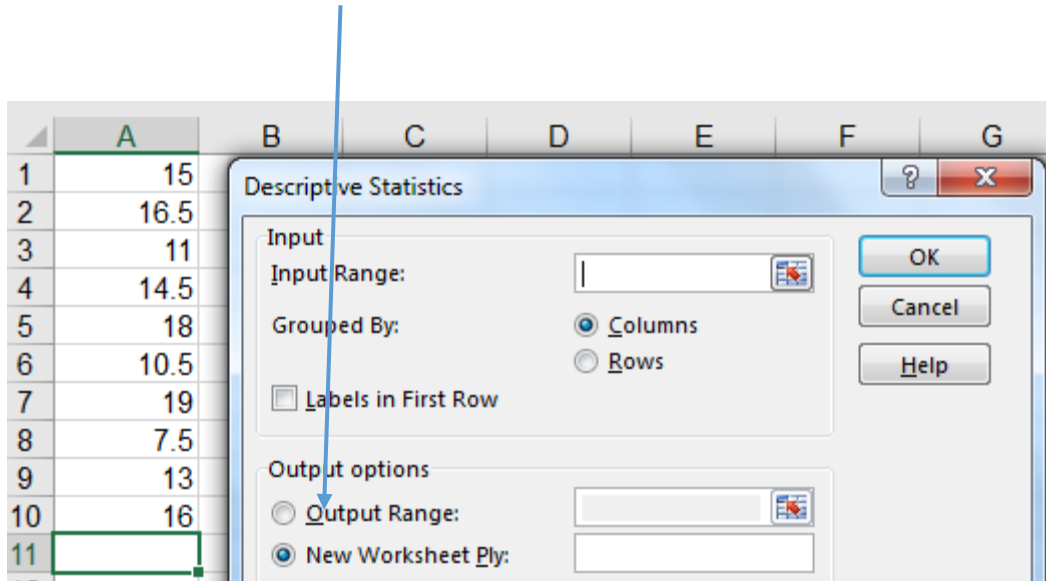
	A
1	15
2	16.5
3	11
4	14.5
5	18
6	10.5
7	19
8	7.5
9	13
10	16
11	
12	
13	
14	
15	
16	
17	
18	
19	

The 'Descriptive Statistics' dialog box is open, showing the following settings:

- Input Range:** (Empty)
- Grouped By:** ☒ Columns, ☐ Rows
- ☐ Labels in First Row
- Output options:**
 - ☐ Output Range: (Empty)
 - ☒ New Worksheet Ply: (Empty)
 - ☐ New Workbook
 - ☐ Summary statistics
 - ☐ Confidence Level for Mean: 95 %
 - ☐ Kth Largest: 1
 - ☐ Kth Smallest: 1

- در قسمت Input Range ، ناحیه‌ای از کاربرد که باید شاخص‌های آمار توصیفی از آن استخراج شود، تعیین شده است. برای مثال ما این قسمت شامل ناحیه \$A\$1:\$A\$10 است. از آنجایی که گزینه Columns انتخاب شده است، متغیرها به صورت ستونی از مجموعه داده گرفته می‌شوند.
- نکته: با فعال‌سازی گزینه Label in first row ، اکسل متوجه می‌شود که سطر اول در ستون A ، شامل نام متغیر است و نباید از آن در محاسبات استفاده کند.

گزینه Descriptive Statistics



- بخش بعدی مربوط به تعیین محل خروجی محاسبات است. اگر گزینه Output Range را انتخاب کنید، باید ناحیه‌ای از کاربرگ را برای نمایش اطلاعات خروجی مشخص نمایید. با گزینه New Worksheet Ply یک کاربرگ جدید به منظور قرارگیری خروجی تعیین می‌شود. با گزینه New Workbook هم، کارپوشه‌ای جدید (یک فایل جدید اکسلی) برای نمایش اطلاعات جدول‌های آمار توصیفی در نظر گرفته خواهد شد.

- با انتخاب گزینه Summary statistics شاخص‌های مرکزی و پراکندگی به‌مراه شاخص چولگی و برجستگی در خروجی شما نمایش داده خواهد شد.

Descriptive Statistics گزینه خروجی

Column1	
14.1	Mean
1.139688066	Standard Error
14.75	Median
#N/A	Mode
3.604010112	Standard Deviation
12.98888889	Sample Variance
0.440255163	Kurtosis
0.49497544	Skewness
11.5	Range
7.5	Minimum
19	Maximum
141	Sum
10	Count

آزمون فرضیه

Statistical Hypothesis Testing

آماره آزمون تی یک نمونه ای با واریانس نامعلوم

- آزمون فرض تساوی میانگین یک جامعه نرمال با واریانس نامعلوم (S^2)
- این آزمون مشابه آزمون قبلی می باشد با این تفاوت که در اینجا بجای واریانس جامعه از واریانس نمونه استفاده می کنیم. در نتیجه توزیع ملاک آزمون بجای توزیع نرمال دارای توزیع t-student با درجه آزادی $df=n-1$ خواهد بود.

$$t_c = \frac{(\bar{X} - \mu)}{S / \sqrt{n}} \sim t(df)$$

- در آزمون دو دامنه ملاک آزمون را با عدد بحرانی $t_{1-\alpha/2}(df)$ و در آزمون یک دامنه با عدد بحرانی $t_{1-\alpha}(df)$ مقایسه می کنیم.
- اگر ملاک آزمون از عدد بحرانی بزرگتر باشد فرض صفر را رد می کنیم.

سوال:

- اگر متوسط فشارخون مردان 40-50 ساله 14 باشد و محققى بخواهد که تاثیر یک رژیم غذایی خاص را بر روی فشار خون مطالعه نماید. فشار خون 15 نفر از این مردان را که از این رژیم تبعیت می کردند را اندازه گیری کرده است و نتایج زیر به دست آمده است:
- 15 و 14 و 11 و 16 و 14 و 12 و 11 و 10 و 15 و 16 و 17 و 15 و 12 و 11 و 13
- حال هدف این است که با اطمینان 95 درصد آیا رژیم غذایی تاثیری بر فشار خون دارد؟

$$t_c = \frac{(\bar{X} - \mu)}{\frac{S}{\sqrt{n}}} = \frac{13/47 - 14}{2/2\sqrt{15}} = -0.94$$

- در آزمون دو دامنه عدد بحرانی بر اساس نتیجه فرض صفر را نمی توان رد کرد.
- این آزمون به دلیل سادگی در اکسل وجود ندارند.

$t_{0.975}(14)$

t Table

cum. prob	<i>t</i> _{.50}	<i>t</i> _{.75}	<i>t</i> _{.80}	<i>t</i> _{.85}	<i>t</i> _{.90}	<i>t</i> _{.95}	<i>t</i> _{.975}	<i>t</i> _{.99}	<i>t</i> _{.995}	<i>t</i> _{.999}	<i>t</i> _{.9995}
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
Z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%

آزمون فرضیه در مطالعات قبل و بعد (ویژگی کمی)

$$\begin{cases} H_0: \mu_D = 0 \\ H_1: \mu_D \neq 0 \end{cases}$$

- آزمون فرض تساوی میانگین در داده های زوجی

- آزمون آماری

$$t_c = \frac{\bar{D}}{S_D / \sqrt{n}} \sim t_{(n-1)}$$

- ملاک آزمون

- **مثال:** فرض کنید یک نوع رژیم غذایی تحت بررسی باشد. برای بررسی تاثیر این رژیم غذایی بر روی وزن، آزمایشی را به صورت زیر ترتیب می دهیم:

- یک نمونه n تایی از افراد انتخاب می کنیم. وزن همه آنها را قبل و بعد از رژیم ثبت می کنیم.

- توجه می کنیم که این دو مجموعه مستقل از هم نبوده و کاملاً **وابسته** به هم می باشند این نوع نمونه ها را نمونه های **جور شده** (زوج شده) می نامند که مبنای روش های اندازه گیری تکراری هستند.

- در این صورت دیگر نمی توان از روش های آزمون فرض بالا استفاده کرد و به روش جدیدی نیاز است که با ساختار وابسته داده ها همساز باشد.

مثال:

• در مثال رژیم غذایی داده های زیر حاصل شده است:

اختلاف -D	بعد	قبل
4	111	115
1	97	98
3	100	103
2	106	108
1	94	95
4	78	81
4	71	75
2	98	100
3	84	87
1	100	101
5	92	97

$$t_c = \frac{\bar{D}}{S_D / \sqrt{n}} = \frac{2.63}{1.36 / \sqrt{11}} = 6.42$$

$$t_{1-\alpha/2}^{n-1} = t_{0.975}^{10} = 2.228$$

• ملاک آزمون را با عدد بحرانی

• چون آماره آزمون عدد 6.42 از عدد بحرانی به دست آمده از جدول توزیع تی بزرگتر می باشد پس فرض بی اثر بودن رژیم غذایی رد می شود.

تمرین: در یک مطالعه برای بررسی رژیم کم نمک بر روی فشارخون افراد 60-70 داده های روبرو به دست آمده است؟ آیا این رژیم باعث کاهش فشار خون شده است؟ با اطمینان 95 درصد.

بعد از رژیم	قبل از رژیم
16	16.5
13.5	14
15	15.5
14.5	15
16	17
14	14.5
12.5	13
13	13.5
14	14
15	15.5
14	13.5
13.5	14
16	16.5
14.5	15
15	15.5
13.5	13

$$t_c = \frac{\bar{D}}{S_D / \sqrt{n}} = \frac{0.375}{0.3873 / 4} = \frac{0.375}{0.09683} = 3.87$$

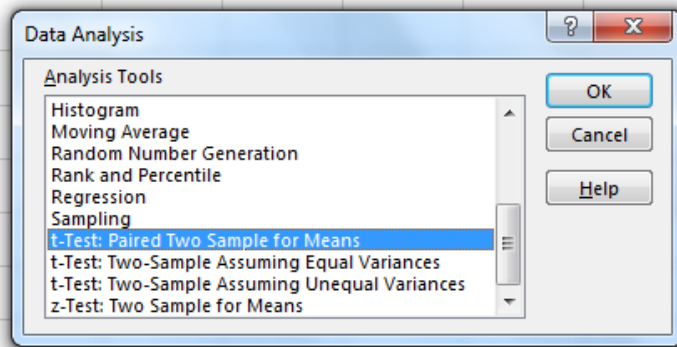
$$t_{1-\alpha}^{n-1} = t_{0.95}^{15} = 1.75$$

چون آماره آزمون 3.87 از عدد بحرانی 1.75 بزرگتر هست پس فرض صفر را رد می کنیم و می پذیریم که رژیم غذایی مورد نظر باعث گکاهش فشار خون شده است.

آزمون تی زوجی در اکسل

- برای انجام این آزمون در اکسل پس از ورود داده ها به اکسل بصورت رو به رو داده ها از منوی Data زیر منوی Data Analysis را باز می کنیم. در پنجره باز شده گزینه T-Test Paired Two Sample for Means را انتخاب و دکمه OK را می زنیم.

	قبل	بعد
1		
2	115	111
3	98	97
4	103	100
5	108	106
6	95	94
7	81	78
8	75	71
9	100	98
10	87	84
11	101	100
12	97	92



آزمون تی زوجی در اکسل

- در پنجره باز شده آدرس ستون مقادیر قبل و بعد را وارد می کنیم. و سپس دکمه OK را می زنیم.

	A	B
1	قبل	بعد
2	115	111
3	98	97
4	103	100
5	108	106
6	95	94
7	81	78
8	75	71
9	100	98
10	87	84
11	101	100
12	97	92

t-Test: Paired Two Sample for Means

Input

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

☐ Labels

Alpha:

Output options

☐ Output Range:

☒ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

خروجی آزمون تی زوجی

t-Test: Paired Two Sample for Means		
Variable 2	Variable 1	
93.72727273	96.36363636	Mean
141.8181818	134.6545455	Variance
11	11	Observations
	0.993625723	Pearson Correlation
	0	Hypothesized Mean Difference
	10	df
	6.42070792	t Stat
	3.81386E-05	P(T<=t) one-tail
	1.812461123	t Critical one-tail
	7.62772E-05	P(T<=t) two-tail
	2.228138852	t Critical two-tail

آماره آزمون

مقدار عددی توزیع تی

چون آماره آزمون از عدد بحرانی به دست آمده از جدول توزیع تی بزرگتر می باشد پس فرض بی اثر بودن رژیم غذایی رد می شود.

آزمون F

- آزمون فرض تساوی واریانس دو جامعه

- اگر یک نمونه تصادفی از جامعه اول و یک نمونه تصادفی از جامعه دوم داشته باشیم، آنگاه با استفاده از مشاهدات، $\bar{X}_1$ را به عنوان برآوردی از μ_1 ، S_1^2 را به عنوان برآوردی از σ_1^2 ، $\bar{X}_2$ را به عنوان برآوردی از μ_2 و S_2^2 را به عنوان برآوردی از σ_2^2 محاسبه می‌کنیم.

- حال تصمیم داریم دستگاه آزمون فرض زیر را در سطح اطمینان $1-\alpha$ آزمون کنیم:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

- ملاک آزمون بصورت $F_c = \frac{S_1^2}{S_2^2} \approx F_{(1-\alpha)(n_1-1, n_2-1)}$ می باشد که دارای توزیع F می باشد.

- برای این منظور از توزیع F و آماره مربوطه استفاده می‌کنیم.

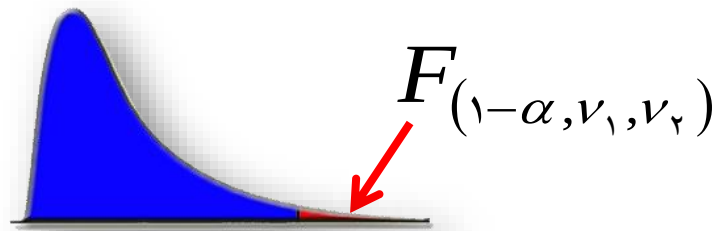
- توجه: واریانس بزرگتر باید در صورت کسر قرار گیرد.

- تحت فرض صفر انتظار می‌رود مقدار آماری F برابر ۱ باشد.

- برای مثال فرض کنید جراحی می‌خواهد بداند که آیا واریانس استحکام دو نوع نخ جراحی را یکی بداند؟ بدین منظور یک $n_1=16$ و $n_2=11$ از دو نوع نخ انتخاب کرده و با یک روش معین میزان استحکام نخ‌ها اندازه‌گیری می‌شود و مشاهده می‌شود که $S_1^2 = 4$ و $S_2^2 = 5$ می‌باشد.

- ملاک آزمون $F_c = \frac{S_1^2}{S_2^2} = \frac{5}{4} = 1.25$ به دست می‌آید که با در نظر خطای ۰/۰۵

- با مقایسه با عدد بحرانی $F_{0.95}^{(10,15)} = 2.54$ فرض برابری واریانس در دو گروه پذیرفته می‌شود.



منحنی توزیع F

صورت کسر: d1

F Distribution: Critical Values of F (5% significance level)

مخرج کسر: d2

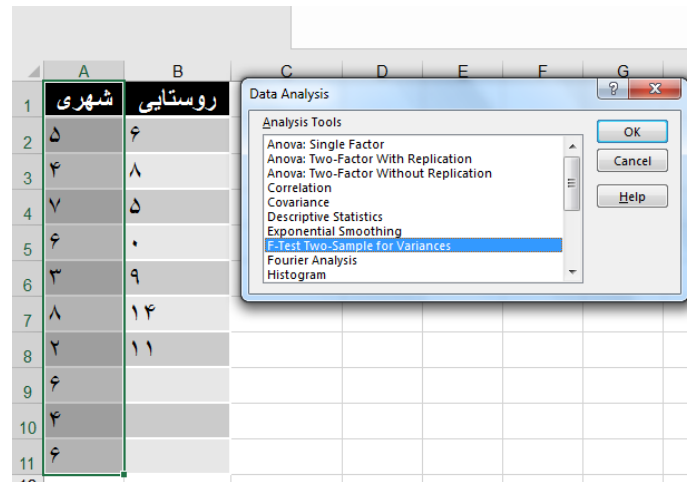
ν_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
ν_2															
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.36	246.46	247.32	248.01
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.44	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	8.67	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	5.82	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	4.58	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	3.90	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.51	2.48	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33

آزمون F : فرض برابری واریانس ها

• مثال: در مطالعه ای میزان شاخص DMFT در کودکان شهری و روستایی مورد بررسی قرار گرفته است. نتایج در جدول روبه رو مشاهده می شود. آیا میزان پراکندگی در دو گروه یکسان است.

• پس از وارد داده به اکسل و باز کردن زیر منوی Data Analysis در پنجره باز شده گزینه F-Test Two Sample for Variances را کلیک کرده و دکمه OK را می زنیم تا پنجره مربوطه باز شود.

روستایی	شهری
6	5
8	4
5	7
0	6
9	3
14	8
11	2
	6
	4
	6



آزمون F : فرض برابری واریانس ها

- در پنجره باز شده آدرس خانه هایی که داده ها در آن وارد شده را وارد می کنیم. دقت کنید چون پراکندگی در گروه روستایی بیشتر است در جای خالی اول آدرس ستون روستایی و در جای خالی دو آدرس ستون شهری را وارد می کنیم. سپس دکمه OK را می زنیم. خروجی در یک صفحه گسترده به صورت زیر ارائه می شود.
- چون آماره آزمون از عدد بحرانی بزرگتر می باشد پس فرض برابری واریانس ها پذیرفته نیست.

	A	B
1	شهری	روستایی
2	۵	۶
3	۴	۸
4	۷	۵
5	۶	۰
6	۳	۹
7	۸	۱۴
8	۲	۱۱
9	۶	
10	۴	
11	۶	

F-Test Two-Sample for Variances

Input

Variable 1 Range:

Variable 2 Range:

☐ Labels

Alpha:

Output options

☐ Output Range:

☒ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

روستایی
شهری

آماره آزمون

عدد بحرانی

F-Test Two-Sample for Variances		
Variable 2	Variable 1	
5.1	7.571428571	Mean
3.433333	20.28571429	Variance
10	7	Observations
9	6	df
	5.908460472	F
	0.009422223	P(F<=f) one-tail
	3.373753647	F Critical one-tail

آزمون فرض تساوی میانگین در دو جامعه با واریانس معلوم

- دستگاه آزمون فرض دوطرفه به صورت زیر باشد:

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

- اگر یک نمونه تصادفی از جامعه اول و یک نمونه تصادفی از جامعه دوم باشند، آنگاه با استفاده از مشاهدات، $\bar{X}_1$ را به عنوان برآوردی از μ_1 و $\bar{X}_2$ را به عنوان برآوردی از μ_2 محاسبه می‌کنیم.

- در صورت صحیح بودن فرض صفر انتظار می‌رود مقدار Z به مقدار نزدیک باشد.

$$Z_c = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

مثال:

- فرض کنید محقق می خواهد بداند که آیا میانگین هموگلوبین در مردان سیگاری و غیر سیگاری یکسان است؟ ضمناً از تجارب گذشته می دانیم توزیع میزان هموگلوبین در هر دو گروه دارای توزیع نرمال با واریانس معلوم 4 می باشد. برای انجام آزمون فرضیه یک نمونه 12 تایی از مردان سیگاری و یک نمونه 8 تایی از مردان غیرسیگاری انتخاب کرده و مشاهده می شود که میانگین هموگلوبین در مردان سیگاری 15/5 و در مردان غیر سیگاری 13/5 است.

• ملاک آزمون به صورت زیر به دست می آید:

$$Z_c = \frac{(15.5 - 13.5)}{\sqrt{\frac{4}{12} + \frac{4}{8}}} = 2.19$$

- چون ملاک آزمون از عدد بحرانی 1/96 بزرگتر است، فرض برابری میزان هموگلوبین در دو گروه رد می شود.

•

آزمون فرض تساوی میانگین در دو جامعه با واریانس نامعلوم ولی مساوی

• دستگاه آزمون فرض دوطرفه به صورت زیر باشد:

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

• در صورتی که واریانس ها نامعلوم باشد از واریانس های نمونه برای برآورد واریانس جامعه استفاده می کنیم. با پذیرش پیش فرض برابری واریانس آماره آزمون بصورت زیر به دست می آید:

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1 + n_2 - 2)}$$

• بطوری که

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

سوال:

- می خواهیم فرض برابری متوسط قد را در زنان دارای تالاسمی ماژور و مینور را با هم مقایسه کنیم، مطالعه ای انجام داده ایم که نتایج حاصل از آن بصورت زیر است. (فرض برابری واریانسها را آزمون می کنیم)

مینور	ماژور
158	145
160	155
156	152
162	150
150	147
155	156
161	153
164	154
168	144
170	159

- فرض آماری: برابری متوسط قد در دو گروه
- ابتدا فرض برابری واریانس ها را چک می کنیم

$$F_c = \frac{S_1^2}{S_2^2} = \frac{35.79}{24.28} = 1.47$$

- در مقایسه با عدد بحرانی $F_{0.975}^{(9,9)} = 4.03$
- فرض برابری واریانس ها رد نمی شود.

آزمون مقایسه میانگین در دو گروه (واریانس نا معلوم و مساوی)

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{9 \times 35.79 + 9 \times 24.28}{18} = 30.035$$

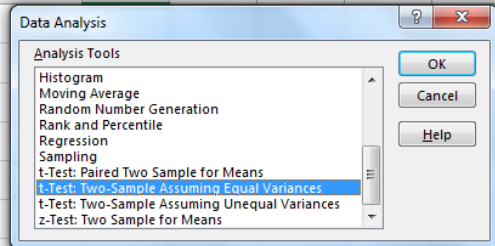
$$t_c = \frac{(\bar{X}_1 - \bar{X}_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(160.3 - 151.5)}{5.48 \times \sqrt{\frac{1}{10} + \frac{1}{10}}} = 3.59$$

$$t_{0.975}^{18} = 2.101$$

فرض برابری میانگین در دو گروه رد می شود

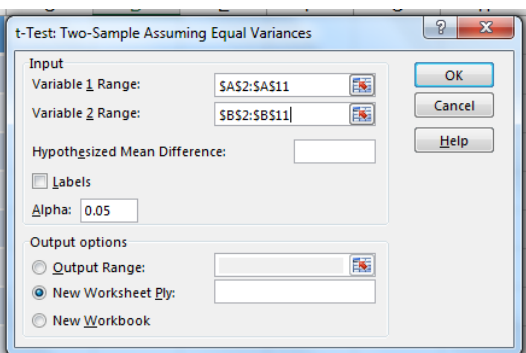
آزمون مقایسه میانگین در دو گروه (واریانس نا معلوم و مساوی) با اکسل

	A	B
1	مینور	ماژور
2	158	145
3	160	155
4	156	152
5	162	150
6	150	147
7	155	156
8	161	153
9	164	154
10	168	144
11	170	159



- پس از ورود داده به اکسل و باز کردن زیر منوی Data Analysis در پنجره باز شده گزینه T-Test: Two Sample Assuming Equal Variance را کلیک کرده و دکمه OK را می زنیم تا پنجره مربوطه باز شود. در پنجره باز شده آدرس خانه هایی که داده ها در آن وارد شده را وارد می کنیم و دکمه OK را می زنیم. خروجی بصورت زیر ظاهر می شود. اختلاف جزئی مشاهده شده در آماره آزمون به دلیل خطای رند کردن می باشد.

	A	B
1	مینور	ماژور
2	158	145
3	160	155
4	156	152
5	162	150
6	150	147
7	155	156
8	161	153
9	164	154
10	168	144
11	170	159



t-Test: Two-Sample Assuming Equal Variances		
Variable 2	Variable 1	
151.5	160.4	Mean
24.27778	36.488889	Variance
10	10	Observations
	30.383333	Pooled Variance
	0	Hypothesized Mean Difference
	18	df
	3.6104165	t Stat
	0.0010002	P(T<=t) one-tail
	1.7340636	t Critical one-tail
	0.0020003	P(T<=t) two-tail
	2.100922	t Critical two-tail

آماره آزمون

عدد بحرانی

آزمون مقایسه میانگین در دو گروه (واریانس نا معلوم و نامساوی)

کبد چرب	سالم
266	244
317	156
440	198
325	208
389	221
414	166
279	193
309	166
461	
371	
287	

- وقتی در مقایسه میانگین در دو گروه فرض برابری واریانس ها رد می شود دیگر از فرمول قبلی نمی توان استفاده نمود و چون آماره آزمون در حالت کمی پیچیده است فقط چگونگی انجام آن در اکسل گفته می شود.
- مثال: در مطالعه ای خواهیم متوسط میزان قند کلسترول در دو از افراد با کبد چرب و سالم را با اطمینان 95 درصد آزمون نماییم؟

فرض برابری واریانس

طبق شکل زیر داده ها را وارد کرده و آزمون برابری واریانس ها را انتخاب می کنیم و در جاهای خالی آدرس سلول های مورد نظر را وارد می کنیم و دکمه OK را کلیک می کنیم.

چون آماره آزمون (4.98) از عدد بحرانی (3.64) بزرگتر شده است پس فرض برابری واریانس ها رد می شود.

	A	B	C	D	E	F	G	H
	کبد چرب	سالم						
1								
2	۲۶۶	۲۴۴						
3	۳۱۷	۱۵۶						
4	۴۴۰	۱۹۸						
5	۳۲۵	۲۰۸						
6	۳۸۹	۲۲۱						
7	۴۱۴	۱۶۶						
8	۲۷۹	۱۹۳						
9	۳۰۹	۱۶۶						
10	۴۶۱							
11	۳۷۱							
12	۲۸۷							

F-Test Two-Sample for Variances

Input

Variable 1 Range: SAS2:SAS12

Variable 2 Range: BS2:BS9

☐ Labels

Alpha: 0.05

Output options

☐ Output Range:

☒ New Worksheet Ply:

☐ New Workbook

OK

Cancel

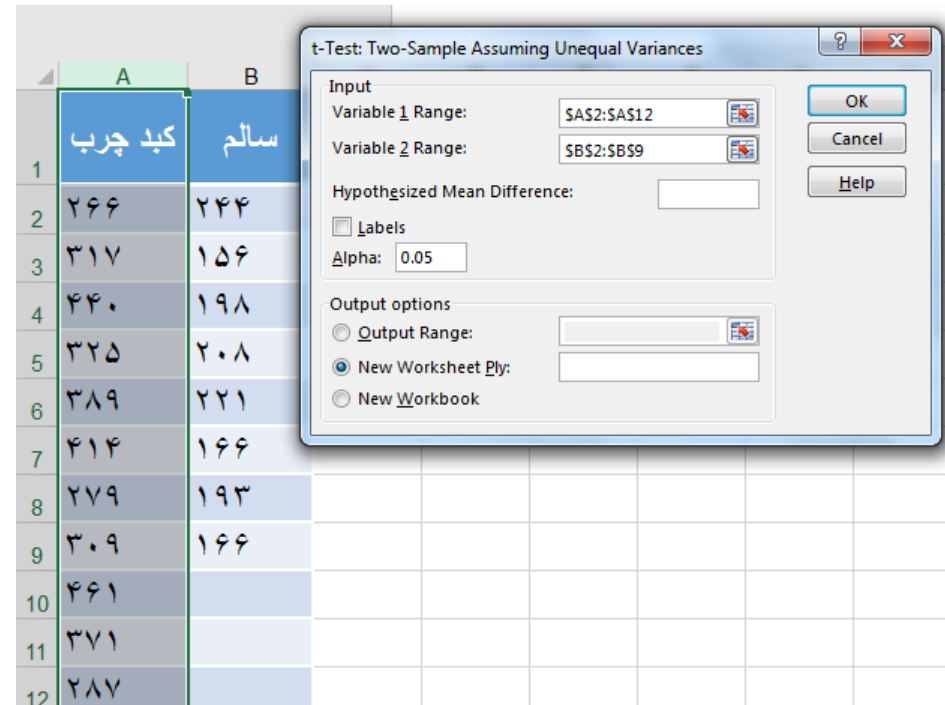
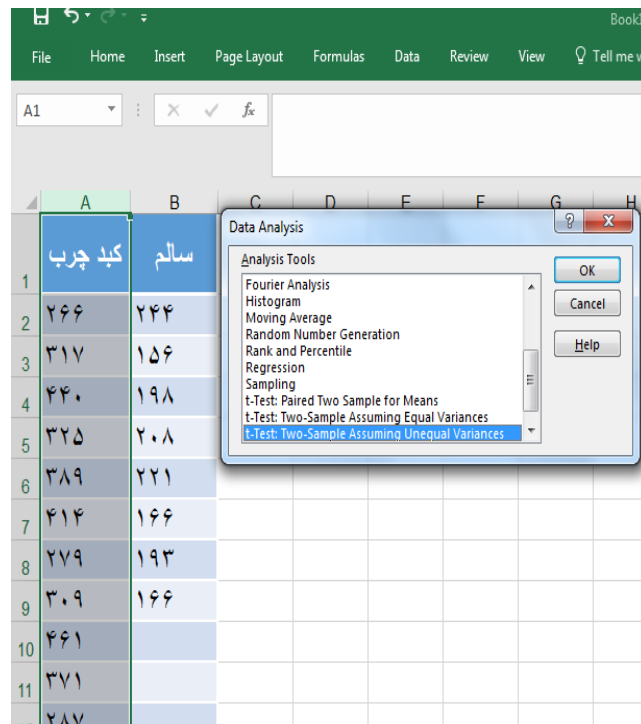
Help

F-Test Two-Sample for Variances		
Variable 2	Variable 1	
194	350.7273	Mean
922	4593.418	Variance
8	11	Observations
7	10	df
	4.982015	F
	0.02213	P(F<=f) one-tail
	3.636523	F Critical one-tail

آزمون مقایسه میانگین در دو گروه (واریانس نا معلوم و نامساوی)

- برای انجام آزمون تی مستقل با عدم فرض برابری واریانس ها در منوی Data Analysis گزینه T-Test: Two Sample Assuming Unequal Variances را انتخاب می کنید و دکمه OK را می زنیم.

در پنجره باز شده آدرس های مورد نظر را وارد کرده و باز OK می کنیم.



خروجی اکسل مقایسه میانگین در دو گروه (واریانس نا معلوم و نامساوی)

چون آماره آزمون 6.79 از عدد بحرانی 2.13 بزرگتر است پس با اطمینان 95 درصد فرض برابری میانگین کلسترول در دو گروه رد می شود.

t-Test: Two-Sample Assuming Unequal Variances		
Variable 2	Variable 1	
194	350.7273	Mean
922	4593.418	Variance
8	11	Observations
	0	Hypothesized Mean Difference
	15	df
	6.789673	t Stat
	3.05E-06	P(T<=t) one-tail
	1.75305	t Critical one-tail
	6.1E-06	P(T<=t) two-tail
	2.13145	t Critical two-tail

آماره آزمون

عدد بحرانی

آنالیز واریانس یکطرفه (مقایسه میانگین یک ویژگی کمی در بیش از دو گروه)

- مثال: در مطالعه متوسط مدت زمان بهبود (هفته) بر اساس سه نوع پانسمان برای زخم دیابت مورد نظر می باشد. اطلاعات در جدول روبه رو نشان داده شده است. آیا متوسط مدت زمان بهبودی با اطمینان 95 در سه گروه با هم برابر است؟
- چون بیش از دو گروه داریم دیگر نمی توان از آزمون تی مستقل استفاده کرد.
- در این مثال حجم نمونه در هر سه گروه برابر است ولی در عمل ممکن است به دلایل مختلف تعداد افراد در گروهها با هم برابر نباشد ولی اکسل فقط توانایی انجام آزمون با حجم برابر را دارد.

بیاتین	نقره	کامفیل
16	7	8
11	11	13
17	9	15
15	8	10

آزمون آنالیز واریانس را مشخص می کنیم.

The screenshot displays an Excel spreadsheet with data for three groups: کامفیل (Kamfil), نقره (Nargh), and بیاتین (Biatin). The data is organized in columns A, B, and C, with rows 1 through 5 containing numerical values. The Data Analysis dialog box is open, showing the 'Analysis Tools' list. The 'Anova: Single Factor' option is selected, indicated by an orange arrow pointing from the text above. The dialog box also includes buttons for OK, Cancel, and Help.

	A	B	C
1	کامفیل	نقره	بیاتین
2	۸	۷	۱۶
3	۱۳	۱۱	۱۱
4	۱۵	۹	۱۷
5	۱۰	۸	۱۵

Data Analysis

Analysis Tools

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram

OK Cancel Help

ستونی بودن داده ها مشخص می شود. آدرس داده ها در این خانه داده می شود.

	A	B	C
1	کامفیل	نقره	بیاتین
2	۸	۷	۱۶
3	۱۳	۱۱	۱۱
4	۱۵	۹	۱۷
5	۱۰	۸	۱۵
6			
7			
8			
9			

Anova: Single Factor

Input
Input Range: 

Grouped By:
☒ Columns
☐ Rows

☐ Labels in First Row
Alpha:

Output options
☐ Output Range: 
☒ New Worksheet Ply:
☐ New Workbook

OK
Cancel
Help

خروجی آنالیز واریانس در اکسل

چون آماره آزمون از عدد بحرانی بزرگتر شده است با اطمینان 95 درصد، حداقل متوسط زمان بهبود در یک گروه با بقیه تفاوت معنی داری دارد.

						Anova: Single Factor
						SUMMARY
		Variance	Average	Sum	Count	Groups
		9.666667	11.5	46	4	Column 1
		2.916667	8.75	35	4	Column 2
		6.916667	14.75	59	4	Column 3
						ANOVA
F crit	P-value	F	MS	df	SS	Source of Variation
4.256495	0.026882	5.551282	36.08333	2	72.16667	Between Groups
			6.5	9	58.5	Within Groups
				11	130.6667	Total

عدد بحرانی

آماره آزمون