

آمار توصیفی در Minitab

(آمار توصیفی داده های گسسته و پیوسته)

تدوین: مرکز آماری خوارزمی

www.kharazmi-statistics.ir

مرکز آماری خوارزمی

آمار توصیفی در MINITAB

مقدمه:

ابتدایی ترین تحلیل های و محاسبات آماری در پروژه ها بدست آوردن آمارهای توصیفی داده های مدنظر است. تعیین فراوانی، میانگین، میانه و ... از جمله مواردی است که در آمار توصیفی بدست می آید و در تحلیل های ابتدایی به کاربران اطلاعات خوبی را می دهد. در این فایل تلاش شده تا با طرح چند مثال به بیان آمار توصیفی به وسیله ی نرم افزار MINITAB بپردازیم.

آمار توصیفی برای داده های کیفی:

مثال: سی نفر از مردان و زنان که دارای ناراحتی چشم بودند مورد معاینه یک چشم پزشک قرار گرفته اند. نتایج معاینه به صورت کد های آستیگمات (A)، آب مروارید (C)، آب سیاه (G)، صدمه دیده (I)، لوچ (S) و سایر ناراحتی ها (O) به همراه جنسیت آنها در زیر فهرست شده است.

ناراحتی چشم	
A	A
O	I
G	A
A	C
O	I
G	S
S	C
C	A
A	A
O	S
S	S
S	C
A	A
A	A
I	S
S	S
O	C
جنسیت	
f	f
f	m
m	m
f	f
m	f
f	m
m	m
m	f
f	f
m	f
m	m
f	m
f	f
m	f
m	f
f	f
f	f

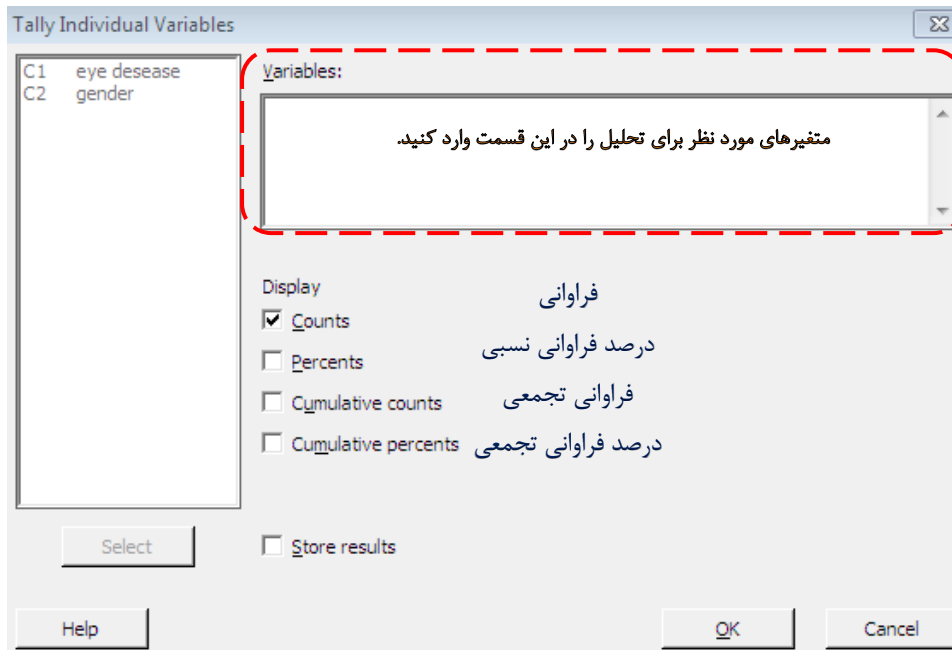
↓	C1-T	C2-T
	eye disease	gender
1	c	f
2	o	f
3	s	f
4	s	m
5	i	m
6	a	f
7	a	m

در ابتدا داده ها را به صورت شکل روبرو در پنجره ی Worksheet وارد نمایید.

برای بدست آوردن جدول فراوانی برای (متغیرهای کیفی) جنسیت و بیماری چشم مسیر زیر را طی نمایید.

Stat > Tables > Tally Individual Variables

با طی کردن مسیر بیان شده پنجره ی پایین باز می شود. دو متغیر C1,C2 را به مستطیل Variables منتقل نمایید(بر روی متغیرها کلیک کرده سپس دکمه ی select بزنید، متغیرها به مستطیل روبرو منتقل می شود). با زدن تیک در هر یک از گزینه های display خروجی مورد نظر نمایش داده خواهد شد. برای ایجاد یک جدول فراوانی کامل باید هر چهار گزینه را تیک بزنید.



gr	Count	Percent	CumCnt	CumPct
g	2	6.67	17	56.67
i	3	10.00	20	66.67
o	4	13.33	24	80.00
s	6	20.00	30	100.00
N=	30			

gender	Count	Percent	CumCnt	CumPct
f	17	56.67	17	56.67
m	13	43.33	30	100.00
N=	30			

پس از ok کردن پنجره session باز می شود. در این پنجره جدول فراوانی مد نظر رسم شده است. و تمام موارد خواسته شده(در اینجا چهار گزینه تیک دار شده اند) نمایش داده شده است. همان طور که مشاهده می کنید فراوانی ها به تفکیک بیماری، جنسیت و همچنین درصدها نمایش داده شده است.

در ادامه به بررسی های کمی پیوسته می پردازیم.

مثال: در مجله ی PC در مقاله ای، برای بررسی سالانه سرعت چاپگرها، سرعت ۳۱۲ نوع مختلف چاپگر بیان شده است. یک نمونه تصادفی ۴۶ تایی از چاپگرهای سوزنی انتخاب شده است. داده های زیر سرعت این

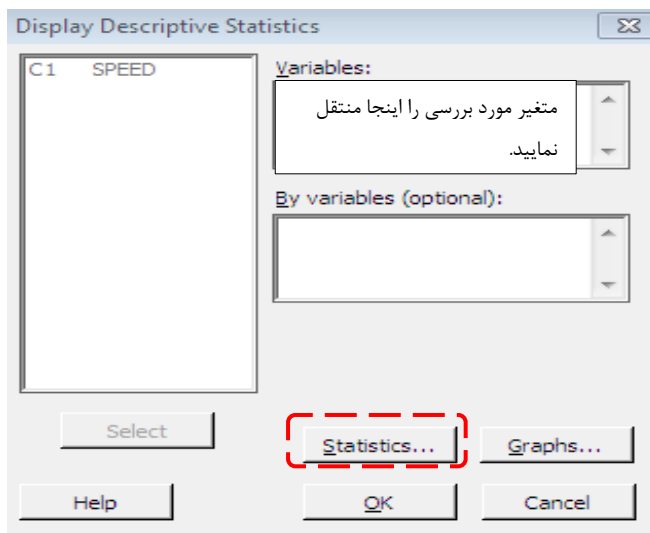
چاپگرها را بر حسب کارکتر در ثانیه می دهد. در این مثال می خواهیم جدول فراوانی را برای داده های پیوسته رسم نماییم.

۱۳۸,۱۵۴,۱۲۸,۱۰۲,۳۷,۸۶,۲۴۶,۱۹۵,۱۴۵,۹۱,۱۰۰,۱۱۱,۲۰۰,۹۹,۵۳,۶۰,۱۲۴,۱۳۳,۱۷۳,۴۶,۱۱۵,۹۳,
 ۱۳۳,۸۹,۱۴۷,۱۷۳,۱۱۱,۱۵۷,۱۵,۱۵۰,۱۰۰,۱۳۱,۱۳۶,۱۵۴,۱۱۱,۱۳۸,۱۶۳,۱۰۵,۱۴۱,۱۳۸,۱۶۳,۱۰۵,۱۴۱
 ۱۷۳,۱۵۷,۱۵۰

جهت رسم جدول داده های فراوانی برای داده های پیوسته ابتدا باید داده ها را به داده های گسسته تبدیل کنیم که برای این منظور می بایست دامنه تغییرات، طول طبقات و تعداد آنها را برای داده های پیوسته اصلی به دست آوریم.

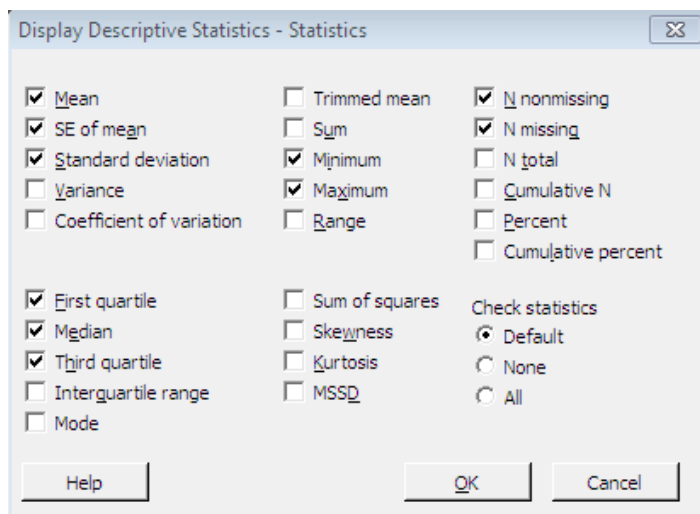
برای تعیین بیشینه، کمینه، و دامنه تغییرات می بایست مسیر زیر را طی نمایید.

Stat > Basic Statistics > Display Descriptive Statistics



سپس پنجره ی روبرو باز می شود.

متغیر مربوط به سرعت را به مستطیل Variable منتقل نمایید. سپس برای انتخاب عملیات های آماری دکمه ی Statistics را بزنید تا پنجره ی پایین باز شود.



با زدن تیک هر گزینه محاسبات لازم توسط نرم افزار انجام شده و در پنجره ی نهایی نمایش داده می شود. در زیر گزینه های پنجره روبرو توضیح داده شده است.

Mean = میانگین، SE of mean = استاندارد میانگین، Standard deviation = اختلاف استاندارد، Variance = واریانس، Coefficient of variation = ضریب تغییرات، Sum = مجموع، Maximum = بیشینه، Minimum = کمینه، Range = دامنه تغییرات، N nonmissing = تعداد داده های موجود، N missing = N missing، N total = تعداد کل داده ها، Percent = درصد فراوانی نسبی، Cumulative percent = درصد فراوانی تجمعی، First quartile = چارک اول، Median = میانه، چارک دوم، Third quartile = چارک سوم، Interquartile range = دامنه میان چارکی، Mode = مد، Sum of squares = مجموع مربعات، Skewness = چولگی، Kurtosis = کشیدگی

با انتخاب گزینه هایی از پنجره بالا و زدن دکمه ی ok خروجی زیر نمایش داده می شود.

Descriptive Statistics: speed

Variable	N	N*	Mean	CoefVar	Minimum	Maximum	Range
speed	46	0	126.30	34.71	15.00	246.00	231.00

حال با مشخص شدن دامنه تغییرات از فرمول $2^k \geq n$ تعداد طبقات را تعیین می کنیم که در آن n حجم نمونه است و k تعداد طبقات، اولین عدد طبیعی است که در نامساوی فوق صدق می کند. سپس از $c = range/k$ فاصله طبقات را تعیین می کنیم.

$$231 = \text{دامنه تغییرات} \quad \longrightarrow \quad 231/6 = 38.5 = \text{طول طبقات}$$

$6 = \text{تعداد طبقات}$

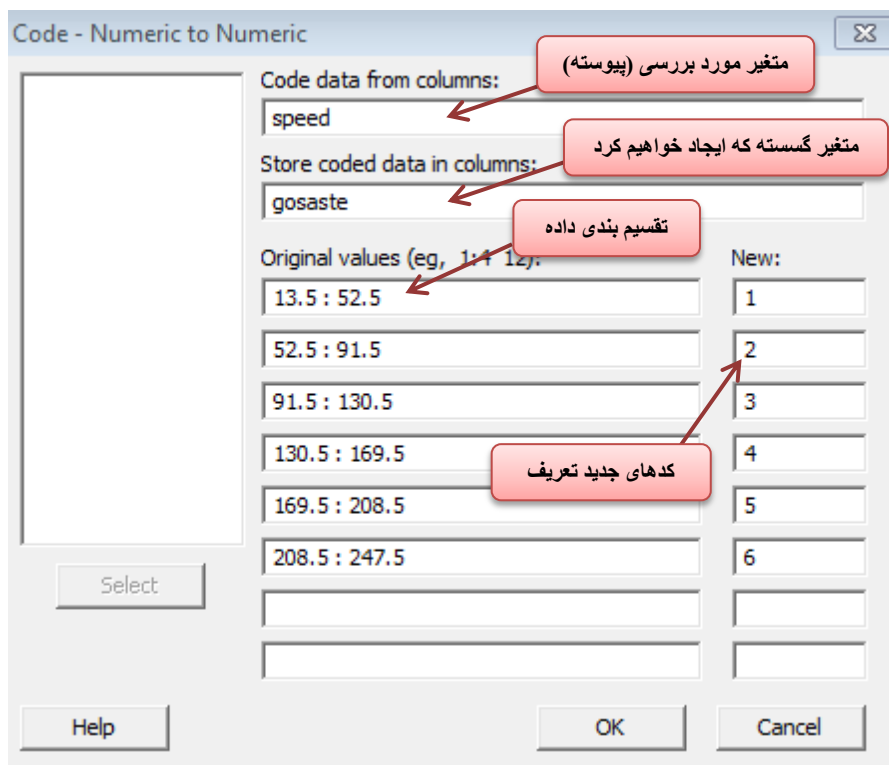
بعد از تعیین بیشینه و کمینه و فاصله و تعداد طبقات، حدود طبقات را به طور دستی تعیین می کنیم و پس از تعیین حدود طبقات به روش زیر داده های پیوسته را به گسسته تبدیل می کنیم. به عنوان مثال فرض کنید حدود طبقات عبارتند از:

- ۱ ۵۲.۵-۱۳.۵
- ۲ ۹۱.۵-۵۲.۵
- ۳ ۱۳۰.۵-۹۱.۵
- ۴ ۱۶۹.۵-۱۳۰.۵
- ۵ ۲۰۸.۵-۱۶۹.۵
- ۶ ۲۴۷.۵-۲۰۸.۵

برای تعریف و معرفی داده های گسسته ۱، ۲، ۳، ۴، ۵، ۶ به عنوان طبقات مسیر زیر را طی کنید.

Date > Code > Numeric To Numeric

سپس پنجره ی زیر باز می شود.



نحوه ی نوشتن اعداد در طبقات به صورت زیر است:


13.5 : 52.5

پس از تکمیل تمامی موارد مورد نظر کلید ok را می زنیم. ستون جدیدی با نام gosaste ایجاد شده است. حال می توان برای داده های گسسه همانند داده های کیفی جدول فراوانی ترسیم کرد. (همانطور که در بالا

توضیح داده شده است.) که در این حالت به جای اعداد طبقات ۱ و ۲ و ... قرار داده شده است.

Stat > Tables > Tally Individual Variables

و متغیر `gosaste` را به مستطیل روبرو منتقل کرده و همانند قبل عمل می کنید. خروجی نهایی در تصویر پایین نمایش داده شده است.

 Session

Tally for Discrete Variables: gosaste

gosaste	Count	Percent	CumCnt	CumPct
1	3	6.52	3	6.52
2	5	10.87	8	17.39
3	13	28.26	21	45.65
4	19	41.30	40	86.96
5	5	10.87	45	97.83
6	1	2.17	46	100.00
N=	46			

منبع:

- آموزش Minitab، آقای سعید ابراهیم نژاد ۸۹-۹۰
- Help نرم افزار Minitab