



دانشگاه صنعتی اصفهان

دانشکده ریاضی

استاد: مریم کلکین نما

تهیه کنندگان:

سید امیررضا زاهدی نژاد (40009843)

مسیح تنورساز (40006133)

بخش اول

لازم به ذکر است دیتاست ما شامل سن، جنسیت (مرد=0 و زن=1)، قد، وزن و بی ام آی است. (4 متغیر کمی پیوسته و یک متغیر کیفی)

هدف از این گردآوری دیتا، بررسی اضافه وزن و چاقی در افراد است. در واقع گلوگاه داده ها در این دیتاست BMI است. BMI مخفف Body Mass Index به معنی شاخص توده بدنی است. یک معیار برای سنجش و اندازه گیری چربی بدن بر اساس قد و وزن هر فرد است. این شاخص در مورد بدن هر زن و مرد بالغی صدق میکند. محاسبه شاخص توده بدنی، با استفاده از یک فرمول ریاضی است و برآورد میکند که در هر متر مربع از بدن چقدر وزن موجود است که در اینجا منظور از وزن چربی است و نه ماهیچه و... به طور عامیانه اگر وزن افزایش داشته باشد و قد افزایش زیادی نداشته باشد به مرور زمان BMI افزایش پیدا میکند و احتمالاً وارد ناحیه خطر بشود.



جدول فراوانی: در این نمودار شاهد هستیم که جمعیت مردان بیشتر از زنان است.

Analyze->Discriptive->Frequencies

Statistics					
gender					
N	Valid	13393			
	Missing	0			

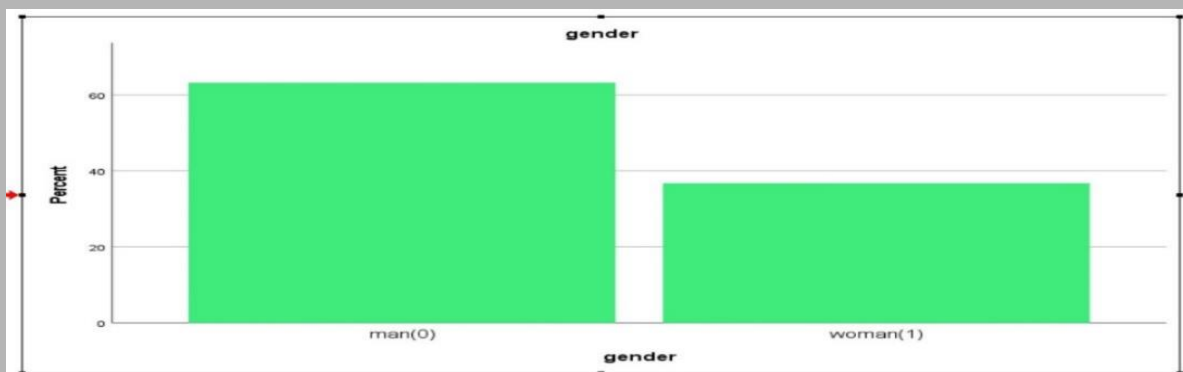
gender					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	8467	63.2	63.2	63.2
	1	4926	36.8	36.8	100.0
	Total	13393	100.0	100.0	

1 ستون فراوانی، درصد فراوانی، توزیع تجمعی

نمودار میله ای:

این نمودار بیانگر درصد افراد جامعه نسبت به کل است. حدود 62 درصد جامعه مردان و حدود 37 درصد جامعه زنان هستند.

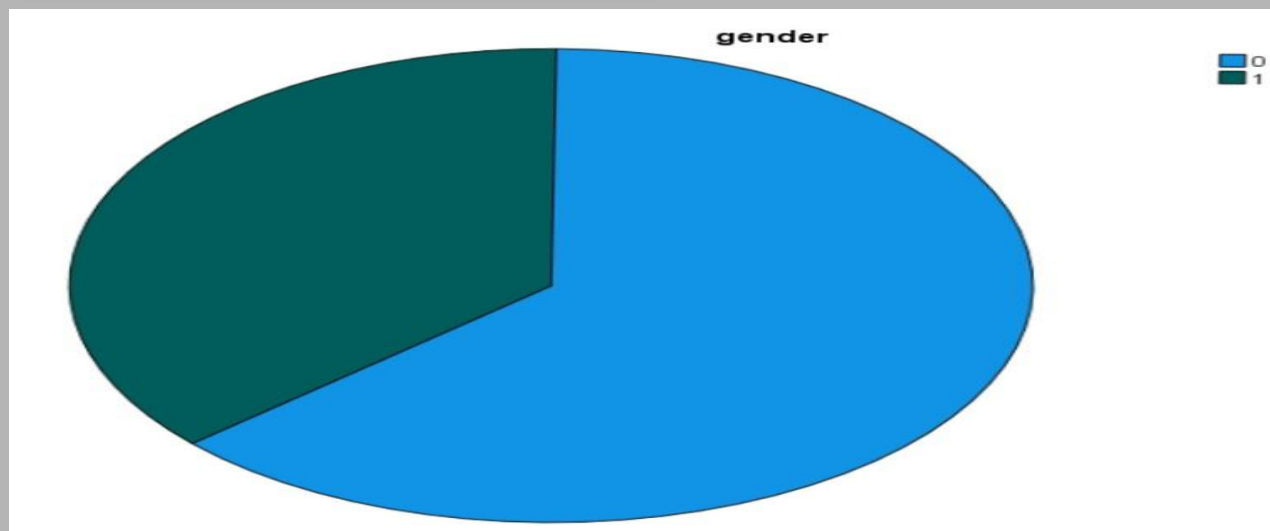
Analyze->Discriptive->Frequencies->Chart->Histogram



1 نمودار میله ای جنسیت

نمودار دایره ای بر حسب فراوانی:

Analyze->Discriptive->Frequencies->Chart->Pie



1 نمودار دایره ای بر حسب فراوانی

شاخص های تمرکز:

در این نمودار میانگین هر یک از متغیرهای پیوسته نوشته شده. مثلاً میانگین سن افراد این دیتاست چیزی حدود 37 سال است. میانه (میانۀ عددی است که یک جمعیت آماری به دو قسمت مساوی تقسیم میکند) متغیرها نیز نشان داده شده. در این دیتاست اکثر افراد سنی معادل 21 سال دارند. قد اکثر افراد 170 سانتی متر است. وزن بیشتر افراد چیزی حدود 70 کیلوگرم است و بی ام آی اکثر افراد عددی حدود 24 است. نکته جالب در جدول میانگین BMI است. هرگاه عدد BMI از 18.5 بیشتر و از 24.9 کم تر باشد بیانگر وزن نرمال است، که در این جامعه بیشتر (MODE) افراد BMI=24.32 دارند. که خوب این خبر خوبی برای این جامعه است...

Analyze->Discriptive->Frequencies->statistics->mean_mode_median

Statistics					
		age	height_cm	weight_kg	BMI
N	Valid	13393	13393	13393	13393
	Missing	0	0	0	0
Mean		36.78	168.560	67.4473	23.60601410
Median		32.00	169.200	67.4000	23.46351349
Mode		21	170.0	70.50	24.32423402

شاخص های پراکندگی:

بزرگترین و کوچکترین داده در این جدول کاملاً روشن بیان شده است. اختلاف بزرگترین و کوچکترین داده نیز با **range** نشان داده شده است. انحراف معیار که ریشه دوم واریانس است بیانگر این است که به‌طور میانگین داده‌ها به چه اندازه از مقدار متوسط فاصله دارند. کمترین انحراف معیار مربوط به BMI است، یعنی داده‌های حاصل از BMI بسیار به میانگین نزدیک هستند.

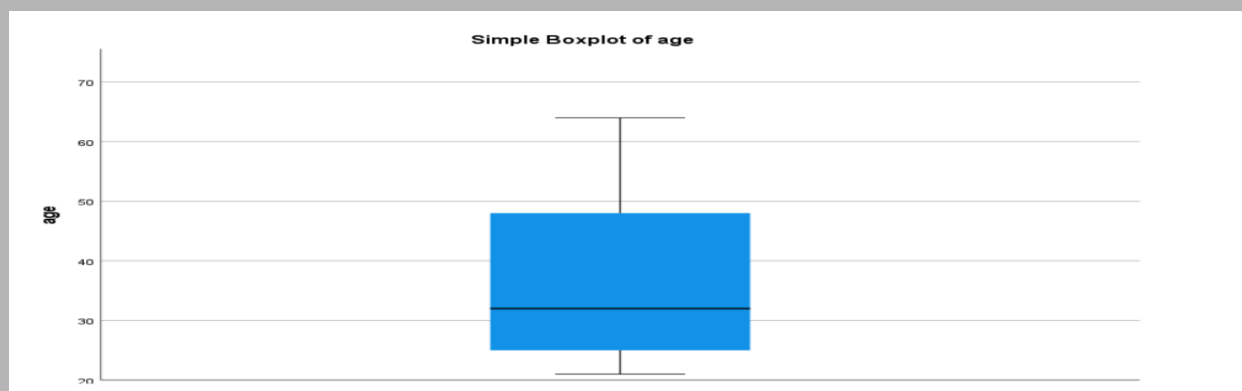
Analyze->Discriptive->Frequencies->statistics->Variance_Range_Min_Max_Deviation

Statistics					
		age	height_cm	BMI	weight_kg
N	Valid	13393	13393	13393	13393
	Missing	0	0	0	0
Std. Deviation		13.626	8.4266	2.940936101	11.94967
Variance		185.658	71.007	8.649	142.795
Range		43	68.8	31.80253302	111.80
Minimum		21	125.0	11.10397586	26.30
Maximum		64	193.8	42.90650888	138.10

نمودار جعبه ای:

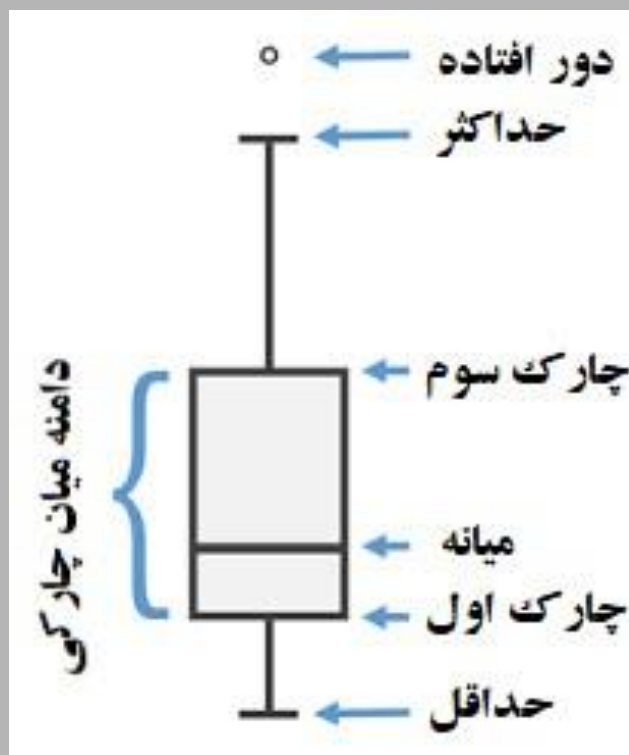
سن (سال) : کمترین سن چیزی حدود 20 سال، میانه سن افراد چیزی حدود 30 سال و بیشترین داده عددی بین 60 تا 65 است.

Graphs->ChartBuilder->Choose from: Box plot->...



در داده‌های مربوط به سن افراد **دادهای پرت** وجود ندارد.

در نمودارهای جعبه ای داده ها (نمودار) به صورت زیر تفسیر میشود:

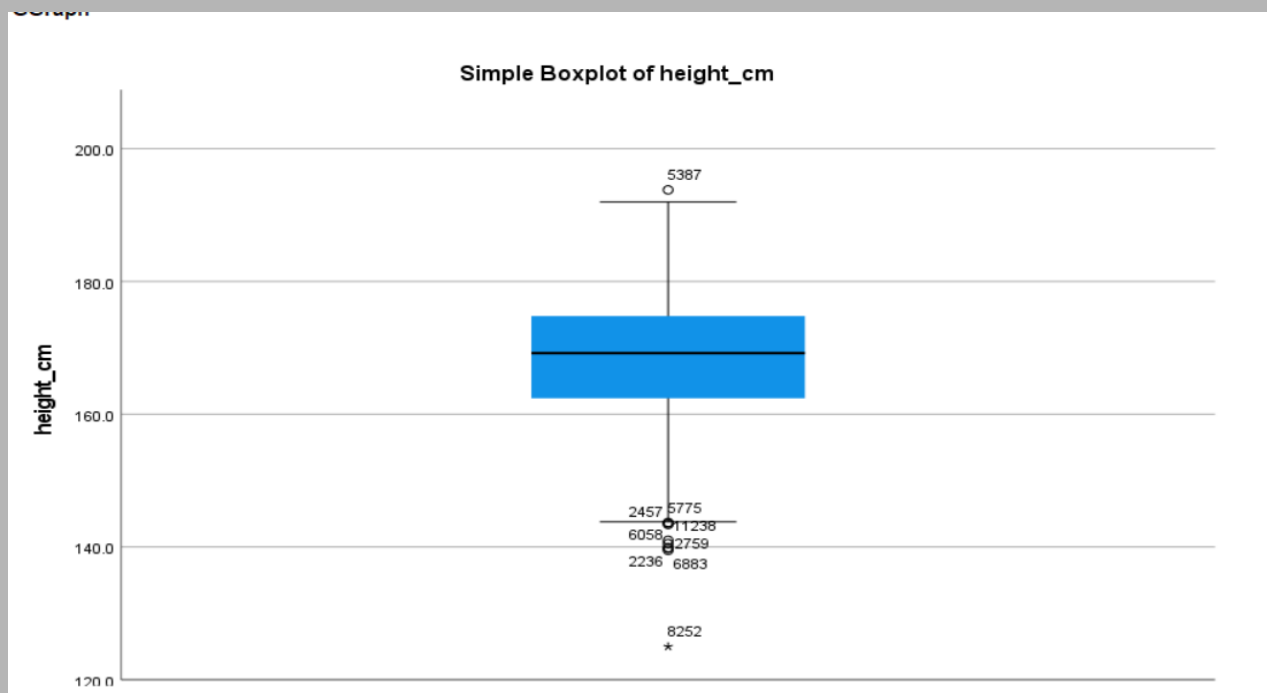


قد (cm): میانه قد افراد چیزی حدود 165 است. بزرگترین قد نزدیک عدد 180 و کوچکترین قد نزدیک به 130 است.

اعداد با ستاره یا دایره نشان داده شده است که به بیان کاربرد آن میپردازیم.

*در یک نمودار جعبه ای، دو نوع مقادیر پرت وجود دارد. یکی اندازه هایی که به شکل دایره نشان داده شده اند، این ها مقادیری هستند که بین 1.5 تا 3 برابر طول جعبه از لبه های بالا و پایین قرار دارند و دیگری اندازه هایی که به شکل ستاره نشان داده شده اند. این مقادیر در فاصله بیش از 3 برابر طول جعبه قرار گرفته اند.

وقتی با این گونه از داده ها مواجه میشویم با استفاده از شماره ای که روی آن ها قرار دارد و بیانگر ردیف آن ها در فایل داده می باشد، آن ها را پیدا میکنیم....



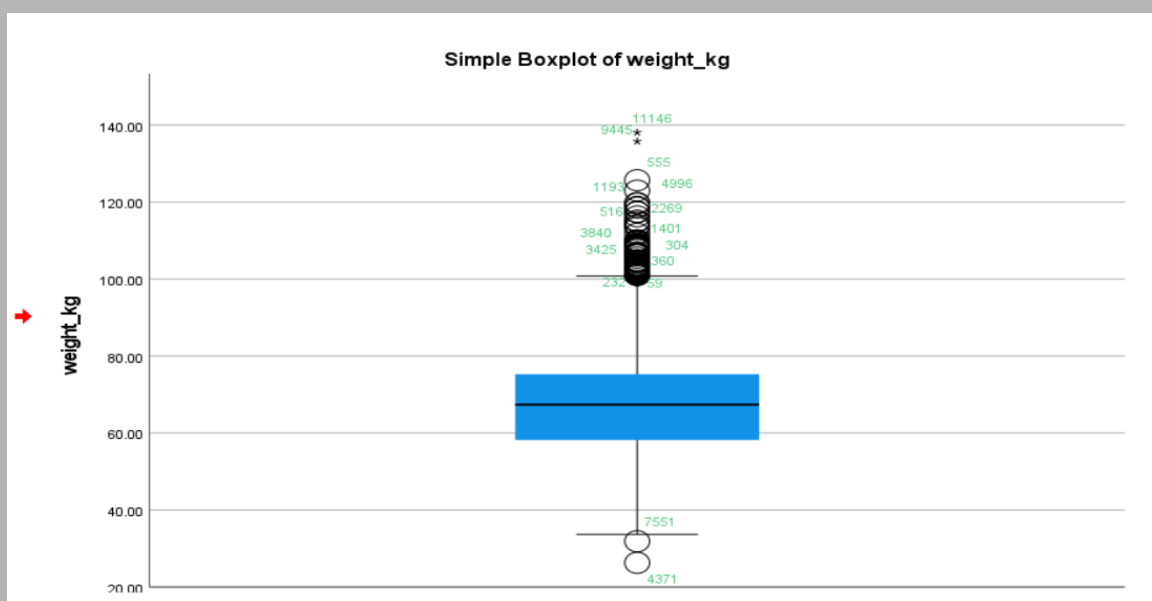
وزن(kg):

میانه = 65

ماکسیمم = 101

مینیمم = 37-38

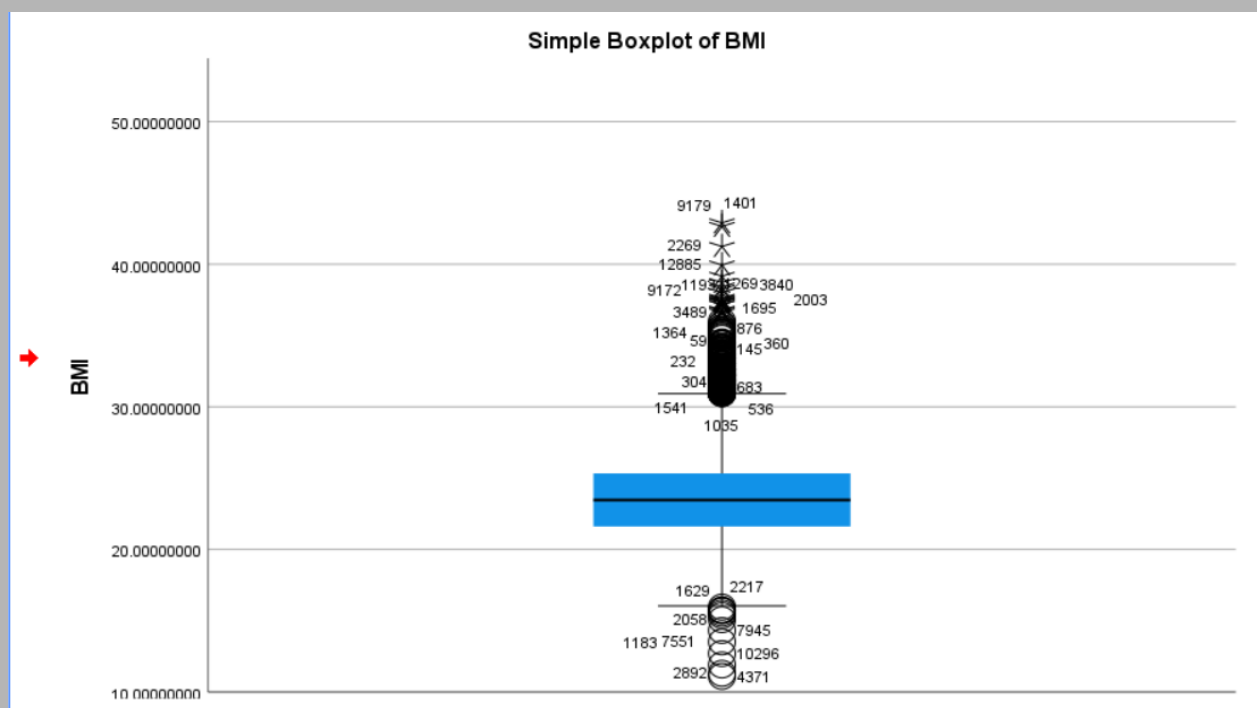
یکی از مزیت‌های مهم نمودارهای جعبه‌ای یافتن داده‌های پرت در یک‌سری از مشاهدات میباشد، که به خوبی در پایین نمایان شده.



BMI شاخص توده بدنی (Body Mass Index):

یک معیار مهم برای اندازه گیری تناسب قد و وزن افراد میباشد و بکمک این شاخص میتوان مقدار اضافه وزن یا کمبود وزن را اندازه گرفت.

در این نمودار بیشترین bmi بزرگتر از 30 و نزدیک 32 است که بیانگر این است که وزن فرد به بسیار بالاست و باید به پزشک مراجعه کند.

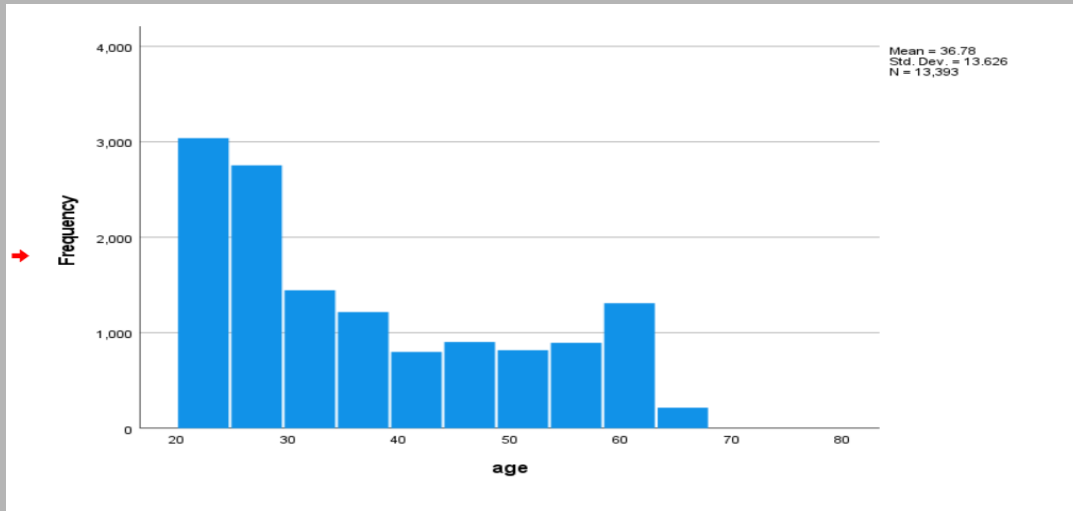


نمودار مستطیلی:

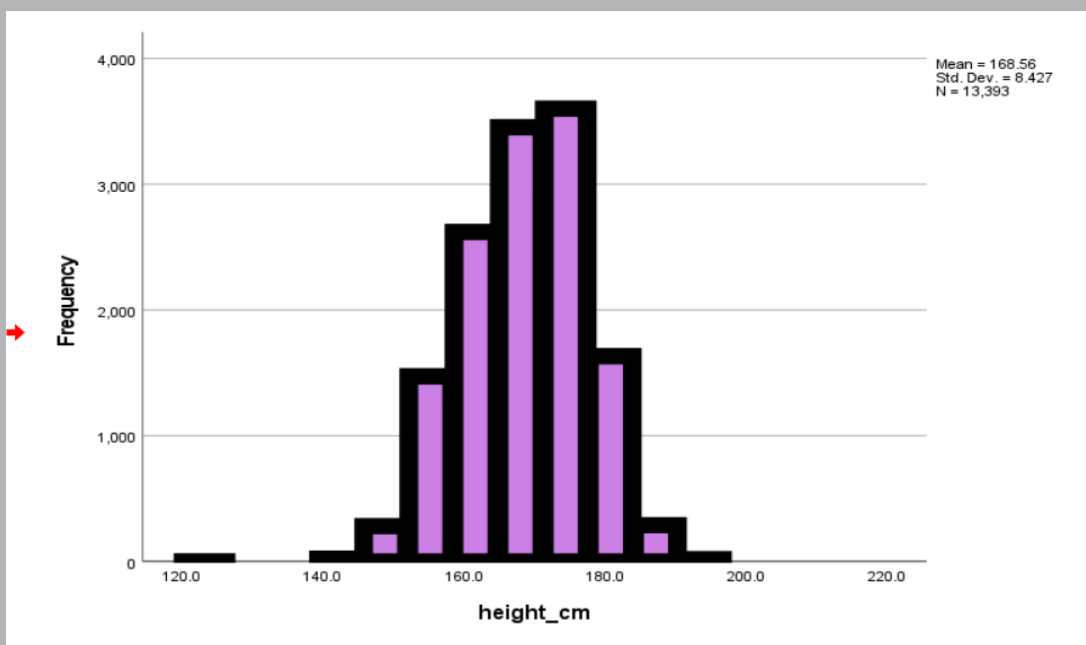
Graphs->Legacy dialogs->Histogram->Variable

Right click->properties window->binning

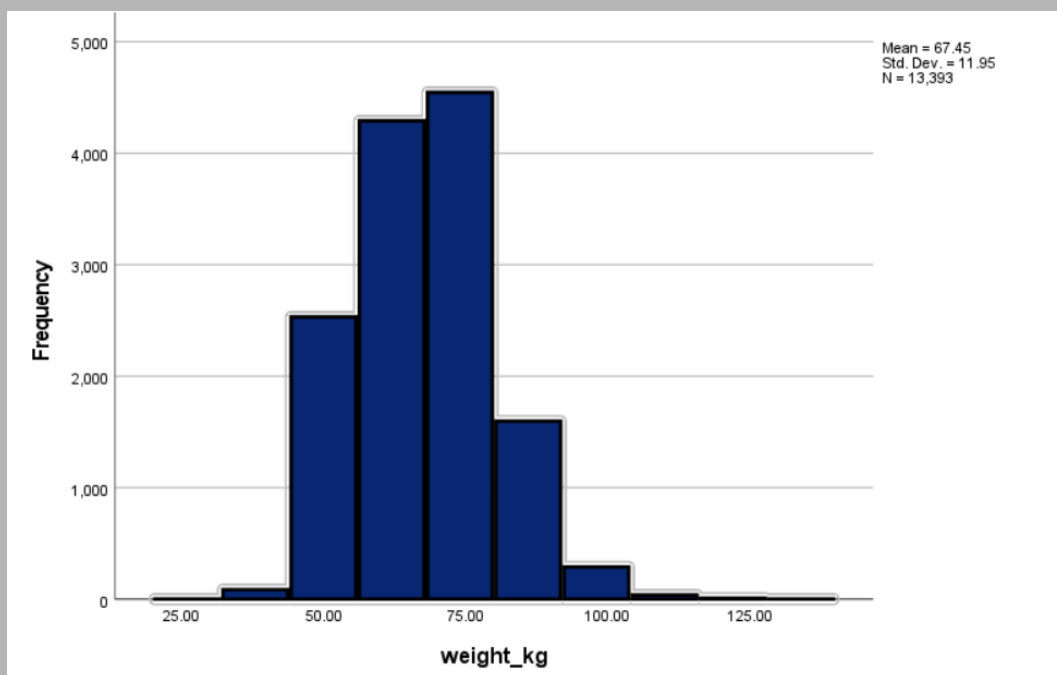
سن:



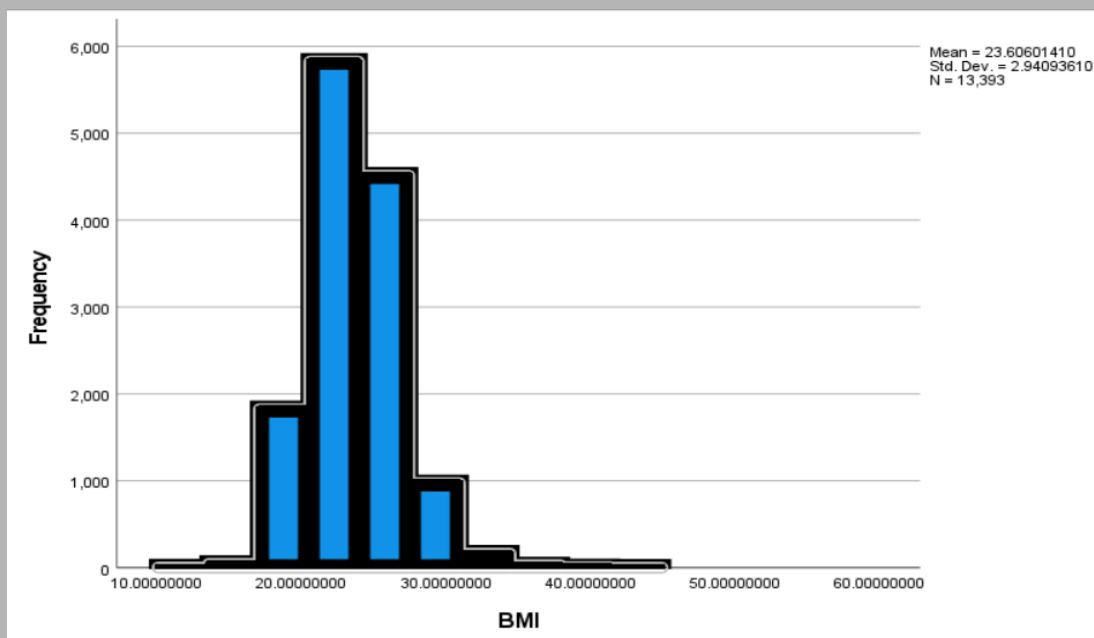
قد:



وزن:



BMI شاخص توده بدنی (Body Mass Index):



کواریانس و ضریب همبستگی:

Correlations			
		BMI	weight_kg
BMI	Pearson Correlation	1	.837**
	Sig. (2-tailed)		.000
	Sum of Squares and Cross-products	115828.816	393762.487
	Covariance	8.649	29.403
	N	13393	13393
weight_kg	Pearson Correlation	.837**	1
	Sig. (2-tailed)	.000	
	Sum of Squares and Cross-products	393762.487	1912304.288
	Covariance	29.403	142.795
	N	13393	13393

** . Correlation is significant at the 0.01 level (2-tailed).

به کمک جدول بالا میتوانیم کواریانس و ضریب همبستگی را حساب کنیم.

فرمول: $\text{Sum of Squares and Cross-products} / N = \text{covariance}$

مثال:

$$\text{COV (BMI, Weight)} = 393762.487 / N (13393) = 29.40;$$

اگر کواریانس دو متغیر مثبت باشد یعنی یک رابطه مستقیم (هم جهت) بین آنها وجود دارد. خلاف این موضوع نیز وجود دارد. اگر کواریانس منفی باشد یعنی دو متغیر رابطه معکوس دارند.

چون عدد کواریانس بین BMI و Weight مثبت شده یعنی رابطه ای مستقیم بین این دو وجود دارد و هرچه قدر وزن بیشتر میزنیم باشد BMI نیز بالاتر است.

*ضریب همبستگی: در جدول بالا ضریب همبستگی دو متغیر از تقاطع **Pearson Correlation** متغیر اول و اسم متغیر دوم به دست می آید.

مثال:

ضریب همبستگی (Weight, BMI): 0.837

نکته: در اینجا فقط ضریب همبستگی پیروسون محاسبه شده.

توجه: اگر ضریب همبستگی دو پارامتر با یکدیگر مثبت باشد، به این معناست که در فضایی که مطالعه و بررسی انجام شده، افزایش یک پارامتر با افزایش پارامتر دیگر و نیز کاهش آن پارامتر با کاهش پارامتر دیگر همراه است.

بخش دوم

1. فاصله اطمینان 95 درصدی برای میانگین:

One-Sample Test						
Test Value = 0						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
age	312.346	13392	.000	36.775	36.54	37.01

مقادیر Lower و Upper بیانگر سر و ته فاصله اطمینان 95 درصدی است. با احتمال 95 درصد مقدار میانگین جامعه هدف بین کران پایین و بالا است.

2. فاصله اطمینان 95 درصدی برای واریانس:

Stat->Basic statistics-> Variance...

Spss امکان محاسبه فاصله اطمینان 95 درصدی برای واریانس رو نداشت. این محاسبات با Minitab انجام شده است.

Method

σ : standard deviation of age

The Bonett method is valid for any continuous distribution.

The chi-square method is valid only for the normal distribution.

Descriptive Statistics

N	StDev	Variance	95% CI for σ	
			using Bonett	Chi-Square
13393	13.6	186	(13.5, 13.7)	(13.5, 13.8)

در این جدول اولین ستون از سمت راست مورد بررسی ما است. البته اینجا فاصله اطمینان برای انحراف معیار اندازه گیری شده است. پس اعداد به توان دو می‌رسند. با احتمال 95 درصد مقدار واریانس جامعه هدف بین کران پایین و بالا است.

3. آزمون:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
age	13393	36.78	13.626	.118

One-Sample Test

Test Value = 46.78

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
age	-84.976	13392	.000	-10.005	-10.24	-9.77

نتیجه: آزمون میانگین در سطح معنادار 0.05 برای متغیر سن با $\text{test_value} = 46.78$ انجام شده است.

همان طور که در جدول میبینیم مقدار $P\text{value}(\text{sig})$ از آلفا که پنج صد است کوچکتر است. پس H_0 قطعاً رد میشود و H_1 قبول میشود.

و این بدان معناست که 46.78 برابر میانگین نیست و بزرگتر از آن است.

مسیر مورد استفاده:

Analyze->Compare means->one sample T test...

منبع دیتاست استفاده شده:

[اینک](#)

لازم به ذکر است در این دیتاست تغییراتی اعمال شده...