

**دانشگاه صنعتی اصفهان**

**دانشکده ریاضی**

**استاد: مریم کلکین نما**

تهیه کنندگان:

سید امیررضا زاهدی نژاد (40009843)

مسیح تنورساز (40006133)

**بخش اول**

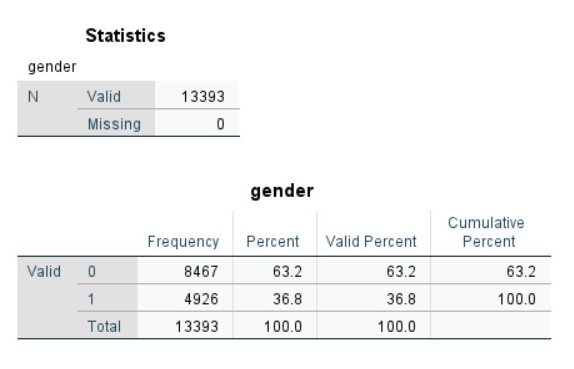
لازم به ذکر است دیتاست ما شامل سن، جنسیت (مرد=0 و زن=1) ، قد، وزن و بی ام آی است.(4متغیر کمی پیوسته و یک متغیر کیفی)

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 براورد 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 احتمالا وارد ناحیه خطر بشود.



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

Analyze->Discriptive->Frequencies

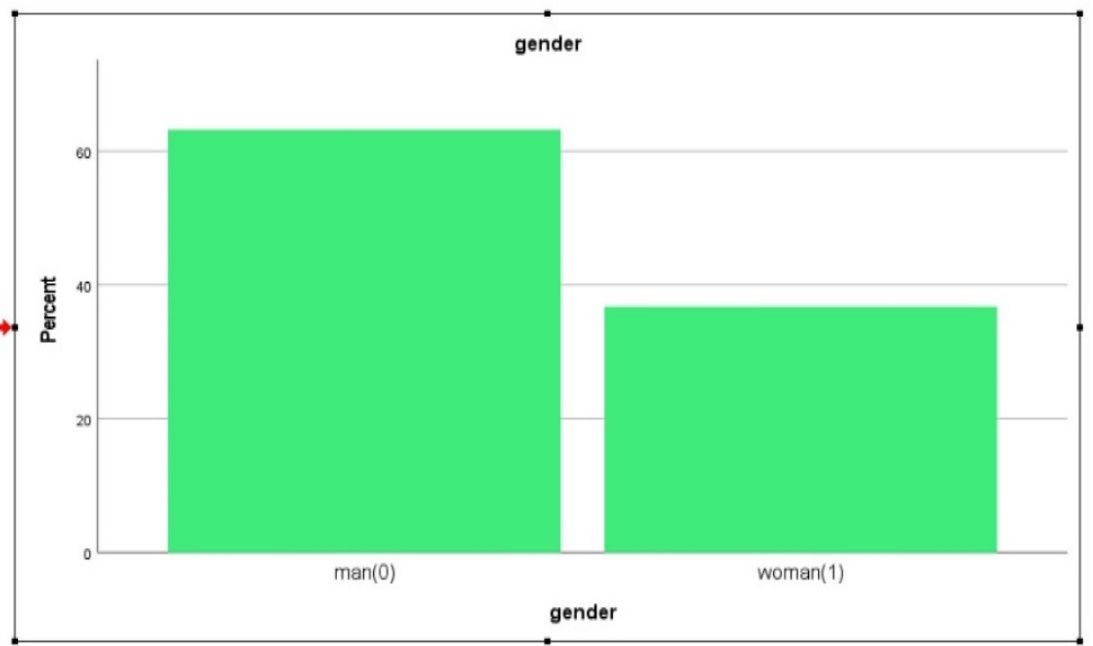


**ستون فراوانی،درصد فراوانی،توزیع تجمعی 1**

**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.

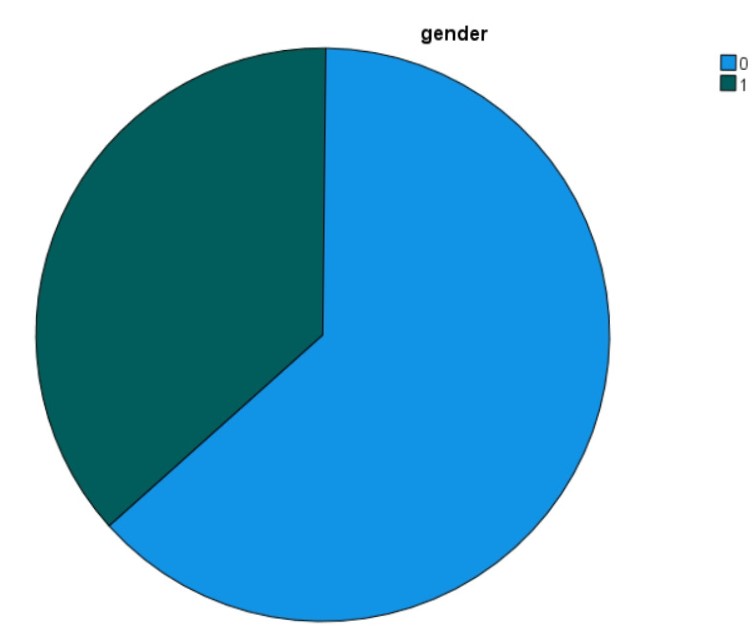
Analyze->Discriptive->Frequencies->Chart->Histogram



**نمودار میله ای جنسیت 1**

**نموداردایره ای بر حسب فراوانی:**

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



**نمودار دایره ای بر حسب فراوانی 1**

**شاخص های تمرکز:**

در این نمودار میانگین هر یک از متغیرهای پیوسته نوشته شده.مثلا میانگین سن افراد این دیتاست چیزی حدود 37 سال است.میانه(میانه 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



**شاخص های پراکندگی:**

بزرگترین و کوچکترین داده دراین جدول کاملا 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 مقدارمتوسط فاصله دارند. کمترین انحراف معیار مربوط به BMI است، یعنی داده های حاصل از BMIبسیار به میانگین نزدیک هستند.

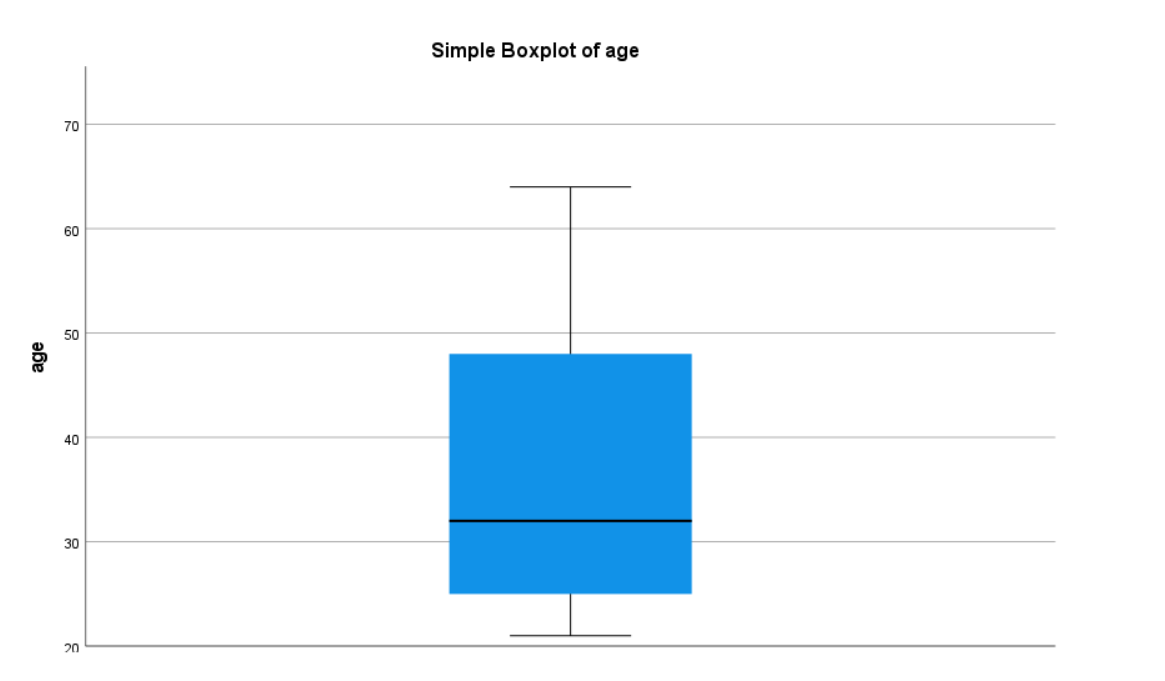
Analyze->Discriptive->Frequencies->statistics->Variance\_Range\_Min\_Max\_Deviation



**نمودار جعبه ای:**

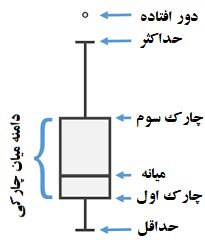
سن(سال) : کمترین سن چیزی حدود 20 سال، میانه سن افراد چیزی حدود 30 سال و بیشترین داده عددی بین 60 تا 65 است.

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



در داده های مربوط به سن افراد دادهای پرت وجود ندارد.

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

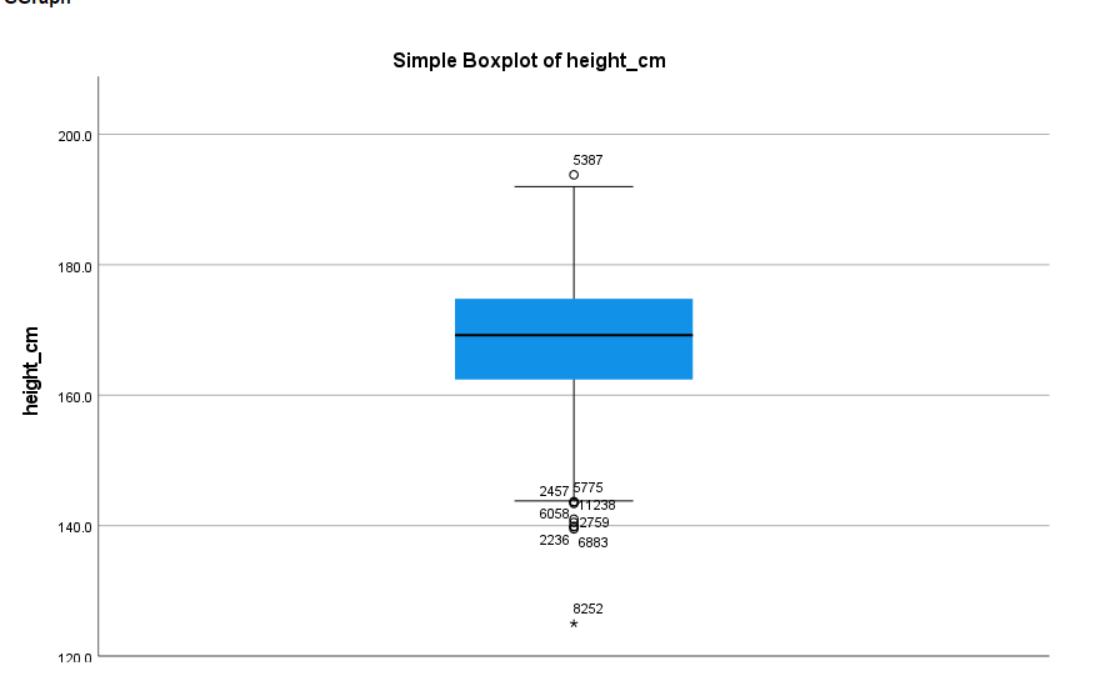


قد(cm): میانه قد افراد چیزی حدود 165 است. بزرگترین قد نزدیک عدد 180 و کوچکترین قد نزدیک به 130 است.

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

\*در یک نمودار جعبه‌ای، دو نوع مقادیر پرت وجود دارد. یکی اندازه‌هایی که به شکل دایره نشان داده شده‌اند، این‌ها مقادیری هستند که بین 1.5 تا 3 برابر طول جعبه از لبه‌های بالا و پایین قرار دارند و دیگری اندازه‌هایی که به شکل ستاره نشان داده شده‌اند. این مقادیر در فاصله بیش از 3 برابر طول جعبه قرار گرفته‌اند.

وقتی با این‌گونه از داده‌ها مواجه میشویم با استفاده از شماره‌ای که روی آن‌ها قرار دارد و بیانگر ردیف آن‌ها در فایل داده می‌باشد، آن‌ها را پیدا میکنیم....

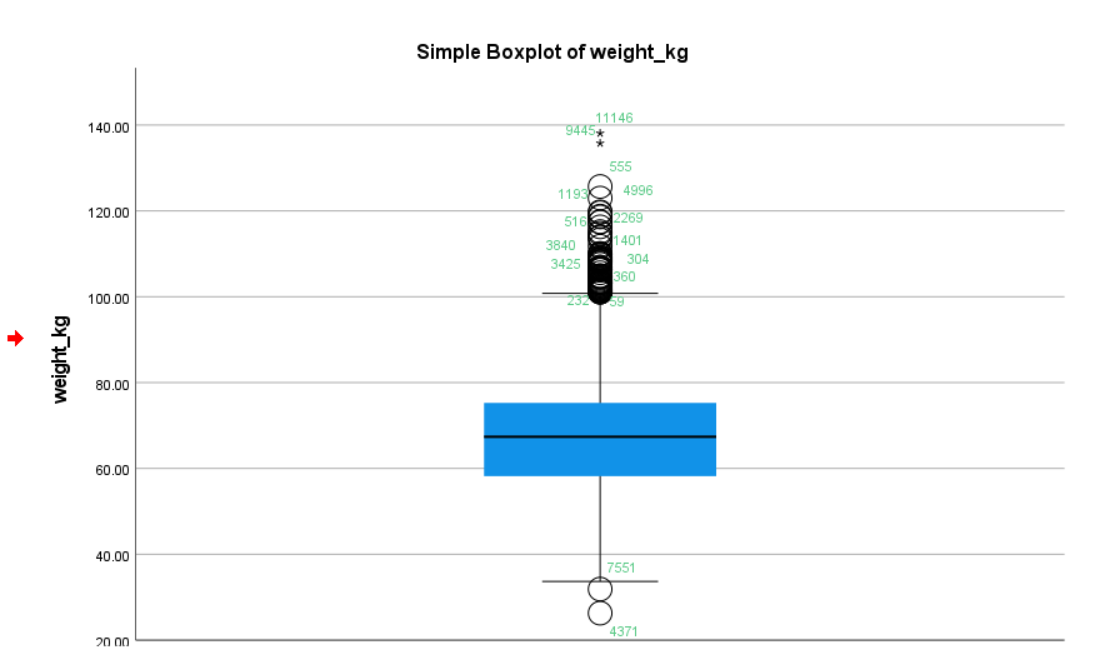


وزن(kg):

میانه = 65

ماکسیمم = 101

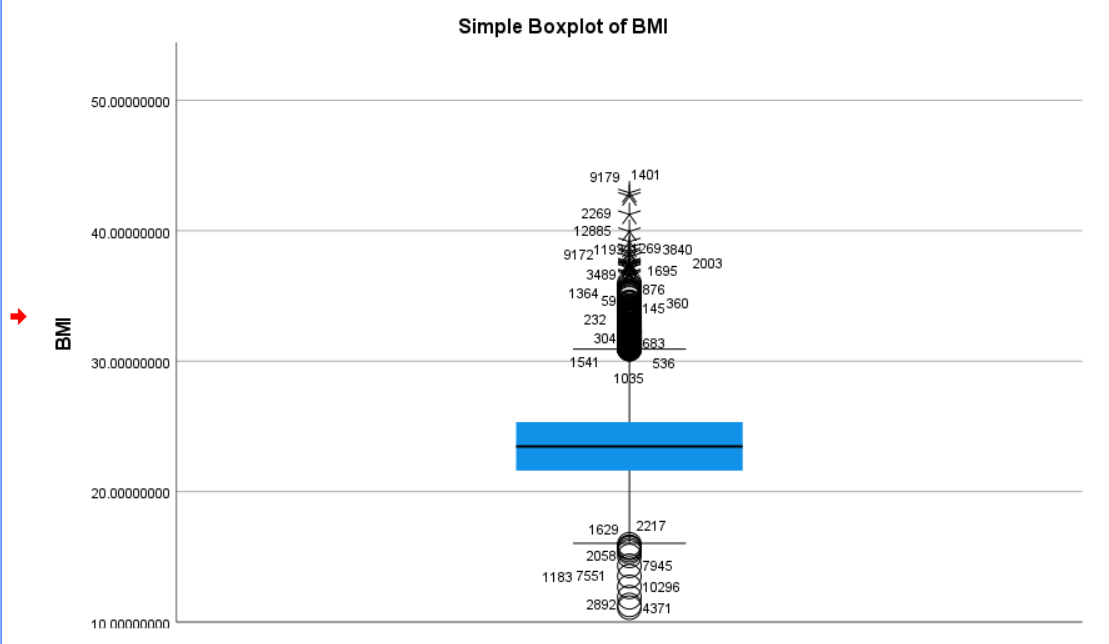
مینیمم = 37-38

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.

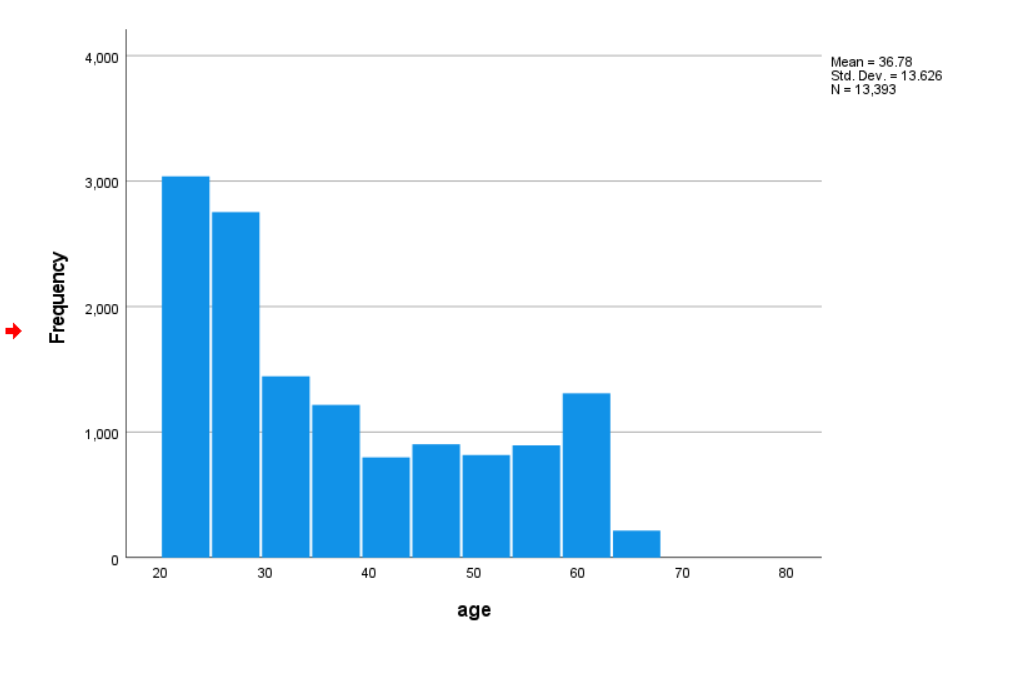


**نمودار مستطیلی:**

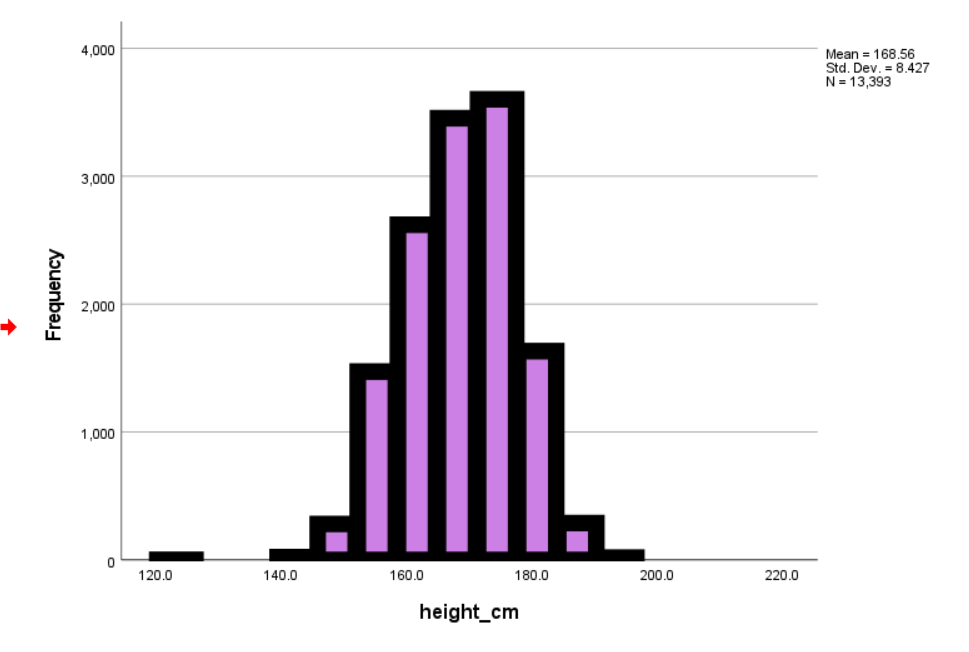
Right click->properties window->binning

Graphs->Legacy dialogs->Histogram->Variable

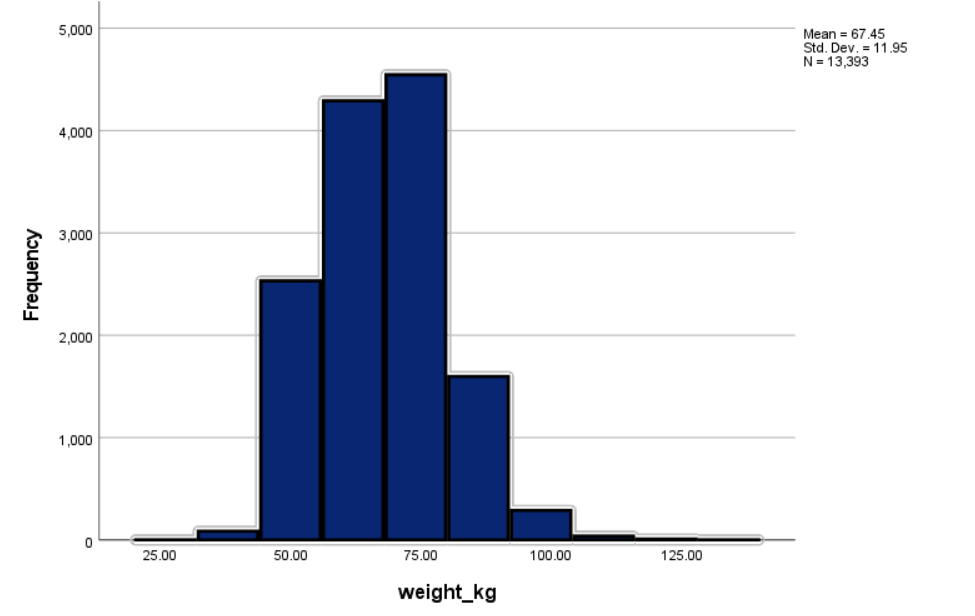
**سن:**



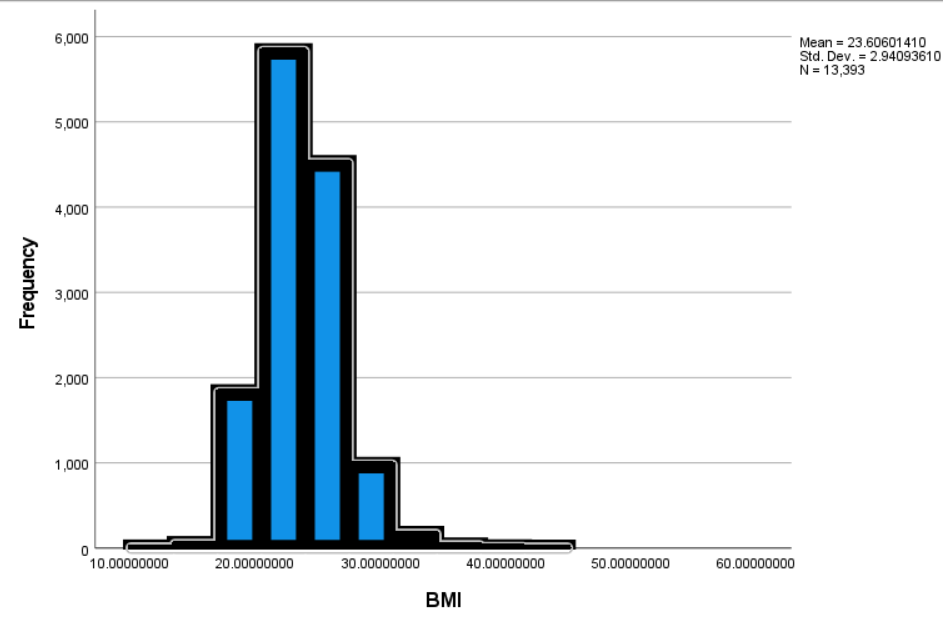
**قد:**



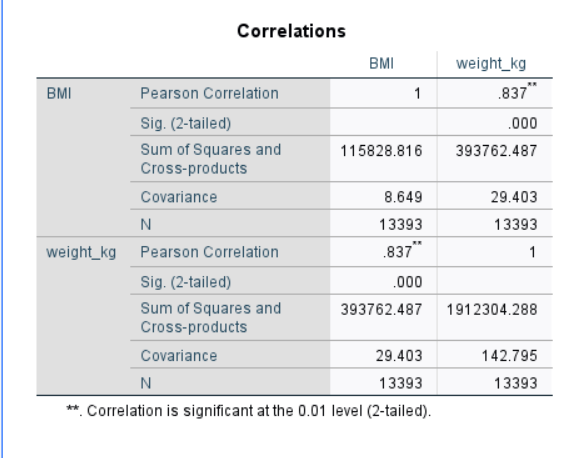
**وزن:**



BMI شاخص توده بدنی (Body Mass Index):



**کواریانس و ضریب همبستگی:**



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

فرمول Sum of Squares and Cross-products / N = covariance :

مثال:

COV (BMI, Weight) = 393762.487/N (13393) = 29.40;

اگر کواریانس دو متغیر مثبت باشد یعنی یک رابطه مستقیم(هم جهت) بین انها 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:

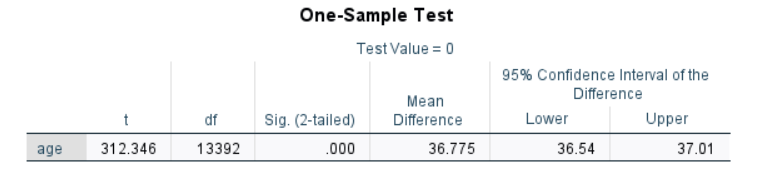
ضریب همبستگی(Weight, BMI):0.837

نکته: در اینجا فقط ضریب همبستگی پیرسون 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:

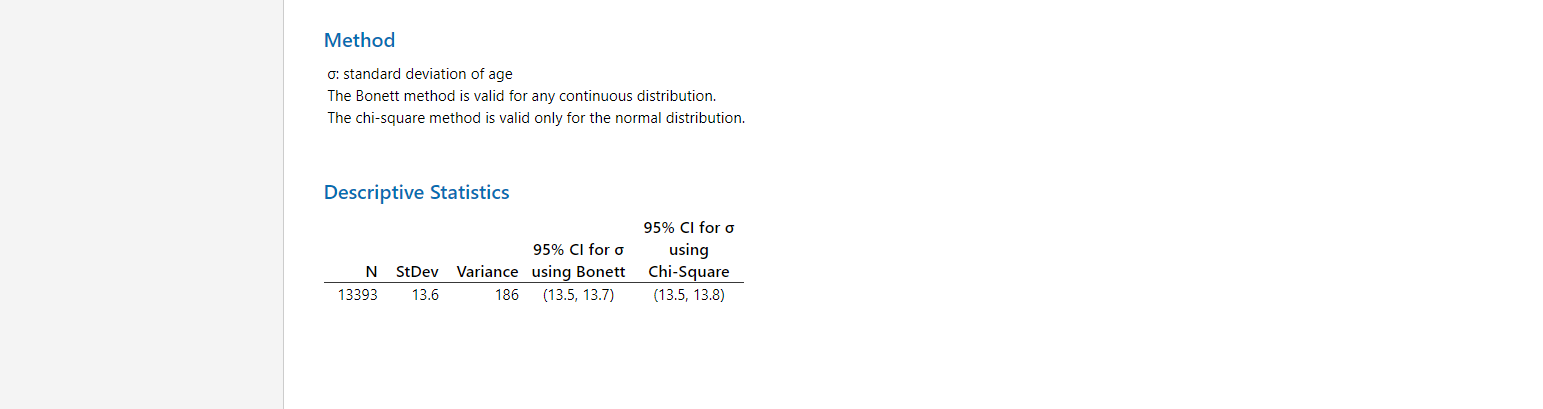


مقادیر Lower و Upper بیانگر سر و ته فاصله اطمینان 95 درصدی است. با احتمال 95 درصد مقدار میانگین جامعه هدف بین کران پایین و بالا است.

1. فاصله اطمینان 95 درصدی برای واریانس:

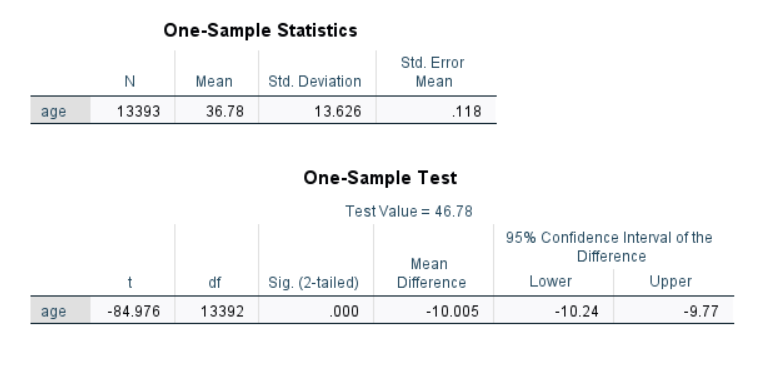
Spss امکان محاسبه فاصله اطمینان 95 درصدی برای واریانس رو نداشت. این محاسبات با Minitab انجام شده است.

Stat->Basic statistics-> Variance…



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.

1. **آزمون:**



نتیجه: آزمون میانگین در سطح معنادار 0.05 برای متغیر سن با test\_value = 46.78 انجام شده است.

همان طور که در جدول میبینیم مقدار Pvalue(sig) از آلفا که پنج صدم است کوچکتر است. پس H0 قطعا رد میشود و H1 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:

[لینک](https://www.kaggle.com/datasets/hamza256/bodyperformance)

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