دکتر محمد امینی مباحث ویژه آماری دانشگاه فردوسی مشهد.

۱۳۹۸

آمار توصیفی

آمار چیست؟

آمار علمی است که پیرامون جمع آوری و تنظیم و تحلیل و تفسیر اطلاعات عددی سخن می گوید.

موضوع آمار چیست؟

موضوع آمار عبارتست از هنر و علم جمع آوری، تعبیر و تجزیه و تحلیل داده ها و استخراج نتایج منطقی در مورد پدیده های تحت بررسی.

اهداف علم آمارچیست؟

هدف اول: توصیف ساده، صریح و قابل فهم مشاهده ها که معمولاً در نمونه مناسبی از جامعه صورت می گیرد.

هدف دوم: تعمیم نتایج مشاهده های مزبور به جامعه ای که نمونه از آن برگرفته شده است.

ابزار های اصلی علم آمارکدامند ؟

۱- آمار توصیفی که ویژگی های کلی تعدادی از داده ها را (که معمولاً به نمونه تعلق دارند) در قالب نمودار؛ جداول و یا یک عدد بیان می کند. آمار توصیفی به عنوان یک ابزار مفید از اهداف اولیه علم آمار است.

۲- آمار استنباطی که بر اساس ویژگیهای مشاهده شده در نمونه، ویژگی های جامعه را برآورد می کند. آمار استنباطی شامل ابزار های مفید برای هدف دوم علم آمار است.

آمار توصیفی:

آمار توصیفی شامل ابزار بسیار مفیدی است که به ساده گی میتوان داده های آماری را تفسیر؛ توصیف و تحلیل نماییم. در این بخش این ابزار ها به طور مختصر قابل استفاده معرفی و علاوه بر این روش اجرای این ابزار با استفاده از نرم افزار SPSS برای چندین نمونه داده تحلیل می شود.

۱-جداول آماری:

الف: جدول فروانی ساده: این جدول برای داده های گسسته؛ کیفی و پیوسته با انواع کم رسم می شود.

ب: جدول فراوانی طبقه بندی: برای داده های پیوسته، گسسته با انواع زیاد رسم می شود.

رسم جداوی آماری در نرم افزار SPSS

۱- رسم جدول فراوانی ساده

Step 1. Start the SPSS software, choose type in data and enter your data or open the data file.

Step2. In the SPSS data editor window, choose Analyze then Descriptive Statistics, then Frequencies to open the Frequencies dialog box.

Step3. In the Frequencies dialog box: select the variable name of your data and transfer it into the variables: check the display frequency table box and the Ok.

۲- رسم جدول فراوانی طبقه ای

Step 1. In the SPSS data editor window choose Transform, then Record, Into Different variables to open the Record into Different variables dialog box.

Step2. In the Record into Different variables dialog box select. variable name of your data and transfer it into the Input variable-output variable box.

Type class in the name: box and then click on change, then old and new values, to open the Record Different variables: Old and New values dialog box.

Step 3. In the Record into Different variables. Old and New values select Range and type c1=min x-i-a to the left of through and c1+h to the right of through. Check the Output variables are strings box and type 11 in the width. Also type ci-c(i+1) in the box for New value. Then click on Add and repeat this action until all the range are set.

Then Continue, then Analyze, then Frequencies, then the file of class transfer into Variables, check Display frequency tables and then OK.

۲-نمودارهای آماری

الف: نمودار هیستوگرام: برای بررسی شکل توزیع به خصوص نرمال بودن یا غیر نرمال بودن داده ها رسم می شود.

ب:نمودار جعبه ای: یک نکایش تصویری برای بررسی تقارن یا عدم تقاراز ماکزیمم؛ مینیمم و چارکهای توزیع بیان می کند معمولا برای مقایسه حداقل دو سری داده استفاده میشود.

ج:نمودار ساقه و برگ: یک نمودار میله ای ستونی از داده ها را ارایه می دهد و داده های واقعی در هر لحظه در اختیار پژوهشگر هستند به عبارتی دیگر تصویر واقعی از توزیع داده ها را نشان می دهد.

د:نمودار دایره ای: برای تحلیل توصیفی داده های کیفی

رسم نمودارها در نرم افزار SPSS

۱- رسم نمودار های آماری برای جداول فراوانی ساده

Analyze---Descriptive Statistics---Frequencies----Transfer file of data into Variables, then click Charts and then select your charts based on frequencies or Percentages.

۲- رسم نمودار ساقه و برگ

Analyze-----Descriptive Statistics----Explore-----Transfer file of data into Dependent list, then select Plots for display. Click on Plots to open the Explore: Plots dialog box. Select None for Boxplots and Stem-and-leaf for Descriptive and then click on continue to return to the Explorer dialog box. Then click on OK.

۳-(شاخص های مرکزی)

شاخصهای مرکزی یا تمرکز برای مطالعه مرکز داده ها که در تحلیل های آماری نقش مهمی دارد بررسی می شوند. در ادامه شاخصهای معروف و پر کاربرد بیان می شود.

میانگین حسابی (Mean) :

به نقطه تعادل یا مرکز ثقل توزیع ، در داده هایی که بصورت منظم بر روی یک محور ردیف شده باشند ، میانگین

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + ... + X_n}{n}$$

(**Mean**) اطلاق می شود

میانه (Median) :

مقداری است که نیمی (۵۰٪) مشاهدات کمتر از آن مقدار باشد.

اگر **n** فرد باشد خواهیم داشت :

میانه = $Md = X_{(\frac{n+1}{2})}$

اگر **n** زوج باشد

میانه = $Md = \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}$

مُد (نما) (Mode) :

به مقداری گفته می شود که در میان سایر مقادیر توزیع ، بیشترین تکرار را داشته باشد ، مد را با Mo نشان می دهند. اشکالی که در این شاخص وجود دارد این است که امکان دارد در یک جامعه بیش از یک مُد داشته باشیم.

چارکها (Quartiles): چارکها مقادیری هستند که داده ها را به چهار قسمت مساوی تقسیم میکنند.

چارک اول: ۲۵ درصد داده ها کوچکتر یا مساوی آن هستند. با $Q1$ نشان میدهند

چارک دوم: ۵۰ درصد داده ها کوچکتر یا مساوی آن هستند. با $Q2$ نشان می دهند.

چارک سوم: ۷۵ درصد داده ها کوچکتر یا مساوی آن هستند. با $Q3$ نشان میدهند.

۴-شاخص‌های پراکندگی : کمیتهایی هستند که میزان تراکم یا عدم تراکم داده ها را اندازه گیری و بیان می کنند.

دامنه تغییرات: (Range)

برابر است با اختلاف کمترین مقدار مشاهده با بیشترین مقدار مشاهده. از معایب دامنه تغییرات این است که فقط از دو مقدار کمترین و بیشترین تأثیر می‌پذیرد و بقیه مشاهدات در نشان دادن وضعیت پراکندگی نقشی ندارند.

$$R_n = \max\{x_1, x_2, ..., x_n\} - \min\{x_1, x_2, ..., x_n\}$$

متوسط انحرافات از میانگین :

این شاخص از تقسیم مجموع قدر مطلق انحرافات تک تک مشاهدات از میانگین شان بر تعداد مشاهدات بدست می آید.

$$M.D = \frac{\sum_{i=1}^{n} | x_i - \bar{x} |}{n}$$

واریانس (Variance)

واریانس داده ها که میزان تراکم داده حول مرکز ثقل را اندازه گیری و بیان می نماید؛ به دو صورت تعریف و در کاربردها استفاده می شود.

۱- این رابطه یک تخمین ناریب برای واریانس جامعه ارایه می دهد.

$$S_1^2 = \frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)} = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

۲- فرمول زیر یک تخمین اریب برای واریانس جامعه ارایه می دهد.

$$S_2^2 == \frac{\sum (x_i - \bar{x})^2}{n} = \frac{1}{n}\left\{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right\}$$

انحراف معیار (Standard Deviation)

مقادیر این معیار مشابه واریانس تعبیر می شود با این تفاوت که واحد انحراف معیار همان واحد داده هاست. به صورت زیر تعریف می شود

$$S = \sqrt{S^2} \qquad S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

اختلاف میان چارکی (Inter Quintile Range)

این کمیت با حذف داده های کوچک ۲۵ درصد پایینی و حذف داده های بزرگ ۲۵ درصد بالایی به صورت $IQR = Q_3 - Q_1$ **محاسبه می شود. در واقع ۵۰ درصد داده ها در دامنه** $(Q_1,\ Q_3)$ **قرار میگیرند. مقادیر کوچک معیار اختلاف تراکم داده ها و مقادیر بزرگ پراگنده گی داده ها اطراف میانه را نشان می دهند.**

توجه: اگر فاصله های $(Q_1,\ Q_2)$ **و** $(Q_2,\ Q_3)$ **تقریبا برابر باشند توزیع متقارن است.**

اگر فاصله $(Q_1,\ Q_2)$ **کوچکتر از** $(Q_2,\ Q_3)$ **باشد توزیع نامتقارن و دارای چولگی مثبت و دم راست کشیده می باشد.**

اگر فاصله $(Q_1,\ Q_2)$ **بزرگتر از** $(Q_2,\ Q_3)$ **باشد توزیع نامتقارن و دارای چولگی منفی و دم چپ کشیده می باشد.**

انواع حالات توزیع ها :

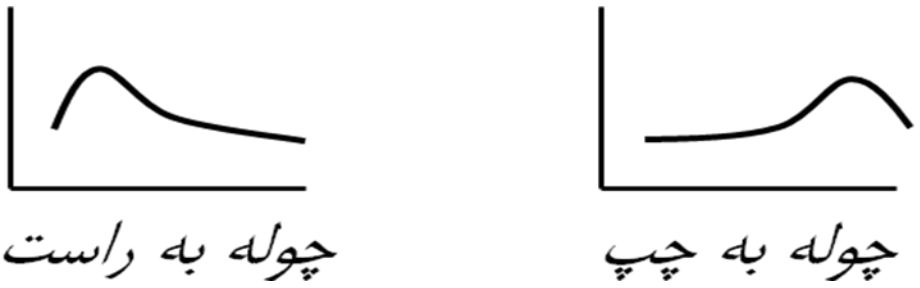
۱- متقارن (نرمال) : مد = میانه = میانگین

۲- چولـه به راسـت : مد < میانه < میانگین

۳- چـولـه بـه چـــپ : مد > میانه > میانگین

مفهوم چولگی :(Skewness)

اگر دم توزیع جامعه به سمت راست باشد ، توزیع را چوله به راست و در صورت عکـس ، آن را چوله به چپ می نامند.



تفسیر مقادیر مختلف ضریب چولگی (SK) :

۱- صفر : در صورت متقارن بودن توزیع جامعه

۲- مثبت : در صورت چوله به راست بودن توزیع جامعه

۳- منفی : در صورت چوله به چپ بودن توزیع جامعه

تفسیر مقادیر SK:

۱- $|SK| \le 0.1$ **، جامعه تقریباً نرمال**

۲- $0.1 < |SK| \le 0.5$ **، تفاوت اندک با توزیع نرمال**

۳- $|SK| > 0.5$ **، تفاوت معنی دار با توزیع نرمال**

فرمول های محاسبه ضریب چولگی (SK) :

۱- ضریب چولگی گشتاوری $SK = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{S^3}$

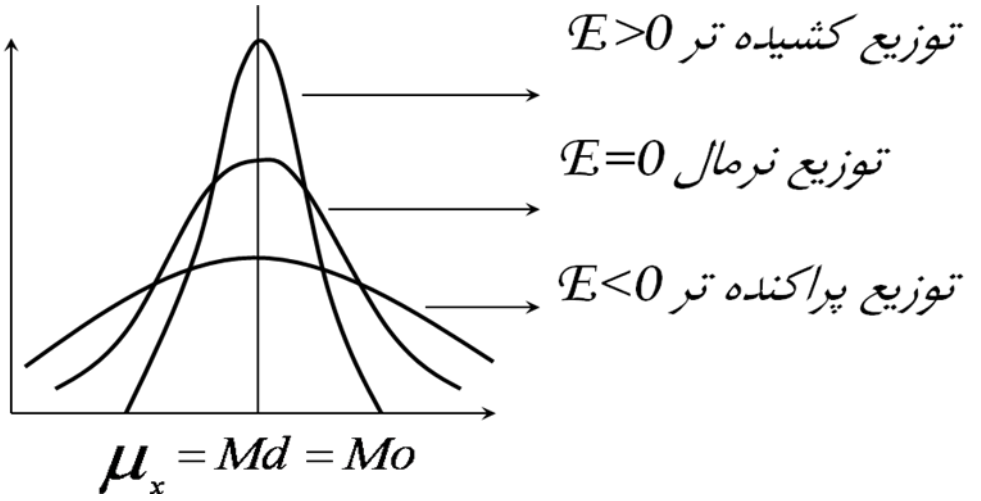
۲- ضریب های چولگی پیرسون

$S \cdot K_1 = \dfrac{(\bar{X} - Mo)}{S}$ $\qquad$ $S \cdot K_2 = \dfrac{3(\bar{X} - Md)}{S}$

***کشیده گی* (Kurtosis)**

این ضریب که به صورت زیر تعریف می شود ارتفاع توزیع داده ها را نسبت به توزیع نرمال استاندارد اندازه گیری و بیان می نماید.

$$K = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4 / S^4 - 3$$



۳- شاخصهای مرکزی و پراکنده گی برای داده های جدول فراوانی ساده در نرم افزار

Select Analyze----Descriptive Statistics------Frequencies-----Transfer file of data into Variables, select Statistics and then choose in Dialog box of Statistics then select Quartiels-Centeral Tendency—Dispersion---and Skew ness, Kurtosis.

۴-شاخصهای مرکزی و پراکنده گی برای داده های جدول فراوانی طبقه ای در نرم افزار

Now, you have variables class midpoint and frequency in SPSS Data Editor;

Select Data in menu----weight cases---- check weighted cases by and then Transfer frequency into weighted cases by then click OK.
Choose Analyze-----Desc.Stat-----Frequencies-----In dialog box of frequencies transfer midpoint into Variables------click Statistics: In this dialog box check Values are group midpoints. you can choose Quartiles—Dispersion—central Tendency –Skewness and Kurtosis then click continue and then Ok..

۴-تبدیل داده ها:

تبدیل داده ها برای بدست اوردن شکلهای استاندارد داده ها در تحلیل داده ها مفید است. سه روش ساده تبدیل در ادامه بیان میشود.

الف: مرکزی کردن: اگر $x_1, x_2, \ldots\ldots x_n$ **داده های آماری در دسترس باشند آنگاه تبدیل ساده** $x_i - \bar{x},\ i = 1,2,\ldots\ldots n$ **را داده های مرکزی شده نامند. توجه نمایید که در داده های مرکزی شده میانگین برابر صفر و واریانس تغییر نمی کند.**

ب: استاندارد سازی: اگر $x_1, x_2, \ldots\ldots x_n$ **داده های آماری در دسترس باشند آنگاه با تبدیل ساده** $z_i = \frac{x_i - \bar{x}}{s},\ i = 1,2,\ldots\ldots n$ **داده ها به داده هایی با میانگین صفر و واریانس ۱ تبدیل می شوند که در بسیاری موارد در تحلیل ها از این داده ها استفاده می شود. داده های جدید را داده های استاندارد شده نامند.**

پ: نرمال سازی: اگر داده های مورد مطالعه از توزیع نرمال پیروی نکنند در این صورت با استفاده از یک تبدیل مناسب داده به داده های با توزیع نرمال تبدیل میشوند. یک تبدیل کلی برای نرمال سازی داده ها تبدیل باکس و کاکس است که به صورت $y_i = \frac{x_i^{\lambda} - 1}{\lambda},\ i = 1,2,\ldots\ldots n$ **تعریف می شود. برخی از این تبدیلات در جدول زیر مشاهده میشود:**

λ	-۱	-۱/۲	۰	۱/۲	۱
y_i	$\frac{1}{x_i}$	$\frac{1}{\sqrt{x_i}}$	$\ln(x_i)$	$\sqrt{x_i}$	x_i

مثال ۱- به منظور کنترل عملیات خاکی در یک محل، آزمایش دانسیته در محل در ۴۰ نقطه انجام شده است. نتایج دانسیته خشک در محل و رطوبت معادل آن ، مطابق جدول زیر می باشد . مطلوب است :

۱- رسم جدول فراوانی داده ها و خلاصه نمودن داده ها

۲- تعیین شاخص های مرکزی شامل میانگین ، میانه ، مد و شاخص های پراکندگی نظیر انحراف معیار ، واریانس، دامنه ، حداکثر و حداقل و همچنین چگونگی توزیع نمودار با توجه به مقادیر کشیدگی و چولگی رسم نمودار هیستوگرام

نتایج دانسیته در محل

ضخامت خاک نباتی	درصد رطوبت (٪)	دانسیته خشک (gr/cm^3)	سمت	محل آزمایش (km)

۹۰.۰۰	۱.۹۰	۲.۹۰	right	۰+۰۰۰
۷۰.۰۰	۱.۸۷	۲.۶۰	center	۰+۱۰۰
۶۰.۰۰	۲.۰۸	۲.۲۰	left	۰+۲۰۰
۵۵.۰۰	۲.۰۳	۲.۶۰	center	۰+۳۰۰
۵۶.۰۰	۱.۸۸	۳.۴۰	right	۰+۴۰۰
۶۰.۰۰	۱.۵۹	۴.۶۰	center	۰+۵۰۰
۶۵.۰۰	۱.۷۸	۳.۳۰	left	۰+۶۰۰
۴۵.۰۰	۱.۸۴	۱.۱۰	center	۰+۷۰۰
۵۰.۰۰	۲.۰۶	۲.۵۰	right	۰+۸۰۰
۳۰.۰۰	۱.۹۰	۲.۸۰	center	۰+۹۰۰
۸۰.۰۰	۲.۰۸	۲.۱۰	left	۱+۰۰۰
۲۰.۰۰	۱.۹۹	۲.۶۰	center	۱+۱۰۰
۲۰.۰۰	۱.۸۶	۱.۴۰	right	۱+۲۰۰
۳۵.۰۰	۱.۸۲	۲.۸۰	center	۱+۳۰۰
۴۵.۰۰	۱.۸۸	۲.۴۰	left	۱+۴۰۰
۶۰.۰۰	۱.۹۲	۲.۴۰	center	۱+۵۰۰
۴۵.۰۰	۱.۸۱	۳.۷۰	right	۱+۶۰۰
۱۰.۰۰	۱.۸۰	۲.۶۰	center	۱+۷۰۰
۲۵.۰۰	۱.۸۱	۳.۱۰	left	۱+۸۰۰
۲۵.۰۰	۱.۹۳	۴.۱۰	center	۱+۹۰۰
۳۵.۰۰	۱.۹۵	۲.۴۰	right	۲+۰۰۰
۴۵.۰۰	۱.۷۶	۵.۴۰	center	۲+۱۰۰
۲۰.۰۰	۱.۹۸	۲.۵۰	left	۲+۲۰۰
۵۵.۰۰	۱.۹۶	۲.۱۰	right	۲+۳۰۰
۵۰.۰۰	۲.۰۳	۲.۹۰	center	۲+۴۰۰
۲۰.۰۰	۱.۹۹	۲.۷۰	left	۲+۵۰۰
۳۵.۰۰	۱.۸۳	۱.۷۰	center	۲+۶۰۰

۲+۷۰۰	left	۲.۳۰	۱.۸۱	۱۰.۰۰
۲+۸۰۰	right	۳.۰۰	۱.۸۹	۱۵.۰۰
۲+۹۰۰	center	۲.۶۰	۲.۰۲	۳۵.۰۰
۳+۰۰۰	left	۴.۳۰	۱.۸۹	۲۵.۰۰
۳+۱۰۰	right	۴.۱۰	۱.۷۹	۵۵.۰۰
۳+۲۰۰	left	۳.۸۰	۱.۵۲	۲۵.۰۰
۳+۳۰۰	center	۳.۶۰	۱.۷۷	۲۰.۰۰
۳+۴۰۰	left	۴.۳۰	۱.۸۳	۳۵.۰۰
۳+۵۰۰	right	۲.۸۰	۱.۹۶	۴۰.۰۰
۳+۶۰۰	right	۲.۳۰	۱.۶۸	۲۰.۰۰
۳+۷۰۰	center	۳.۰۰	۱.۸۱	۱۵.۰۰
۳+۸۰۰	left	۵.۴۰	۱.۸۶	۲۰.۰۰
۳+۹۰۰	right	۳.۶۰	۱.۶۷	۲۵.۰۰

مثال۲: یک نوع رسوب را به روش غربال کردن آنالیز کرده و نتایج بدست آمده در جدول زیر ارائه شده است. با استفاده از این اطلاعات، شاخ ص های زیر را محاسبه کنید.

الف) میانگین، میانه ،مد، انحراف معیار، مینیمم، ماکزیمم، بازه اطمینان برای میانگین، چولگی و کشیدگی را تعیین کنید.

اندازه غربال ها به فی	وزن ذرات در هر غربال
-۰/۵	۰/۱۵۲
-۰/۵	۰/۱۵۲
-۰/۵	۰/۱۵۲
-۰/۵	۰/۱۵۲
-۰/۵	۰/۱۵۲
-۰/۵	۰/۱۵۲

-۰/۵	۰/۱۵۲
-۰/۵	۰/۱۵۲
-۰/۵	۰/۱۵۲
-۰/۵	۰/۱۵۲
-۰/۵	۰/۱۵۲

مثال ۳. در یک آزمایش درصد چروک خورده گی ۵۰ نمونه سفال خشک اندازه گیری شده است داده ها عبارتند از.

۱۵,۶	۱۸,۵	۲۳,۱	۲۱,۲	۱۸,۲
۱۳,۴	۲۱,۲	۱۵,۴	۱۹,۴	۲۰,۸
۱۴,۳	۱۹,۶	۱۸,۲	۱۸,۷	۱۶,۴
۲۰,۲	۱۷,۸	۱۷,۶	۲۴	۱۶,۶
۱۶,۶	۲۰,۳	۱۷,۵	۲۳,۶	۱۷,۴
۱۳,۹	۲۱,۲	۱۹,۳	۱۸,۵	۱۹,۳
۱۸,۴	۲۲,۳	۱۷,۶	۱۹	۲۰,۵
۲۰,۱	۲۰,۳	۲۱,۴	۲۰,۴	۲۱,۵
۲۰,۵	۱۹,۷	۱۴,۸	۲۰,۶	۱۹,۶
۱۷	۲۳,۱	۱۵,۸	۲۰,۸	۱۸

الف- داده ها را در ۵ طبقه تقسیم نمایید

ب- نمودارهای ساقه و برگ و هیستو گرام را رسم نمایید.

ج- شاخصهای مرکزی و پراکنده گی را محاسبه نمایید.

۴-به منظور ارزیابی مقاومت فشاری تک محوری نمونه‌های مغزه در محل سد جامیشان، بر روی ۴۸ نمونه آزمون تک محوری نمونه سنگ انجام گردیده است. مطلوب است محاسبه : جدول فراونی داده ها، شاخص های مرکزی شامل میانگین ، میانه ، مد و شاخص های پراکندگی نظیر انحراف معیار ، واریانس، دامنه ، حداکثر و حداقل برای داده حاصل از آزمایش، همچنین چگونگی توزیع نمودار با توجه به مقادیر کشیدگی و چولگی و رسم نمودار هیستوگرام.

شماره نمونه	UCS (kg/cm2)	شماره نمونه	UCS (kg/cm2)
۱	۴۲۵	۲۵	۲۰۳
۲	۶۸۰	۲۶	۱۷۸
۳	۲۶۰	۲۷	۲۸۵
۴	۶۳۰	۲۸	۲۹۰
۵	۶۳۰	۲۹	۳۴۶
۶	۱۲۸	۳۰	۹۲
۷	۲۶۰	۳۱	۲۰۳
۸	۳۶۰	۳۲	۹۷
۹	۳۸۰	۳۳	۳۰۵
۱۰	۳۸۰	۳۴	۲۰۹
۱۱	۳۴۰	۳۵	۱۱۲
۱۲	۹۴	۳۶	۱۷۸
۱۳	۲۱۷	۳۷	۴۸۳
۱۴	۳۵۸	۳۸	۳۵۶
۱۵	۷۵۲	۳۹	۱۹۳
۱۶	۱۴۲	۴۰	۱۷۸
۱۷	۲۹۷	۴۱	۱۹۳
۱۸	۳۵۳	۴۲	۲۸۰
۱۹	۵۳۳	۴۳	۹۲
۲۰	۳۴۶	۴۴	۲۸۰
۲۱	۱۴۰	۴۵	۲۵۴
۲۲	۲۷۰	۴۶	۳۳۱
۲۳	۱۷۸	۴۷	۴۵۸
۲۴	۱۰۱	۴۸	۶۱