# به نام خدا


نام دانشجو: فاطمه صادقیان

موضوع: بررسی عملکرد آزمون های ناپارامتری برای مقایسه گروه‌ها

استاد راهنما:

دکتر حامد لروند

و پویا علی نیان

فهرست

مقدمه :

در تحلیل داده‌ها، انتخاب روش آماری مناسب برای آزمون فرضیه‌ها نقش کلیدی در نتیجه‌گیری دقیق و معتبر دارد. آزمون‌های آماری به دو دسته اصلی پارامتری و ناپارامتری تقسیم می‌شوند. آزمون‌های پارامتری معمولاً با فرض‌هایی مانند نرمال بودن توزیع داده‌ها و وجود میانگین و واریانس مشخص انجام می‌شوند، در حالی که آزمون‌های ناپارامتری انعطاف‌پذیری بیشتری دارند و می‌توانند در شرایطی که فرضیات پارامتری برقرار نیست، مورد استفاده قرار گیرند. با این حال، مقایسه این دو رویکرد از نظر دقت، قدرت آماری و کاربرد در داده‌های واقعی، موضوعی مهم و مورد توجه پژوهشگران است. در این پروژه، با بررسی چندین آزمون ناپارامتری، معادل پارامتری آن‌ها را نیز معرفی و تحلیل خواهیم کرد تا تفاوت‌ها و شباهت‌های این روش‌ها در شرایط مختلف داده‌ای مشخص شود. هدف این مطالعه، ارائه دیدگاهی جامع برای انتخاب بهینه روش آماری با توجه به ویژگی‌های داده‌ها و اهداف تحلیلی است. پس از معرفی این آزمون‌ها، با استفاده از شبیه‌سازی، سه آزمون تی (پارامتری)، ویلکاکسون و علامت (ناپارامتری) را در شرایط مختلف داده‌ای، از جمله وجود یا عدم وجود چولگی و حضور داده‌های پرت، مقایسه خواهیم کرد و در نهایت بر اساس نتایج به نتیجه‌گیری جامع دست خواهیم یافت.

در ادامه به معرفی آزمون های ناپارامتری و پارامتری می پردازیم.

## آزمون نشانه یک نمونه ای (one sample sign_test)

فرض کنید یک نمونه ی تصادفی n تایی $X_1, X_2, \dots, X_n$ از توزیعی نامعلوم داریم و هدف انجام آزمون زیر است.

$$\begin{cases} H_0 : Q_p = a \\ H_1 : Q_p > a \end{cases}$$

این آزمون مقدار چندک مرتبه p ام با عددی از پیش تعیین شده را می سنجد. تحت فرض صفر برای هر $X_i$ داریم :

$$P_{H_0}(X_i > a) = P_{H_0}(X_i > Q_p) = 1 - p$$

اگر B شماره $X_i$ هایی باشد که از a بزرگترند. آماره B تحت فرض صفر دارای توزیع دوجمله ای Bin(n , 1-p) می باشد. هر کدام از تفاضل های زیر دارای نشانه های + یا – می باشد.

$X_1 - a , X_2 - a , \dots, X_n - a$

B برابر است با شماره + ها. ناحیه بحرانی آزمون به صورت $B \geq k$ که در آن k کوچکترین عدد صحیحی است که برای آن داریم $P( B \geq k \leq \alpha)$

معادل پارامتری این آزمون، آزمون t-test است :

## آزمون t تک نمونه ای (one sample t-test)

فرض کنید یک نمونه ی تصادفی n تایی $X_1, X_2, \dots, X_n$ از توزیعی نرمال داریم (یا توزیع نامعلوم و $n>30$)و هدف انجام آزمون زیر است.

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

این آزمون برابری میانگین با میانگین از پیش تعیین را می آزماید. آماره آزمون t به صورت $T = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$ تعریف می شود. که در آن $\bar{X}$ میانگین نمونه است، $\mu_0$ مقدار میانگین از پیش تعیین شده است، S انحراف معیار نمونه است، n حجم نمونه است و درجه آزادی آن برابر n-۱ است. فرض صفر رد می شود هر گاه داشته باشیم : $|T| > t_{1-\frac{\alpha}{2}}(n-1)$

باید توجه داشته باشیم که برای استفاده از آزمون t داده ها باید از توزیع نرمال پیروی کنند و وقتی که می خواهیم برای جامعه ای که توزیع آن مشخص نیست از این آزمون استفاده کنیم باید نمونه به اندازه کافی بزرگ باشد تا بتوانیم از قضیه حد مرکزی استفاده کنیم. (n>25)

## آزمون تک نمونه ای ویلکاکسون

فرض کنید یک نمونه ی تصادفی n تایی $X_1, X_2, \dots, X_n$ از توزیعی نامعلوم داریم و هدف انجام آزمون زیر است.

$$\begin{cases} H_0 : m = m_0 \\ H_1 : m \neq m_0 \end{cases}$$

آزمون بررسی می کند که آیا میانه با مقدار مورد نظر تفاوت دارد یا خیر. فرض می کنیم مشاهدات نمونه دارای مقادیر مطلق متمایز هستند و هیچ مشاهده برابری ندارند. اگر رتبه های $|X_1|, |X_2|, \dots, |X_n|$ به ترتیب برابر با $R_1, R_2, \dots, R_n$ باشد، آماره این آزمون بصورت $T = \sum_{i=1}^{n} sgn(X_i)R_i$ تعریف می شود. این آماره ارتباط نزدیکی با دو آماره $T^+ = \sum_{1\le i\le n, X_i>0} R_i$ و $T^- = \sum_{1\le i\le n, X_i<0} R_i$ دارد. به عنوان نمونه داریم : $T = T^+ - T^- = 2T^+ - \frac{n(n+1)}{2} = \frac{n(n+1)}{2} - 2T^-$

فرض صفر رد می شود اگر آماره $W = \min(T^+, T^-)$ خیلی کوچک باشد.

معادل پارامتری ای آزمون، آزمون t تک نمونه ای است که قبلا به شرح آن پرداخته ایم.

## آزمون جمعی_رتبه ای ویلکاکسون

با توجه به نمونه های تصادفی مستقل $X_1, X_2, \dots, X_n \sim F(x)$ و $Y_1, Y_2, \dots, Y_m \sim G(y)$ رابطه $G(y) = F(y - c)$ را در نظر می گیریم و هدف انجام آزمون زیر است.

$$\begin{cases} H_0 : c = 0 \\ H_1 : c > 0 \end{cases}$$

می توان نشان داد رابطه $Y = X + c$ برقراراست. اگر فرض صفر برقرار باشد به این معناست که X و Y هم توزیع هستند و درنتیجه میانگین های برابر دارند. c را پارامتر تغییر مبدا می نامند. اگر $R_i$ جمع رتبه های $X_i$ها و $S_i$ جمع رتبه های $Y_i$ ها باشند فرض صفر را رد می کنیم هر گاه برای $W_s = \sum_{i=1}^{n} S_i$ داشته باشیم $\{W_s \ge k\}$

معادل پارامتری این آزمون، آزمون t زوجی است :

## آزمون t زوجی (paired sample t-test)

با توجه به نمونه های تصادفی $X_1, X_2, \dots, X_n \sim F(x)$ و $Y_1, Y_2, \dots, Y_m \sim G(y)$ رابطه $G(y) = F(y - c)$ که مشاهدات X و Y به هم وابسته اند هدف انجام آزمون زیر است.

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$$

در این آزمون یکسان بودن میانگین دو جامعه را بررسی می کنیم.

اماره آزمون t به صورت $T = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_0/\sqrt{n}}$ تعریف می شود.که در آن $\bar{X} - \bar{Y}$ میانگین اختلاف های جفت هاست، $S_0$ انحراف معیار اختلاف های جفت هاست، n تعداد جفت ها است و درجه آزادی آن برابر با n-1 است. فرض صفر رد می شود اگر داشته باشیم $|T| > t_{1-\frac{\alpha}{2}}(n-1)$. می دانیم که برای استفاده از آزمون t-test داده ها باید از توزیع نرمال پیروی کنند یا اندازه ی نمونه بزرگ باشد. (n>۲۵)

## آزمون من-ویتنی

با فرض $X_1, X_2, \dots, X_n \sim F(x)$ و $Y_1, Y_2, \dots, Y_m \sim G(y)$ و در نظر گرفتن $G(y) = F(y - c)$ هدف انجام آزمون زیر است

$$\begin{cases} H_0 : c = 0 \\ H_1 : c > 0 \end{cases}$$

نمادهای زیر را در نظر میگیریم:

$W_{xy}$ :تعداد زوج مرتب هایی که در آن $x_i < y_i$

$W_{yx}$ :تعداد زوج مرتب هایی که در آن $x_i > y_i$

و ناحیه بحرانی آن به صورت { Wxy ≤ k } است. آزمون من-ویتنی و آزمون جمعی-رتبه ای ویلکاکسون معادل می باشند و نتایج یکسانی دارند. اگر فرض صفر برقرار باشد به این معناست که X و Y هم توزیع هستند و درنتیجه میانگین های برابر دارند.

آزمون پارامتری معادل آن آزمون t دونمونه ای مستقل است :

## آزمون t دو نمونه ای مستقل (independent sample t-test)

فرض کنید $X_1, X_2, \dots, X_n \sim F(x)$ و $Y_1, Y_2, \dots, Y_m \sim G(y)$ و دو نمونه از هم مستقل باشند و هدف انجام آزمون زیر است.

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

این آزمون برابری میانگین در دو جامعه را می آزماید. آماره این آزمون به صورت $T = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_p/(\sqrt{n}+\sqrt{m})}$ تعریف می شود. که در آن $\bar{X} - \bar{Y}$ میانگین اختلاف های جفت هاست، $S_p$ انحراف معیار ترکیبی است ($S_p = \frac{((n-1)S_x^2+(m-1)S_y^2}{n+m-2}$) و درجه آزادی برابر است با n+m-۲. برای به کارگیری آزمون t داده ها باید از توزیع نرمال پیروی کنند و یا اندازه داده ها بزرگ باشد. فرض صفر رد می شود اگر داشته باشیم $|T| > t_{1-\frac{\alpha}{2}}(n-1)$

## آزمون نشانه با نمونه ای از دوتایی ها

دو تایی های $(X_1,Y_1),...,(X_n,Y_n)$ را در نظر می گیریم و هدف انجام آزمون زیر است.

$$\begin{cases} H_0 : c = 0 \\ H_1 : c > 0 \end{cases}$$

اگر فرض صفر برقرار باشد یعنی Y-X نسبت به صفر متقارن است. اگر تعداد $Y_i - X_i$ ها را با B نشان دهیم داریم : $B \sim Bin(n,1/2)$
فرض صفر رد می شود هر گاه داشته باشیم $B \geq K$

معادل پارامتری آن آزمون paired t-test است که قبلا به شرح آن پرداختیم.

## آزمون رتبه ای-نشانه ای ویلکاکسون

دو تایی های $(X_1,Y_1),...,(X_n,Y_n)$ که دو متغیر به هم وابسته اند را در نظر می گیریم و هدف انجام آزمون زیر است.

$$\begin{cases} H_0 : c = 0 \\ H_1 : c > 0 \end{cases}$$

$Z_i = Y_i - X_i$ را در نظر بگیرید. در این آزمون از نشانه و همچنین رتبه $Z_i$ ها کمک می گیریم . تحت فرض صفر توزیع $Z_i$ ها نسبت به صفر متقارن هستند.اگر رتبه های $|Z_1|, |Z_2|, ..., |Z_n|$ به ترتیب برابر با $R_1, R_2, ..., R_n$ باشد. آماره آزمون به صورت $W = \sum_{r=1}^{n} u_i R_i$ تعریف می شود . ناحیه بحرانی برابر با $W > K$ است. $W_P$ برابر با مجموع رتبه های جملات مثبت W می باشد.

$$W_p = \frac{W}{2} + \frac{(n+1)n}{4}$$

$W_N$ برابر با مجموع رتبه های جملات منفی W است.

$$W_N = \frac{-W}{2} + \frac{(n+1)n}{4}$$

معادل پارامتری این آزمون paired t-test است که قبلا به شرح و توضیح آن پرداختیم.

## آزمون اسپیرمن

فرض کنید $(X_1 , Y_1),...,(X_n , Y_n)$ نمونه تصادفی از توزیعی دو بعدی باشند و هدف انجام آزمون زیر است:

$$\begin{cases} H_0 : \rho_1 = \rho_2 \\ H_1 : \rho_1 \neq \rho_2 \end{cases}$$

اگر رتبه $X_i$ ها را با $R_i$ نمایش دهیم و رتبه $Y_i$ ها را با $S_i$ نشان دهیم و $D_i = r_i - s_i$ باشد. ضریب همبستگی اسپیرمن $R_s$ به صورت زیر تعریف می شود:

$$R_s = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2 \sum_{i=1}^{n}(S_i - \bar{S})^2}}$$

$$R_s = 1 - \frac{6\sum_{i=1}^{n} D_i}{n(n^2-1)} \qquad -1 \le R_s \le 1$$

این آزمون برای سنجش میزان همبستگی دو متغیر بکار می رود و اگر فرض صفر درست باشد یعنی بین دو متغیر همبستگی وجود ندارد . فرض صفر رد می شود هر گاه داشته باشیم : $D^2 > k$

معادل پارامتری آن آزمون آزمون همبستگی پیرسون است :

## آزمون ضریب همبستگی پیرسون

یک نمونه تصادفی به صورت $(X_1 , Y_1),...,(X_n , Y_n)$ داریم و هدف انجام آزمون زیر است:

$$\begin{cases} H_0 : \rho = \rho_0 \\ H_1 : \rho \neq \rho_0 \end{cases}$$

برآورد ضریب همبستگی را با R نشان می دهیم :

$$R = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}}$$

فرض صفر زمانی رد می شود که : $|Z| > Z_{1-\frac{\alpha}{2}}$ این آزمون همبستگی بین دو متغیر را با مقدار از قبل تعیین شده ای می سنجد و تحت فرض صفر این دو مقدار با هم برابرند.

## آزمون ضریب همبستگی کندال

یک نمونه تصادفی به صورت $(X_1 , Y_1),...,(X_n , Y_n)$ داریم و هدف انجام آزمون زیر است:

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

تحت فرض صفر همبستگی X و Y صفر است .

برآورد نقطه ای کندال : $T = \frac{2P}{\binom{n}{2}} - 1$

که P در آن برابر است با $(y_j - y_i)$ *های مثبت* $i \le j \le n$ . فرض صفر رد می شود هر گاه : $|T| > k$

معادل پارامتری این آزمون آزمون اسپیرمن است که قبلا آنرا شرح دادیم.

## آزمون تصادفی بودن داده های عددی

نمونه $X_1,...,X_n$ را در نظر بگیرید این آزمون بررسی می کند که آیا این نمونه تصادفی است یا خیر. میانه داده ها را با $m$ نشان می دهیم . $x_i < m$ را با ۰ و $x_i \geq m$ را با ۱ مشخص می کنیم. فرض صفر زمانی رد می شود که تعداد دو ها خیلی کم یا زیاد باشد. یعنی ناحیه بحرانی به صورت $R \geq k1$ یا $R \leq k2$ است.

این آزمون معادل پارامتری مناسبی ندارد.

## آزمون مساوی بودن دو توزیع

با فرض $X_1,X_2,...,X_n \sim F(x)$ و $Y_1,Y_2,...,Y_m \sim G(y)$ و مستقل بودن این دو نمونه، هدف انجام ازمون زیر است

$$\begin{cases} H_0 : F(t) = G(t) \\ H_1 : F(t) \neq G(t) \end{cases}$$

یافته های این دو نمونه را ترکیب می کنیم و به ترتیب صعودی مرتب می کنیم سپس در این نمونه مرتب به جای xi ها ۰ و به جای yi ها ۱ قرار می دهیم. ناحیه بحرانی به صورت $R \leq K$ است.

معادل پارامتری آن آزمون independent two-sample t-test است که قبلا به شرح و توزیح آن پرداختیم.

## آزمون $X^2$ برای نیکویی برازش

$$H_0 : \begin{cases} \text{مجموعه} : C1, \dots, Ck \\ \text{احتمال} : p1, \dots, pk \end{cases}$$

نتایج یک آزمایش تصادفی را به صورت رده های دو به دو جدای C۱ ,...,Ck مشاهده می کنیم احتمال مشاهده نتیجه این آزمایش را به صورت p۱ +...+ pk = ۱ نشان می دهیم.این آزمایش را n بار تکرار می کنیم. اگر Oi بار نتیجه را در Ci مشاهده کنیم. متغیر تصادفی Oi را فراوانی مشاهده شده می نامند. و همچنین داریم ei = npi. $V_k = \frac{\sum_{i=1}^{n}(o_i - e_i)}{e_i}$ آماره آزمون است و به آن آماره پیرسن می گویند. فرض صفر زمانی رد می شود که داشته باشیم :$V_k > X^2_{(1-\frac{\alpha}{2})}(k-1)$

معادل پارامتری این آزمون , آزمون نسبت درستنمایی است :

## آزمون نسبت درستنمایی

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_0^c \end{cases}$$

آماره این آزمون به صورت زیر تعریف می شود :

$$\lambda = \frac{(\sup \theta \in \Theta_0 \, L(\theta|x))}{(\sup \theta \in \Theta_0^c \, L(\theta|x))}$$

$L(\theta|x)$ تابع درستنمایی داده است. فرض صفر رد می شود هرگاه : $\lambda(x) \leq c$ این آزمون خوب بودن تناسب دو مدل آماری نسبت به احتمالات آنها ارزیابی می کند . اگر فرض صفر درست باشد این دو احتمال یکسان اند.

## مقایسه آزمون علامت و آزمون t

ابتدا بسته های مورد نیاز را نصب و فراخوانی می کنیم.

```
install.packages("BSDA")
library(BSDA)
library(moments)
```

تابع param را تعریف می کنیم. در این تابع نام توزیع، پارامتر ها و مقدار اختلاف آزمون از پارامتر اصلی را می گیرد و متناسب با آن میانه و میانگینی که باید در آزمون چک شود را به عنوان خروجی می دهد.

```
param=function(dist,parameter,diff){
if(dist=="exp"){
med1=qexp(0.5,parameter)
mu1=med1/log(2)
med=med1+diff
mu=med1/log(2)+diff
}else if(dist=="normal"){
med=mu=parameter[1]+diff
med1=mu1=parameter[1]
}else if(dist=="binom"){
mu=prod(parameter)+diff
med=qbinom(0.5,parameter[1],parameter[2])+diff
mu1=prod(parameter)
med1=qbinom(0.5,parameter[1],parameter[2])
}else if(dist=="pois"){
mu=parameter+diff
med=qpois(0.5,parameter)+diff
mu1=parameter
med1=qpois(0.5,parameter)
}else if(dist=="geom"){
```

```
mu=1/parameter+diff
med=qgeom(0.5,parameter)+diff
mu1=1/parameter
med1=qgeom(0.5,parameter)
}else if(dist=="unif"){
mu=mean(parameter)+diff
med=qunif(0.5,parameter[1],parameter[2])+diff
mu1=mean(parameter)
med1=qunif(0.5,parameter[1],parameter[2])
}else if(dist=="chisq"){
mu=parameter+diff
med=qchisq(0.5,parameter)+diff
mu1=parameter
med1=qchisq(0.5,parameter)
}else if(dist=="t"){
mu=-diff
med=-diff
mu1=0
med1=0
}else if(dist=="lnorm"){
 mu=parameter[1]+diff
 med=exp(parameter[1])+diff
 mu1=parameter[1]
 med1=exp(parameter[1])
}else if(dist=="F"){
 mu=parameter[2]/(parameter[2]-2)+diff
 med=qf(0.5,parameter[1],parameter[2])+diff
 mu1=parameter[2]/(parameter[2]-2)
```

```
  med1=qf(0.5,parameter[1],parameter[2])
}
out=c(mu=mu,med=med,mu1=mu1,med1=med1)
return(out)
}
```

تابع randsamp را تعریف می کنیم. این تابع نیز نمونه تصادفی به اندازه ی n از توزیع با پارامتر های داده شده را به عنوان خروجی می دهد.

```
randsamp=function(dist,parameter,n){
if(dist=="exp"){
x=rexp(n,parameter)
}else if(dist=="normal"){
x=rnorm(n,parameter[1],parameter[2])

}else if(dist=="binom"){
x=c(1,1)
while(var(x)==0){
x=rbinom(n,parameter[1],parameter[2])}
}else if(dist=="pois"){
x=c(1,1)
while(var(x)==0){
x=rpois(n,parameter)
}
}else if(dist=="geom"){
x=c(1,1)
while(var(x)==0){
x=rgeom(n,parameter)}
```

```
}else if(dist=="unif"){
x=runif(n,parameter[1],parameter[2])
}else if(dist=="chisq"){
x=rchisq(n,parameter)
}else if(dist=="t"){
x=rt(n,parameter)
}else if(dist=="lnorm"){
 x=rlnorm(n,parameter[1],parameter[2])
}else if(dist=="F"){
 x=rf(n,parameter[1],parameter[2])
}
return(x)
}
```

تابع evaluation_ttest_sign را تعریف می کنیم. این تابع اسم توزیع، پارامتر ها و مقدار اختلاف مقدار واقعی پارامتر با مقدار بررسی در آزمون را می گیرد و تابع توان و خطای نوع اول را برای دو آزمون علامت و t را به صورت عددی و نموداری می دهد. همچنین برای هر اندازه نمونه نرمال بودن نمونه های تولید شده، میزان چولگی، تعداد داده های دورافتاده و قدر مطلق اختلاف خطای نوع اول با خطای نوع اول اسمی در آزمون t را می دهد.

```
evaluation_ttest_sign=function(dist,parameter,diff){
pvalue=c()
out1=out2=out3=out4=out5=out6=out7=c()
num1 = numeric(500)
num2 = numeric(500)
num3 = numeric(500)
mu=param(dist,parameter,diff)[1]
mu1=param(dist,parameter,diff)[3]
med=param(dist,parameter,diff)[2]
```

```
med1=param(dist,parameter,diff)[4]
number=c(5,10,15,20,25,30,35,40,50,60,70,80,90,100)
alpha=c(0.0625,0.109375000,1.184692e-01,1.153183e-01,
      1.077521e-01,9.873715e-02,8.953108e-02,8.069047e-02,
      1.189205e-01,9.246098e-02,1.196093e-01,9.291188e-02,
      1.133444e-01,8.862608e-02)+0.001
j=1
for(n in number){
z=y=w=s=0
for(i in 1:500){
x=randsamp(dist,parameter,n)
num1[i]=shapiro.test(x)$p.value
num2[i]=skewness(x)
num3[i]=length(boxplot(x,plot=FALSE)$out)
a=SIGN.test(x, md = med, alternative = "two.sided")
b=t.test(x,mu=mu,alternative = "two.sided")
e=SIGN.test(x, md = med1, alternative = "two.sided")
f=t.test(x,mu=mu1,alternative = "two.sided")
z=ifelse(b$p.value<alpha[j],z+1,z)
y=ifelse(a$p.value<alpha[j],y+1,y)
w=ifelse(e$p.value<alpha[j],w+1,w)
s=ifelse(f$p.value<alpha[j],s+1,s)
}
out1[j]=z/500
out2[j]=y/500
out3[j]=w/500
out4[j]=s/500
out5[j]=mean(num1)
```

```
out6[j]=mean(num2)
out7[j]=mean(num3)
j=j+1
}
out=list()
out$par=data.frame(number=number,"power ttest"=out1,
        "ttest error"=out4,
        "power signtest"=out2,
        "signtest error"=out3)
out$static=data.frame(number=number,mean_pvalue=out5,
            mean_skweness=out6,mean_out=out7,
            tdiff=abs(alpha-out4))
par(mar=c(4,4,2.5,0.5))
mm=c(out1,out2,out3,out4)

plot(number,out1,col="blue",type="l",ylim=c(min(mm),min(c(1,1.5*max(mm)))),
   main="مقایسه ی توان و خطای نوع اول برای آزمون تی و علامت",
   xlab="اندازه نمونه",ylab="توان یا خطای نوع اول",pch=1)
points(number,out4,col="red",type="l",pch=2)
points(number,out2,col="green",type="l",pch=3)
points(number,out3,col="black",type="l",pch=4)
legend("topleft",
    legend = c("توان آزمون تی",
          "خطای نوع اول آزمون تی",
          "توان آزمون علامت",
          "خطای نوع اول آزمون علامت"),
    col = c("blue", "red","green","black"),
```

```
    pch=c(1,2,3,4),
    bty = "o",cex=0.5)
return(out)
}
```

چون آزمون علامت ماهیت گسسته دارد (از توزیع دوجمله ای استفاده می کند) همه ی مقادیر p_value را پوشش نمی دهد و ممکن است در عمل به مقدار تنظیم شده دلخواه ما نرسد و محافظه کارانه تر عمل کرده و آزمون ما در سطح خطای نوع اول کمتری بررسی شود. از آنجا که با کاهش خطای نوع اول، خطای نوع دوم افزایش می یابد و در نتیجه توان آزمون کاهش می یابد پس تابع توان آزمون کمتر از مقدار واقعی نشان داده می شود. اما آزمون t که از توزیع نرمال استفاده می کند همه ی مقادیر p_value را در بر می گیرد. برای اینکه مقایسه دو آزمون عادلانه تر شود باید به ازای خطای نوع اول ثابت و یکسان، توان دو آزمون را مقایسه کنیم. پس به دلیل ماهیت گسسته ی آزمون علامت هر دو آزمون در سطح خطای نوع اول یکسانی برای مقایسه نیستند. یکی از کارهایی که می توان انجام داد استفاده از آزمون تصادفی شده است که آزمون گسسته ما در سطح خطای نوع اول دلخواه می رسد. در اینجا ما راه حل متفاوتی را امتحان کردیم. برای هر اندازه نمونه مقدار متفاوتی از $\alpha$ را در نظر گرفتیم؛ $\alpha$ را به گونه ای انتخاب کردیم که برای هر اندازه نمونه دقیقا آن مقادیر آزمون علامت باشد که نزدیکترین عدد به ۰.۱ است. با انجام این کار خطای نوع اول همان مقدار p_value دلخواه ما می شود و برای آزمون t هم به دلیل پیوسته بودن مشکلی پیش نخواهد آمد. توجه داریم که در اینجا مقایسه یک آزمون در اندازه نمونه های مختلف به علت متفاوت بودن خطای نوع اول قابل مقایسه نیست و جزو اهداف ما هم نبوده است. هدف این است که بررسی شود برای هر اندازه نمونه کدام آزمون توان بیشتری دارد

برای مقایسه آزمون ها چهار دسته از توزیع ها را بررسی می کنیم :

دسته اول : در این دسته توزیع هایی که چولگی کمی دارند (کمتر از یک) و داده های دور افتاده ندارند قرار می گیرند مانند توزیع نرمال و توزیع یکنواخت پیوسته.

دسته دوم : اعضای این گروه توزیع های بدون چولگی اند که داده دورافتاده دارند. مانند توزیع t-student با درجه آزادی کمتر از ۳

دسته سوم : گروهی از توزیع ها در این دسته قرار می گیرند که چولگی زیادی دارند اما داده دور افتاده ندارند. مانند توزیع نمایی و توزیع کای دو با درجه آزادی کمتر از ۳.۵

دسته چهارم : توزیع هایی که چولگی زیاد و همچنین داده های دورافتاده زیادی دارند در این دسته قرار می گیرند. مانند توزیع لگ_نرمال با واریانس بیش از ۰.۸ و توزیع F با درجه آزادی دوم کمتر از ۱۰

در ادامه نتایج شبیه سازی آمده است. ابتدا از هر گروه نمونه ای از خروجی R را آورده ایم که مقادیری همچون میانگین پی مقدار، میانگین چولگی، میانگین تعداد داده های دور افتاده و میزان تفاوت خطای نوع اول آزمون تی با مقدار خطای نوع اول اسمی که برای اندازه نمونه های متفاوت به تفکیک داده است.

$static

| | number | mean_pvalue | mean_skweness | mean_out | tdiff |
|---|---|---|---|---|---|
| 1 | 5 | 3.912320e-01 | 0.4700561 | 0.458 | 0.04850000 |
| 2 | 10 | 2.433182e-01 | 0.7996598 | 0.434 | 0.02962500 |
| 3 | 15 | 1.353992e-01 | 1.0021937 | 0.752 | 0.00946920 |
| 4 | 20 | 8.586690e-02 | 1.0715090 | 0.704 | 0.03168170 |
| 5 | 25 | 5.747227e-02 | 1.1352503 | 1.074 | 0.02524790 |
| 6 | 30 | 2.721807e-02 | 1.2363095 | 1.268 | 0.01226285 |
| 7 | 35 | 1.225173e-02 | 1.2939427 | 1.542 | 0.01546892 |
| 8 | 40 | 9.890001e-03 | 1.2847674 | 1.556 | 0.00369047 |
| 9 | 50 | 2.403112e-03 | 1.3647426 | 1.864 | 0.01007950 |
| 10 | 60 | 9.189854e-04 | 1.3815097 | 2.250 | 0.01946098 |
| 11 | 70 | 1.380886e-04 | 1.4434991 | 2.632 | 0.02339070 |
| 12 | 80 | 9.558641e-05 | 1.4617359 | 3.250 | 0.02008812 |
| 13 | 90 | 4.158417e-05 | 1.4325302 | 3.402 | 0.00834440 |
| 14 | 100 | 6.006240e-06 | 1.4871608 | 3.734 | 0.01237392 |

تصویر ۱

$static

| | number | mean_pvalue | mean_skweness | mean_out | tdiff |
|---|---|---|---|---|---|
| 1 | 5 | 0.5198929 | 0.0002706155 | 0.426 | 0.00550000 |
| 2 | 10 | 0.5125720 | 0.0014840812 | 0.310 | 0.00437500 |
| 3 | 15 | 0.5054146 | 0.0048583422 | 0.442 | 0.03146920 |
| 4 | 20 | 0.5058334 | -0.0173607651 | 0.310 | 0.02568170 |
| 5 | 25 | 0.5157823 | 0.0220205853 | 0.598 | 0.01475210 |
| 6 | 30 | 0.5110440 | 0.0022358766 | 0.372 | 0.02226285 |
| 7 | 35 | 0.4845486 | -0.0043620398 | 0.606 | 0.00546892 |
| 8 | 40 | 0.5223794 | 0.0069091049 | 0.488 | 0.01230953 |
| 9 | 50 | 0.5058641 | -0.0021052734 | 0.536 | 0.01592050 |
| 10 | 60 | 0.5016307 | 0.0059748929 | 0.650 | 0.00346098 |
| 11 | 70 | 0.5164488 | 0.0178502917 | 0.658 | 0.00660930 |
| 12 | 80 | 0.4762153 | -0.0036654746 | 0.850 | 0.00991188 |
| 13 | 90 | 0.5211397 | -0.0003550313 | 0.778 | 0.01565560 |
| 14 | 100 | 0.4879038 | 0.0123229004 | 0.996 | 0.00762608 |

تصویر ۲

$static

| | number | mean_pvalue | mean_skweness | mean_out | tdiff |
|---|---|---|---|---|---|
| 1 | 5 | 0.46206350 | -0.014658721 | 0.558 | 0.01750000 |
| 2 | 10 | 0.37142761 | -0.027106840 | 0.694 | 0.02837500 |
| 3 | 15 | 0.30350293 | -0.024792336 | 1.206 | 0.00946920 |
| 4 | 20 | 0.24882154 | 0.068732506 | 1.258 | 0.02831830 |
| 5 | 25 | 0.20143621 | -0.039220974 | 1.972 | 0.01475210 |
| 6 | 30 | 0.16151324 | 0.043819347 | 2.066 | 0.00373715 |
| 7 | 35 | 0.15165775 | -0.068901936 | 2.462 | 0.01946892 |
| 8 | 40 | 0.11865969 | 0.101999600 | 2.710 | 0.01569047 |
| 9 | 50 | 0.09552360 | 0.048890213 | 3.324 | 0.00807950 |
| 10 | 60 | 0.05197196 | -0.005969589 | 3.968 | 0.01346098 |
| 11 | 70 | 0.05024040 | 0.026321307 | 4.464 | 0.00860930 |
| 12 | 80 | 0.02807582 | -0.014052805 | 5.284 | 0.01991188 |
| 13 | 90 | 0.01855023 | 0.126526511 | 6.084 | 0.00834440 |
| 14 | 100 | 0.01126001 | 0.148414195 | 6.776 | 0.00762608 |

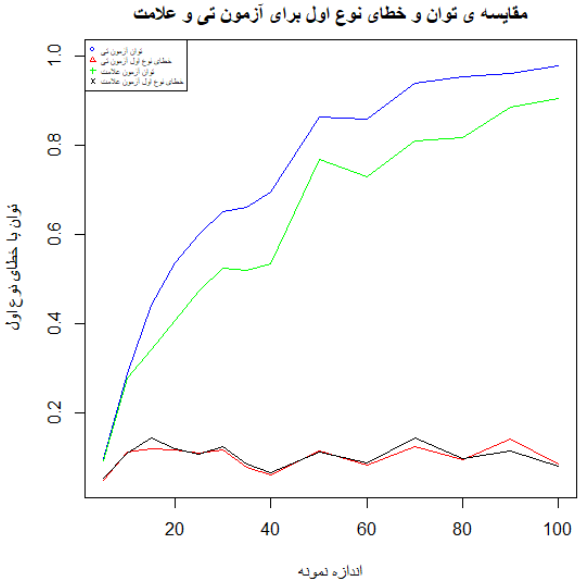
تصویر ۳

$static

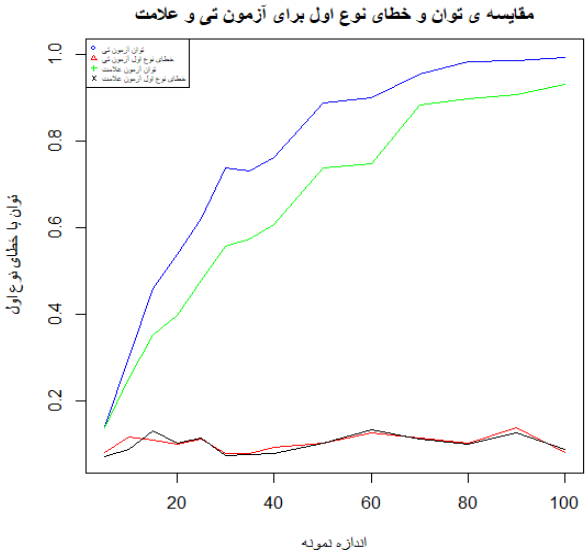
| | number | mean_pvalue | mean_skweness | mean_out | tdiff |
|---|---|---|---|---|---|
| 1 | 5 | 1.442135e-01 | 0.9796149 | 0.666 | 0.0665000 |
| 2 | 10 | 1.790930e-02 | 1.8059812 | 1.088 | 0.4696250 |
| 3 | 15 | 2.840012e-03 | 2.3243936 | 1.796 | 0.6645308 |
| 4 | 20 | 2.416512e-04 | 2.7752895 | 2.368 | 0.7076817 |
| 5 | 25 | 2.380927e-05 | 3.0685438 | 3.332 | 0.7472479 |
| 6 | 30 | 2.606279e-06 | 3.3863148 | 3.760 | 0.7562628 |
| 7 | 35 | 1.518454e-07 | 3.6358341 | 4.636 | 0.7594689 |
| 8 | 40 | 8.604297e-08 | 3.8539938 | 4.998 | 0.7643095 |
| 9 | 50 | 3.520981e-10 | 4.1992215 | 6.558 | 0.8240795 |
| 10 | 60 | 1.884706e-11 | 4.5308592 | 7.818 | 0.8505390 |
| 11 | 70 | 4.752151e-12 | 4.9695243 | 9.018 | 0.8313907 |
| 12 | 80 | 8.126244e-14 | 5.1927825 | 10.402 | 0.8580881 |
| 13 | 90 | 1.000993e-14 | 5.4923970 | 11.786 | 0.8596556 |
| 14 | 100 | 2.065764e-16 | 5.7909392 | 13.392 | 0.8643739 |

تصویر ۴

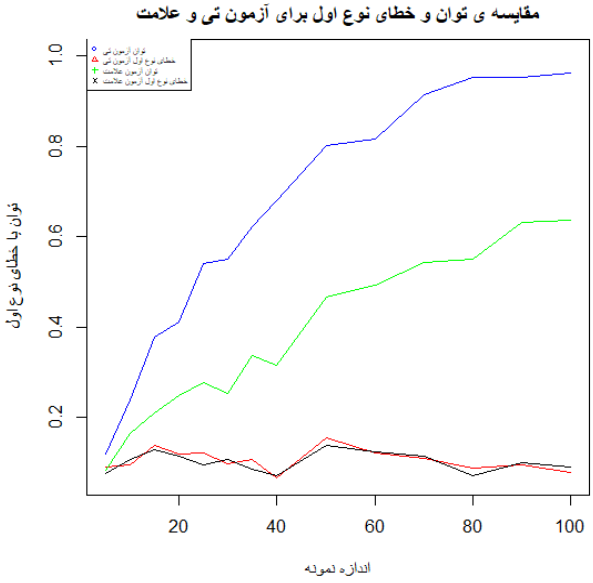
دسته اول (چولگی کم و بدون داده ی دور افتاده)



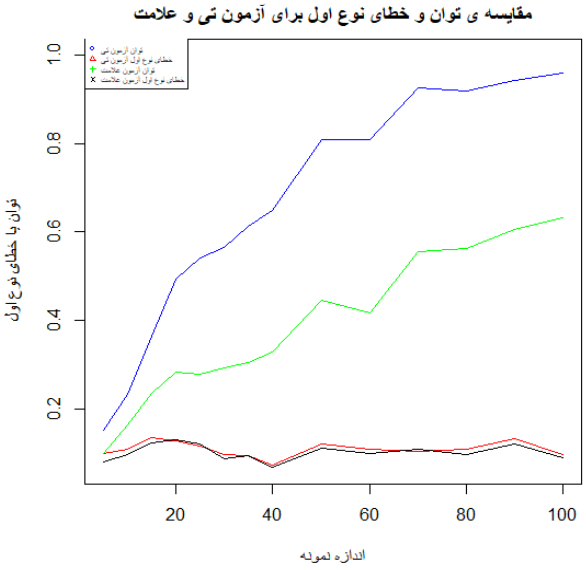

N(2,2)




N(2,5)



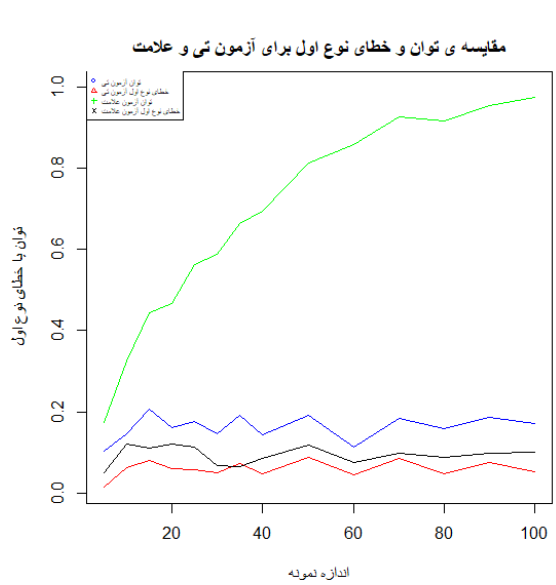

U(-3,7)



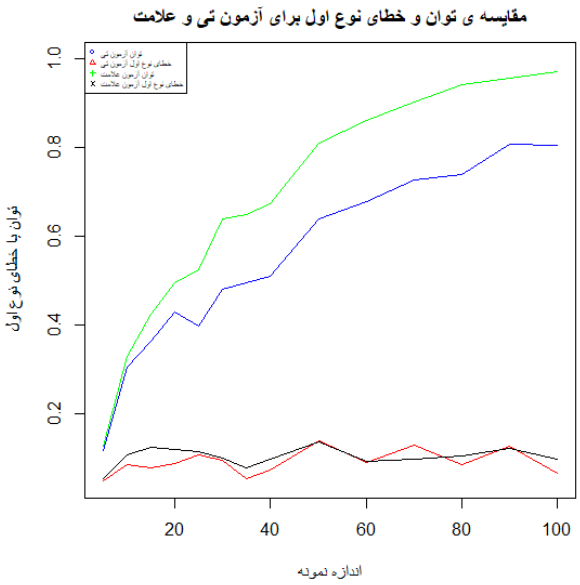

U(0,1)

نمودار های فوق که نمودار مشابه برای توزیع های این گروه هستند نشان می دهند که ابتدا خطای نوع اول دو آزمون تقریبا با هم برابرند پس می توان، توان آنها را به صورت عادلانه مقایسه کرد. توان آزمون آزمون t همواره بهتر از توان آزمون علامت عمل کرده است. به علاوه از آنجا که در این گروه توزیع یکنواخت وجود دارد نشان می دهد در شرایطی که حتی توزیع داده ها نرمال نیست ولی متقارن و بدون داده ی پرت است آزمون t توانایی بالایی دارد و می توان از آن استفاده کرد. برای تمامی موارد میانگین چولگی کمتر از یک شده و نشان می دهد داده های ما چولگی کمی دارند. میانگین داده های دورافتاده برای تمامی اندازه نمونه ها کمتر از یک بوده است پس می توان گفت داده های دور افتاده شانس بسیار کمی برای وقوع داشته اند. تصویر ۱ نماینده ی خروجی برای این گروه بوده است.

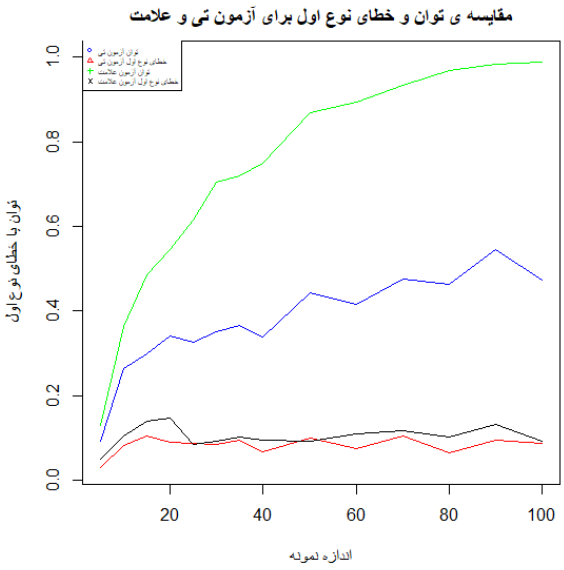
دسته دوم (چولگی کم ولی دارای داده ی دور افتاده)



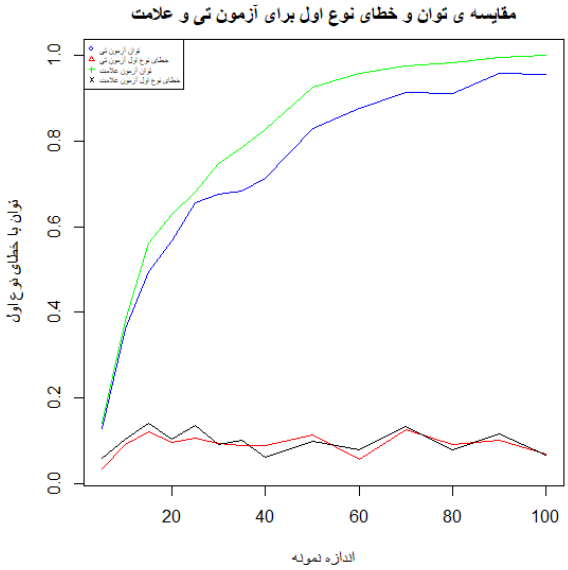

t(1)




t(2.5)



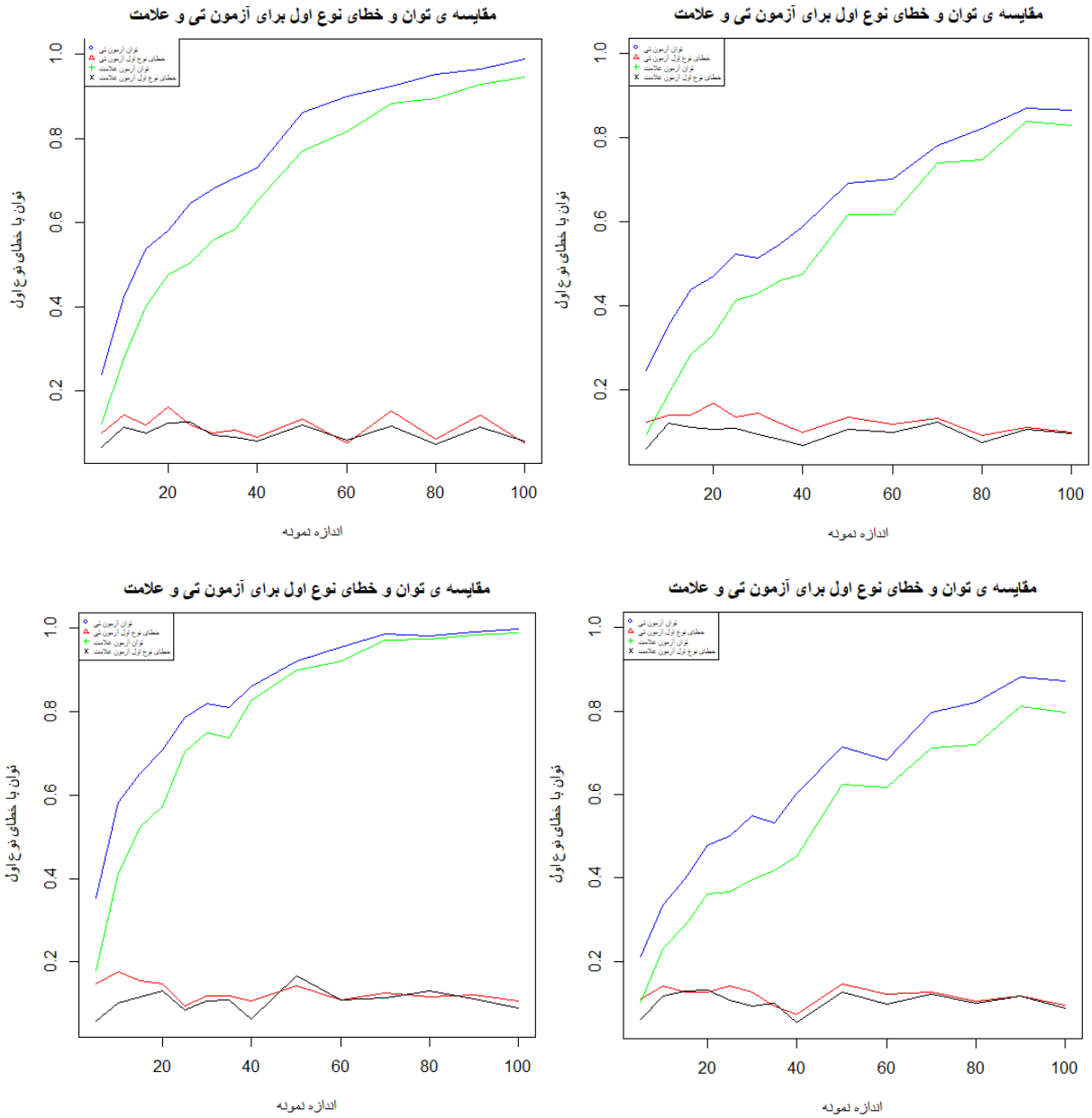

t(1.5)




t(3)

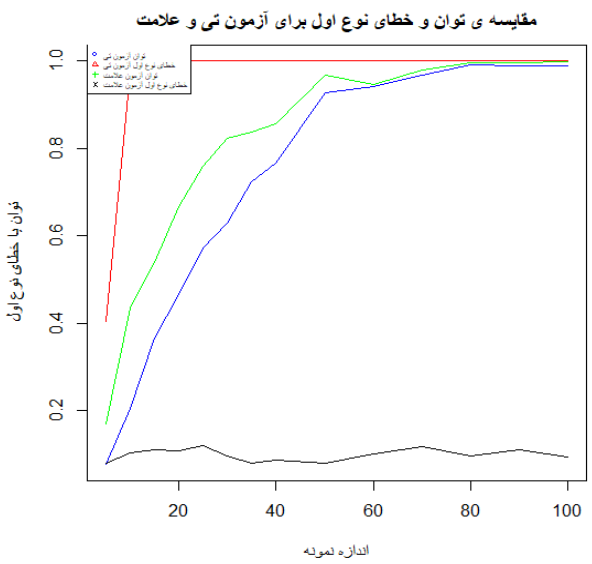
در تصاویر فوق که نمونه ای برای توزیع های این گروه هستند می بینیم که خطای نوع اول دو آزمون تقریبا نزدیک به یکدیگر هستند پس می توانیم دو آزمون را به طور عادلانه مقایسه کنیم. آزمون علامت در همه ی اندازه های نمونه بهتر عمل کرده و توان بالاتری نسبت به آزمون t داشته است. می توان گفت در توزیع هایی که متقارن است ولی داده دورافتاده دارند آزمون علامت مناسب تر از آزمون t است. میانگین چولگی در همه اندازه های نمونه کمتر از یک بوده است که چولگی کم داده ها را مشخص می کند. داده های دورافتاده نسبتا زیاد بوده اند، می دانیم وجود داده های دور افتاده توان آزمون t را کم می کند پس ممکن است دلیل عملکرد بدتر آزمون t در اینجا همین موضوع باشد. تصویر ۲ نماینده ی خروجی برای این گروه بوده است.

دسته سوم (چولگی زیاد و بدون داده ی دور افتاده)

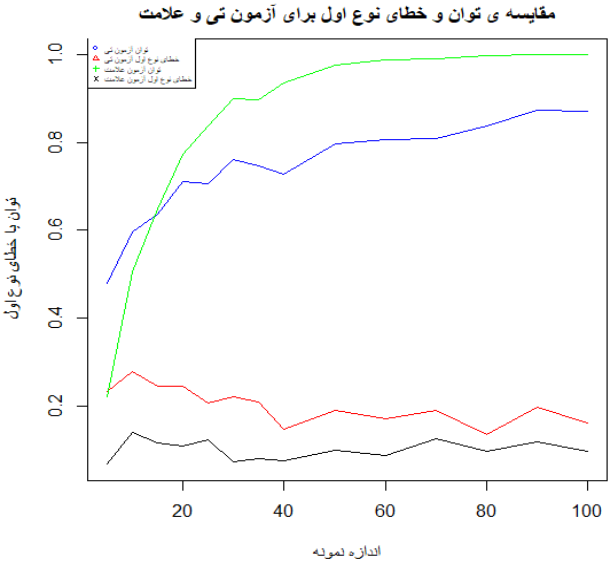



با مشاهده در نمودار های بالا که نمودار مشابه برای توزیع های این گروه هستند اولا می بینیم که خطای نوع اول دو آزمون تقریبا یکسان است پس می توانیم دو آزمون را به طور عادلانه مقایسه کنیم. آزمون t در اینجا بهتر عمل کرده است و توان بیشتری نسبت به آزمون علامت دارد. میانگین چولگی در اکثر مواقع بیشتر از یک است پس می توان گفت اعضای این دسته چولگی زیادی دارند. همچنین میانگین داده های دورافتاده در این دسته از توزیع ها زیاد نیست و تعداد داده های دورافتاده کمی دارند. فرض نرمال برای اعضای این گروه رد می شود، پس می توان گفت حتی مواقعی که داده ها از توزیع نرمال پیروی نمی کنند اما تعداد داده های دور افتاده کم است می توان از آزمون t استفاده کرد و این آزمون می تواند خوب عمل کند. تصویر ۳ نماینده ی خروجی برای این گروه بوده است.

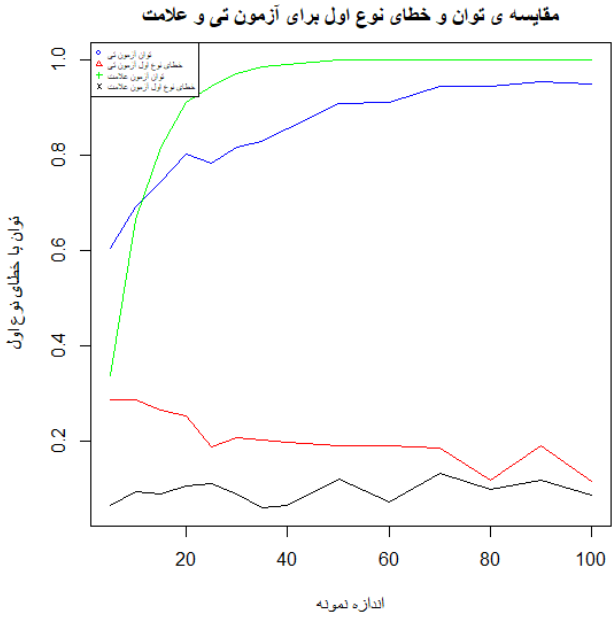
دسته چهارم (چولگی زیاد و دارای داده ی دور افتاده)



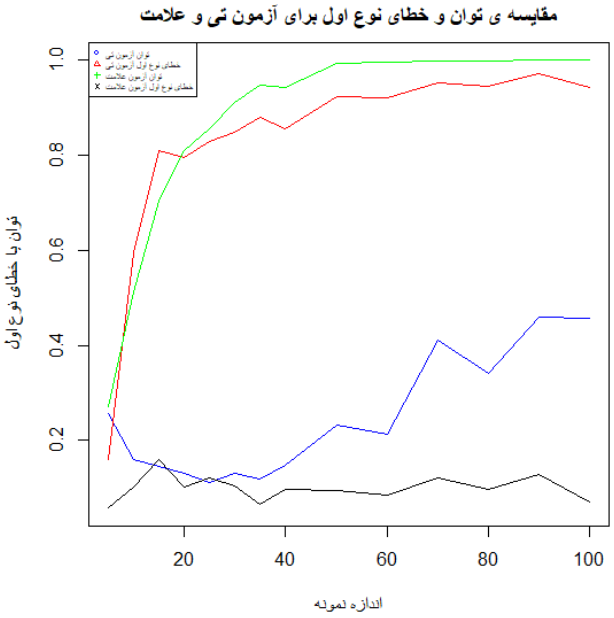

lnorm(2,0.9)




F(2,5)




F(1,6)




lnorm(0,2)

در نمودار های فوق که نمونه ای برای سایر توزیع های این گروه نیز هستند می بینیم که خطای نوع اول دو آزمون بشدت با هم اختلاف دارند پس در این حالت عادلانه نیست اگر بخواهیم توان دو آزمون را با هم مقایسه کنیم. همچنین برای اعضای مختلف این گروه نتایج یکسانی برای توان آزمون ها بدست نمی آید. در واقع زمانی که چولگی زیاد است در آزمون t خطای نوع اول در سطوحی که در نظر گرفتیم ثابت نمی شود. وجود داده های دورافتاده نیز توان آزمون t را می کاهد پس می توان گفت این آزمون در اینجا ضعیف عمل کرده است. برای اعضای این گروه فرض نرمال بودن رد می شود و داده ها از توزیع نرمال پیروی نمی کنند. همچنین این داده ها دارای داده های دورافتاده و چولگی زیادی هستند. تصویر ۴ نماینده ی خروجی برای این گروه بوده است.

## مقایسه آزمون علامت، ویلکاکسون تک نمونه ای و آزمون t

ابتدا بسته های مورد نیاز را نصب و فراخوانی می کنیم.

```
install.packages("BSDA")
library(BSDA)
library(moments)
```

تابع param را تعریف می کنیم. در این تابع نام توزیع، پارامتر ها و مقدار اختلاف آزمون از پارامتر اصلی را می گیرد و متناسب با آن میانه و میانگینی که باید در آزمون چک شود را به عنوان خروجی می دهد.

```
param=function(dist,parameter,diff){
 if(dist=="exp"){
  med1=qexp(0.5,parameter)
  mu1=med1/log(2)
  med=med1+diff
  mu=med1/log(2)+diff
 }else if(dist=="normal"){
  med=mu=parameter[1]+diff
  med1=mu1=parameter[1]
 }else if(dist=="binom"){
  mu=prod(parameter)+diff
  med=qbinom(0.5,parameter[1],parameter[2])+diff
  mu1=prod(parameter)
  med1=qbinom(0.5,parameter[1],parameter[2])
 }else if(dist=="pois"){
  mu=parameter+diff
  med=qpois(0.5,parameter)+diff
  mu1=parameter
  med1=qpois(0.5,parameter)
 }else if(dist=="geom"){
```

```
  mu=1/parameter+diff
  med=qgeom(0.5,parameter)+diff
  mu1=1/parameter
  med1=qgeom(0.5,parameter)
 }else if(dist=="unif"){
  mu=mean(parameter)+diff
  med=qunif(0.5,parameter[1],parameter[2])+diff
  mu1=mean(parameter)
  med1=qunif(0.5,parameter[1],parameter[2])
 }else if(dist=="chisq"){
  mu=parameter+diff
  med=qchisq(0.5,parameter)+diff
  mu1=parameter
  med1=qchisq(0.5,parameter)
 }else if(dist=="t"){
  mu=-diff
  med=-diff
  mu1=0
  med1=0
 }else if(dist=="lnorm"){
  mu=parameter[1]+diff
  med=exp(parameter[1])+diff
  mu1=parameter[1]
  med1=exp(parameter[1])
 }else if(dist=="F"){
  mu=parameter[2]/(parameter[2]-2)+diff
  med=qf(0.5,parameter[1],parameter[2])+diff
  mu1=parameter[2]/(parameter[2]-2)
```

```
  med1=qf(0.5,parameter[1],parameter[2])
 }
 out=c(mu=mu,med=med,mu1=mu1,med1=med1)
 return(out)
}
```

تابع randsamp را تعریف می کنیم. این تابع نیز نمونه تصادفی به اندازه ی n از توزیع با پارامتر های داده شده را به عنوان خروجی می دهد.

```
randsamp=function(dist,parameter,n){
if(dist=="exp"){
x=rexp(n,parameter)
}else if(dist=="normal"){
x=rnorm(n,parameter[1],parameter[2])

}else if(dist=="binom"){
x=c(1,1)
while(var(x)==0){
x=rbinom(n,parameter[1],parameter[2])}
}else if(dist=="pois"){
x=c(1,1)
while(var(x)==0){
x=rpois(n,parameter)
}
}else if(dist=="geom"){
x=c(1,1)
while(var(x)==0){
x=rgeom(n,parameter)}
```

```
}else if(dist=="unif"){
x=runif(n,parameter[1],parameter[2])
}else if(dist=="chisq"){
x=rchisq(n,parameter)
}else if(dist=="t"){
x=rt(n,parameter)
}else if(dist=="lnorm"){
 x=rlnorm(n,parameter[1],parameter[2])
}else if(dist=="F"){
 x=rf(n,parameter[1],parameter[2])
}
return(x)
}
```

تابع evaluation_ttest_sign_wilcox را تعریف می کنیم. این تابع اسم توزیع، پارامتر ها و مقدار اختلاف مقدار واقعی پارامتر با مقدار بررسی در آزمون را می گیرد و تابع توان و خطای نوع اول را برای آزمون های ویلکاکسون ، علامت و t را به صورت عددی و نموداری می دهد. همچنین برای هر اندازه نمونه نرمال بودن نمونه های تولید شده، میزان چولگی، تعداد داده های دورافتاده و قدرمطلق اختلاف خطای نوع اول با خطای نوع اول اسمی در آزمون t را می دهد.

```
evaluation_ttest_sign_wilcox=function(dist,parameter,diff){
 pvalue=c()
 out1=out2=out3=out4=out5=out6=out7=out8=out9=c()
 num1 = numeric(500)
 num2 = numeric(500)
 num3 = numeric(500)
 mu=param(dist,parameter,diff)[1]
 mu1=param(dist,parameter,diff)[3]
 med=param(dist,parameter,diff)[2]
```

```
med1=param(dist,parameter,diff)[4]
number=c(5,10,15,20,25,30,35,40,50,60,70,80,90,100)
alpha=c(0.0625,0.109375000,1.184692e-01,1.153183e-01,
      1.077521e-01,9.873715e-02,8.953108e-02,8.069047e-02,
      1.189205e-01,9.246098e-02,1.196093e-01,9.291188e-02,
      1.133444e-01,8.862608e-02)+0.001
j=1
for(n in number){
 z=y=w=s=q=v=0
 for(i in 1:500){
  x=randsamp(dist,parameter,n)
  num1[i]=shapiro.test(x)$p.value
  num2[i]=skewness(x)
  num3[i]=length(boxplot(x,plot=FALSE)$out)
  a=SIGN.test(x, md = med, alternative = "two.sided")
  b=t.test(x,mu=mu,alternative = "two.sided")
  h=wilcox.test(x,mu=mu,alternative = "two.sided")
  e=SIGN.test(x, md = med1, alternative = "two.sided")
  f=t.test(x,mu=mu1,alternative = "two.sided")
  g=wilcox.test(x,mu=mu1,alternative = "two.sided")
  z=ifelse(b$p.value<alpha[j],z+1,z)
  y=ifelse(a$p.value<alpha[j],y+1,y)
  w=ifelse(e$p.value<alpha[j],w+1,w)
  s=ifelse(f$p.value<alpha[j],s+1,s)
  q=ifelse(h$p.value<alpha[j],q+1,q)
  v=ifelse(g$p.value<alpha[j],v+1,v)
 }
 out1[j]=z/500
```

```
  out2[j]=y/500
  out3[j]=w/500
  out4[j]=s/500
  out5[j]=q/500
  out6[j]=v/500
  out7[j]=mean(num1)
  out8[j]=mean(num2)
  out9[j]=mean(num3)
  j=j+1
 }
 out=list()
 out$par=data.frame(number=number,"power ttest"=out1,
                    "ttest error"=out4,
                    "power signtest"=out2,
                    "signtest error"=out3,
                    "power wilcox"=out5,
                    "error wilcox"=out6)
 out$static=data.frame(number=number,mean_pvalue=out5,
                       mean_skweness=out6,mean_out=out7,
                       tdiff=abs(alpha-out4))
 par(mar=c(4,4,2.5,0.5))
 mm=c(out1,out2,out3,out4,out5,out6)


plot(number,out1,col="blue",xlim=c(0,100),type="l",ylim=c(min(mm),min(c(1,1.5*max(
mm)))),
     main="توان و خطای نوع اول آزمون های علامت، تی و ویلکاکسون",
     xlab="اندازه نمونه",ylab="توان یا خطای نوع اول",lty=1)
```

```
 points(number,out4,col="blue",type="l",lty=2)
 points(number,out2,col="green",type="l",lty=1)
 points(number,out3,col="green",type="l",lty=2)
 points(number,out5,col="red",type="l",lty=1)
 points(number,out6,col="red",type="l",lty=2)
 points(number,alpha,lty=1)
 legend("topleft",
     legend = c("توان تی",
                "خطای تی",
                "توان علامت",
                "خطای علامت",
                "توان ویلکاکسون",
                "خطای ویلکاکسون",
                "خطای نوع اول اسمی"
     ),
     col = c("blue", "blue","green","green","red","red","black"),
     lty=c(1,2,1,2,1,2,1),
     bty = "o",cex=0.75)
 return(out)
}
```

تنظیمات مانند کدهای قبلی است و فقط کدهای مربوط به آزمون ویلکاکسون هم اضافه شده است.

برای مقایسه آزمون ها هم از چهار دسته ی قبلی استفاده کردیم.

دسته اول : در این دسته توزیع هایی که چولگی کمی دارند (کمتر از یک) و داده های دور افتاده ندارند قرار می گیرند مانند توزیع نرمال و توزیع یکنواخت پیوسته.

دسته دوم : اعضای این گروه توزیع های بدون چولگی اند که داده دورافتاده دارند. مانند توزیع t-student با درجه آزادی کمتر از ۳

دسته سوم : گروهی از توزیع ها در این دسته قرار می گیرند که چولگی زیادی دارند اما داده دور افتاده ندارند. مانند توزیع نمایی و توزیع کای دو با درجه آزادی کمتر از ۳.۵

دسته چهارم : توزیع هایی که چولگی زیاد و همچنین داده های دورافتاده زیادی دارند در این دسته قرار می گیرند. مانند توزیع لگ_نرمال با واریانس بیش از ۰.۸ و توزیع F با درجه آزادی دوم کمتر از ۱۰

$static

| | number | mean_pvalue | mean_skweness | mean_out | tdiff |
|---|---|---|---|---|---|
| 1 | 5 | 0.114 | 0.070 | 0.4909535 | 0.00350000 |
| 2 | 10 | 0.320 | 0.094 | 0.5068747 | 0.00437500 |
| 3 | 15 | 0.408 | 0.106 | 0.5039206 | 0.00946920 |
| 4 | 20 | 0.554 | 0.114 | 0.5089019 | 0.00031830 |
| 5 | 25 | 0.636 | 0.112 | 0.5189537 | 0.00324790 |
| 6 | 30 | 0.696 | 0.104 | 0.5006396 | 0.01426285 |
| 7 | 35 | 0.754 | 0.094 | 0.5047189 | 0.00146892 |
| 8 | 40 | 0.762 | 0.078 | 0.4991683 | 0.00230953 |
| 9 | 50 | 0.912 | 0.112 | 0.5161366 | 0.00792050 |
| 10 | 60 | 0.906 | 0.102 | 0.5076161 | 0.00053902 |
| 11 | 70 | 0.952 | 0.120 | 0.5106503 | 0.00139070 |
| 12 | 80 | 0.972 | 0.076 | 0.4868347 | 0.01191188 |
| 13 | 90 | 0.976 | 0.116 | 0.5113994 | 0.00834440 |
| 14 | 100 | 0.992 | 0.082 | 0.4859508 | 0.00362608 |

تصویر ۱

$static

| | number | mean_pvalue | mean_skweness | mean_out | tdiff |
|---|---|---|---|---|---|
| 1 | 5 | 0.136 | 0.044 | 3.271993e-01 | 0.04950000 |
| 2 | 10 | 0.252 | 0.072 | 1.797456e-01 | 0.04837500 |
| 3 | 15 | 0.356 | 0.112 | 8.888685e-02 | 0.04346920 |
| 4 | 20 | 0.418 | 0.124 | 4.278331e-02 | 0.03031830 |
| 5 | 25 | 0.444 | 0.084 | 2.451474e-02 | 0.05675210 |
| 6 | 30 | 0.496 | 0.062 | 6.030982e-03 | 0.03773715 |
| 7 | 35 | 0.480 | 0.104 | 5.831997e-03 | 0.02653108 |
| 8 | 40 | 0.568 | 0.070 | 1.235409e-03 | 0.04369047 |
| 9 | 50 | 0.752 | 0.148 | 8.061062e-05 | 0.01992050 |
| 10 | 60 | 0.776 | 0.080 | 2.177233e-05 | 0.02946098 |
| 11 | 70 | 0.844 | 0.136 | 5.572812e-06 | 0.05060930 |
| 12 | 80 | 0.856 | 0.094 | 2.092072e-07 | 0.03591188 |
| 13 | 90 | 0.902 | 0.116 | 4.089564e-07 | 0.03834440 |
| 14 | 100 | 0.912 | 0.088 | 2.226856e-06 | 0.04362608 |

تصویر ۲

$static

| | number | mean_pvalue | mean_skweness | mean_out | tdiff |
|---|---|---|---|---|---|
| 1 | 5 | 0.212 | 1 | 1.770970e-01 | 0.1225000 |
| 2 | 10 | 0.246 | 1 | 2.155451e-02 | 0.5016250 |
| 3 | 15 | 0.270 | 1 | 3.229539e-03 | 0.6485308 |
| 4 | 20 | 0.324 | 1 | 2.156758e-04 | 0.6656817 |
| 5 | 25 | 0.326 | 1 | 3.186131e-05 | 0.7552479 |
| 6 | 30 | 0.372 | 1 | 4.757844e-06 | 0.7322628 |
| 7 | 35 | 0.396 | 1 | 5.480373e-07 | 0.7834689 |
| 8 | 40 | 0.368 | 1 | 5.869207e-08 | 0.7843095 |
| 9 | 50 | 0.510 | 1 | 4.328392e-10 | 0.8040795 |
| 10 | 60 | 0.540 | 1 | 1.202255e-10 | 0.8205390 |
| 11 | 70 | 0.610 | 1 | 1.875217e-12 | 0.8433907 |
| 12 | 80 | 0.632 | 1 | 2.587565e-13 | 0.8520881 |
| 13 | 90 | 0.684 | 1 | 1.251785e-14 | 0.8596556 |
| 14 | 100 | 0.612 | 1 | 9.494787e-16 | 0.8643739 |

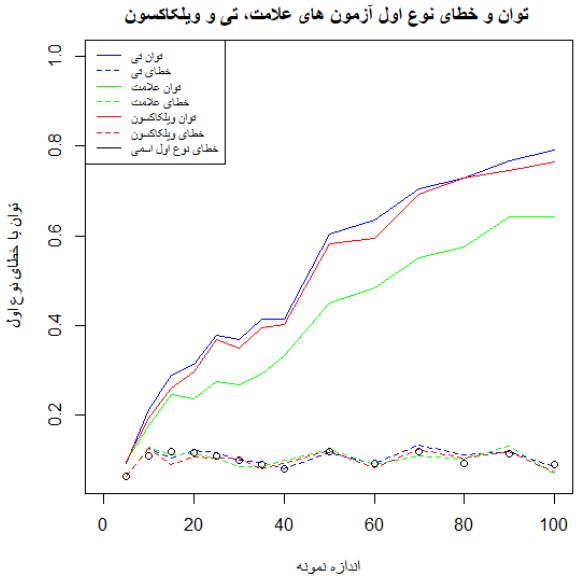
تصویر ۳

$static

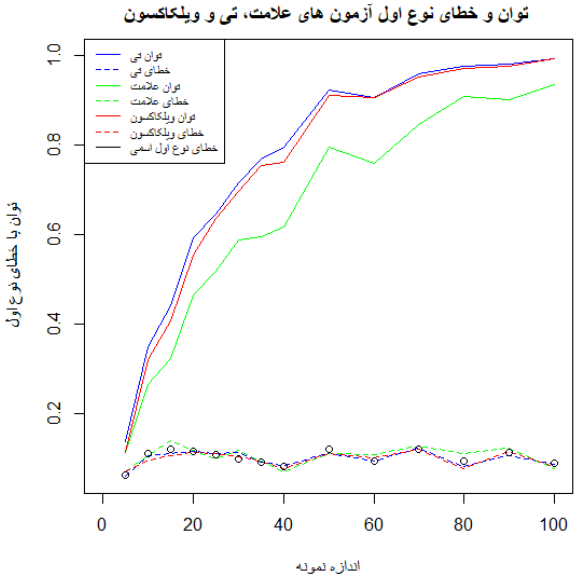
| | number | mean_pvalue | mean_skweness | mean_out | tdiff |
|---|---|---|---|---|---|
| 1 | 5 | 0.190 | 0.090 | 3.964173e-01 | 0.04650000 |
| 2 | 10 | 0.356 | 0.146 | 2.291603e-01 | 0.02762500 |
| 3 | 15 | 0.474 | 0.162 | 1.569778e-01 | 0.01253080 |
| 4 | 20 | 0.640 | 0.200 | 8.921038e-02 | 0.03768170 |
| 5 | 25 | 0.648 | 0.200 | 5.006522e-02 | 0.00324790 |
| 6 | 30 | 0.676 | 0.180 | 2.751323e-02 | 0.00026285 |
| 7 | 35 | 0.792 | 0.196 | 1.856279e-02 | 0.02346892 |
| 8 | 40 | 0.800 | 0.184 | 8.811928e-03 | 0.00369047 |
| 9 | 50 | 0.878 | 0.226 | 3.361355e-03 | 0.01192050 |
| 10 | 60 | 0.890 | 0.268 | 8.519169e-04 | 0.02546098 |
| 11 | 70 | 0.974 | 0.338 | 4.199674e-04 | 0.00860930 |
| 12 | 80 | 0.972 | 0.332 | 1.318633e-04 | 0.00208812 |
| 13 | 90 | 0.982 | 0.368 | 9.706523e-05 | 0.00565560 |
| 14 | 100 | 0.986 | 0.356 | 1.595646e-05 | 0.00837392 |

تصویر ٤

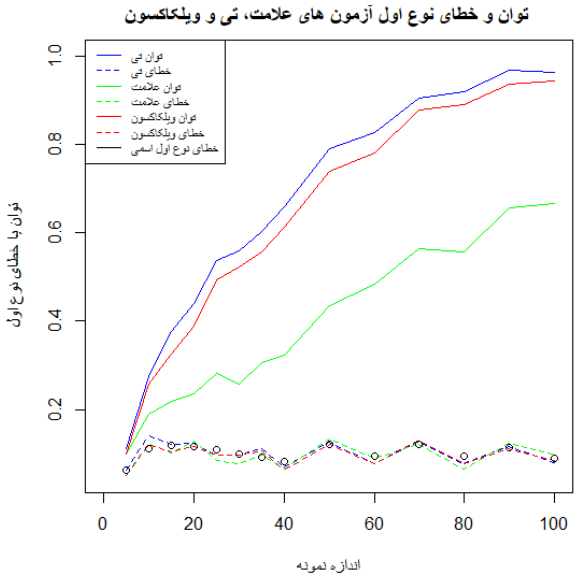
دسته اول (چولگی کم و بدون داده ی دور افتاده)



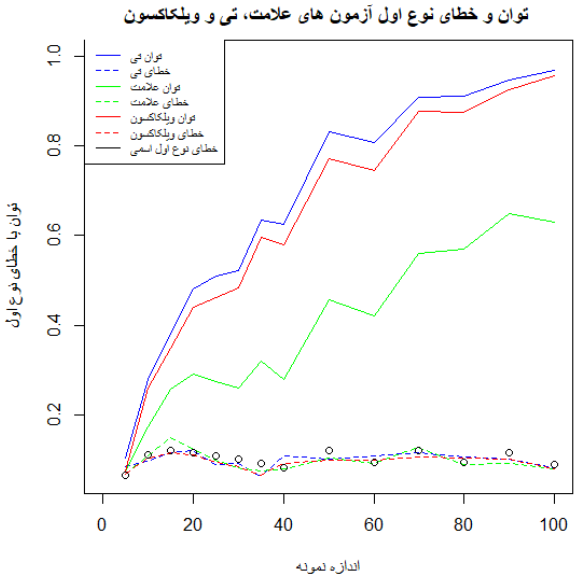

N(2,2)




N(2,5)



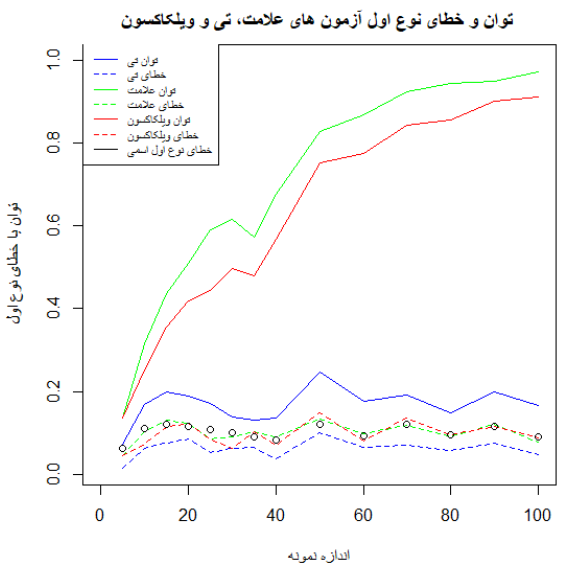

U(-3,7)



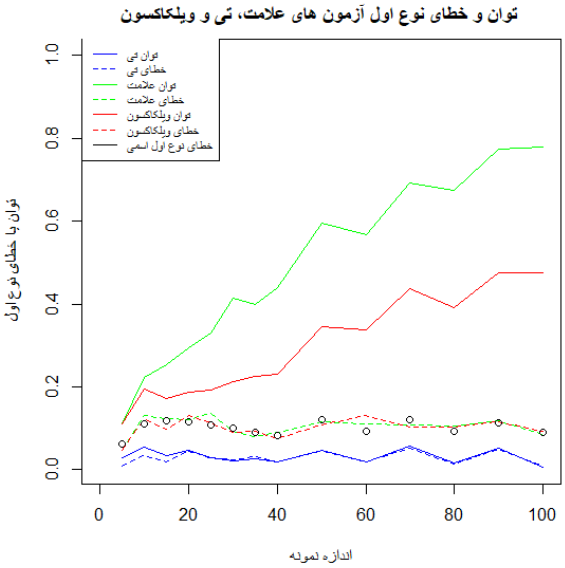

U(0,1)

تصاویر فوق که نمونه ای برای اعضای این دسته هستند نشان می دهد که خطای نوع اول هر سه آزمون نزدیک به هم است پس می توان این آزمون ها را به طور عادلانه مقایسه کرد. آزمون t بهترین عملکرد را داشته و بیشترین توان را دارد و در رده بعدی آزمون ویلکاکسون بهتر عمل کرده است. اگرچه رتبه ی دوم است ولی بسیار به توان آزمون تی نزدیک است. فرض نرمال بودن معمولا برای داده های این گروه تایید می شود و به همین خاطر در این گروه آزمون t عملکرد مطلوبی دارد. میانگین چولگی و میانگین داده های دور افتاده هردو کمتر از یک بوده و نشان می دهد چولگی کم بوده و داده ها بدون داده دورافتاده هستند. تصویر ۱ نماینده ی خروجی برای این گروه بوده است.

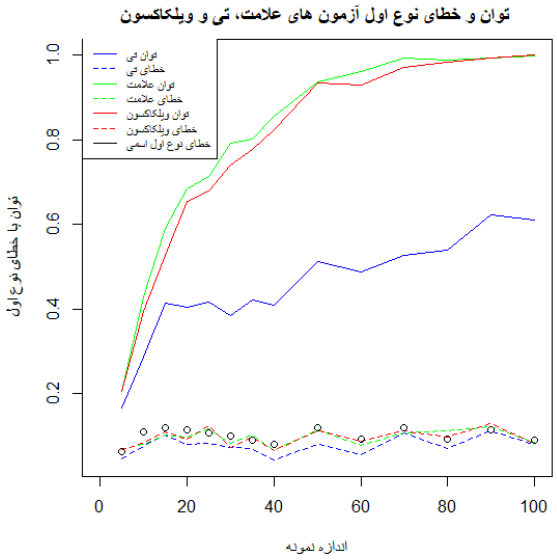
دسته دوم (چولگی کم و دارای داده ی دور افتاده)



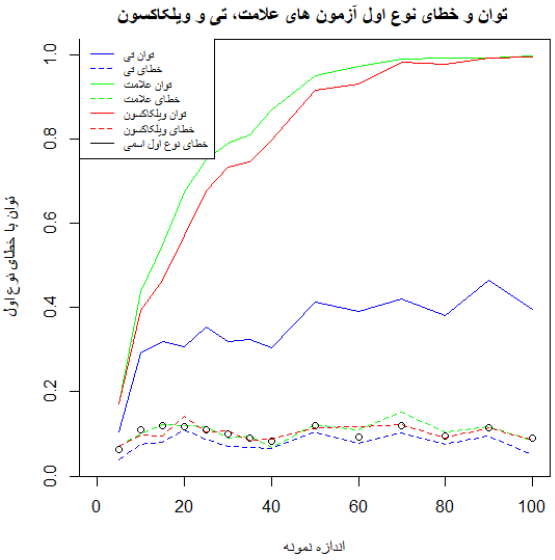

t(1)




t(0.5)



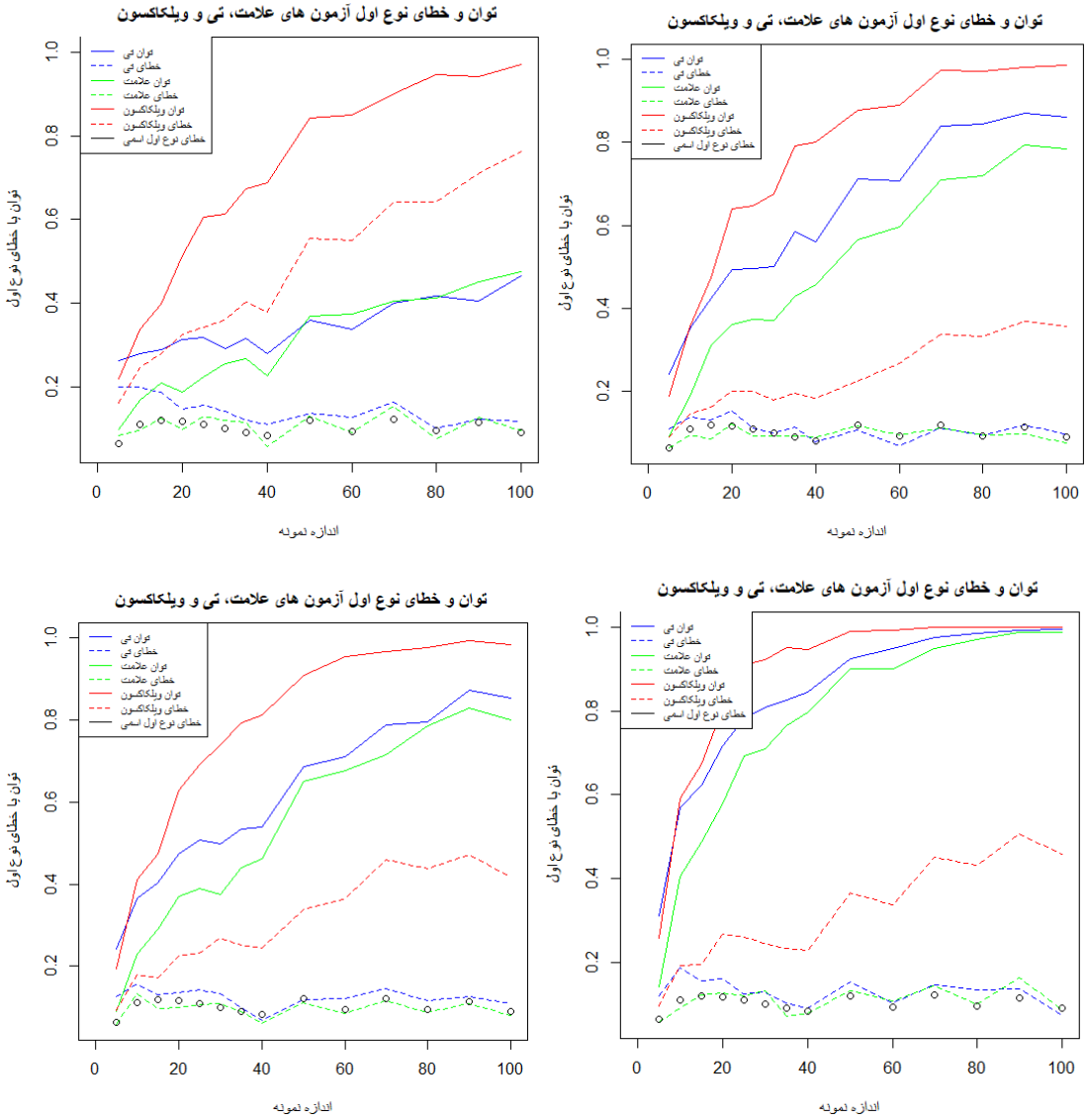

t(1.5)




t(1.25)

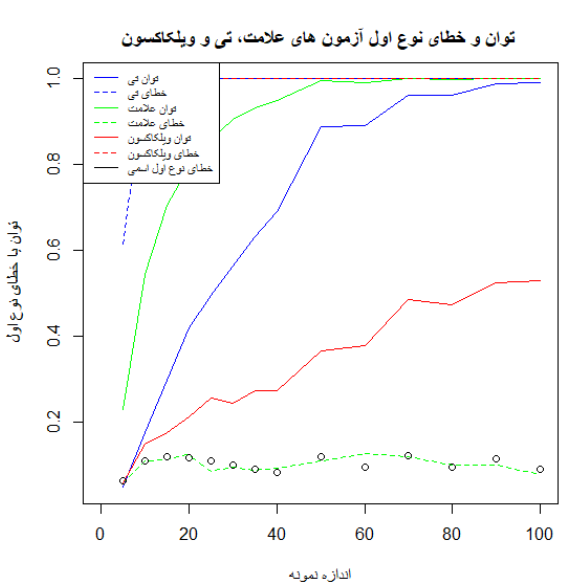
تصاویر فوق نمودار هایی مشابه برای توزیع های این گروه هستند. خطای نوع اول آزمون ها تقریبا برابر اسا پس می توانیم این سه آزمون را به طور عادلانه مقایسه کنیم. آزمون علامت  بیشترین توان را دارد و بدترین عملکرد مربوط به آزمون t است. میانگین چولگی کمتر از یک است و داده ها چولگی کمی دارند. داده های دورافتاده نسبتا زیاد بوده اند که می تواند دلیل عملکرد بد آزمون t باشند چون داده های دور افتاده از توان آزمون t می کاهند. تصویر ۲ نماینده ی خروجی برای این گروه بوده است.

دسته سوم (چولگی زیاد و بدون داده ی دور افتاده)

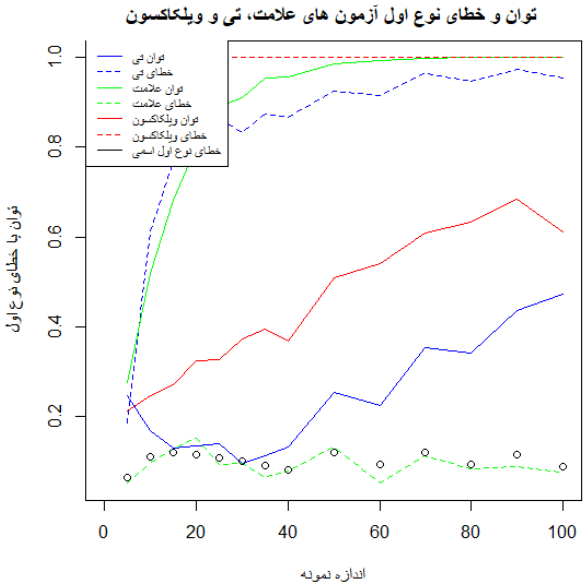



تصاویر فوق نماینده ای برای توزیع های این گروه هستند.خطای نوع اول آزمون ویلکاکسون با افزایش اندازه نمونه از خطای نوع اول دو آزمون دیگر فاصله گرفته است و به همین سبب توان این آزمون نیز افزایش یافته است پس باید از مقایسه آزمون ها بپرهیزیم. یعنی این افزایش توان کاذب است. آزمون ویلکاکسون به شدت به چولگی داده ها حساس است و وجود چولگی باعث نتایج غیر قابل اعتماد از این آزمون می شود ولی میتوان آزمون تی و علامت را مقایسه کرد که در آن آزمون تی بهتر عمل کرده است. تصویر ۳ نماینده ی خروجی برای این گروه بوده است.

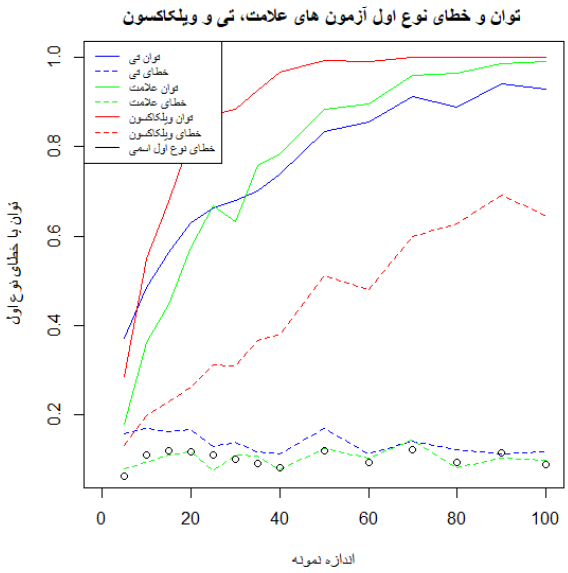
دسته چهارم (چولگی زیاد و دارای داده ی دور افتاده)



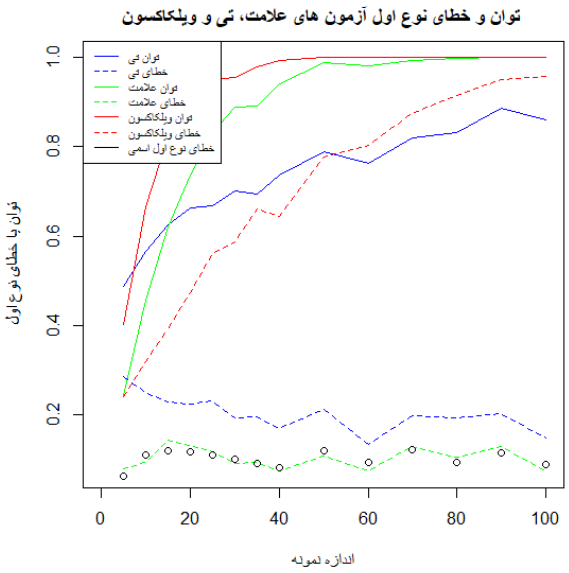

lnorm(0,1)




lnorm(0,2)




F(4,9)




F(1,6)

نمودار های بالا نمونه ای مشابه توزیع های این دسته اند. به دلیل اختلاف زیاد خطای نوع اول آزمون ها و تاثیر آن بر توان آزمون ها از مقایسه آزمون ها اجتناب می کنیم. تصویر ۴ نماینده ی خروجی برای این گروه بوده است.

## نتایج

در مقایسه آزمون‌های ناپارامتری (ویلکاکسون و علامت) با آزمون پارامتری تی ، باید به مفروضات و میزان مقاومت هر آزمون توجه داشت.

آزمون تی، به لطف قضیه حد مرکزی، در برابر نقض فرض نرمال بودن (به خصوص در حجم نمونه‌های بالا) نسبتاً مقاوم است. با این حال، اعتبار و توان این آزمون به شدت تحت تأثیر دو عامل کلیدی است :چولگی شدید و وجود داده‌های دورافتاده.

این دو پدیده نه تنها باعث کاهش توان آزمون تی می‌شوند، بلکه می‌توانند اعتبار آن را نیز زیر سوال ببرند. در چنین شرایطی، ممکن است خطای نوع اول آزمون از سطح اسمی مورد نظر (مثلاً ۰.۰۵) فراتر رود و نتایج آزمون گمراه‌کننده باشد.

در چنین شرایطی، آزمون‌های ناپارامتری گزینه‌های بهتری هستند .آزمون ویلکاکسون، که بر اساس رتبه‌ها عمل می‌کند، در برابر چولگی مقاوم‌تر از آزمون تی است، اما همچنان به میزانی از تقارن در توزیع تفاوت‌ها نیاز دارد و داده‌های پرت می‌توانند بر آن تأثیر بگذارند. مقاوم‌ترین گزینه، آزمون علامت است که به دلیل تمرکز صرف بر میانه و عدم وابستگی به شکل توزیع، در برابر هر دو مشکل چولگی شدید و داده‌های پرت، اعتبار خود را به بهترین شکل حفظ می‌کند.

| سناریو | چولگی کم و بدون داده ی دور افتاده | چولگی کم و دارای داده ی دور افتاده | چولگی زیاد و بدون داده ی دور افتاده | چولگی زیاد و دارای داده ی دور افتاده |
|---|---|---|---|---|
| رتبه اول بهترین | تی | علامت | علامت | علامت |
| رتبه دوم بهترین | ویلکاکسون | ویلکاکسون | - | - |
| رتبه آخر | علامت | تی | - | - |