تحلیل میانگین‌ها برای مشاهدات غیر نرمال

رسول نورالسناء[i]؛ جلال صفری[ii]

چکیده

تحلیل میانگین‌ها یک روش گرافیکی برای بررسی برابری میانگین چند جامعه است. در این روش، پس از تخمین میانگین هر یک از جامعه‌های مورد مطالعه و رسم آنها بر روی نموداری که حدود بالا و پایین دارد، در مورد برابری یا عدم برابری میانگین‌ها تصمیم گیری می‌شود. یکی از مفروضات اصلی در انجام تحلیل میانگین‌ها فرض نرمال بودن مشاهدات است. در صورت نقض این فرض، تحلیل‌های انجام شده می‌تواند به شدت تحت تاثیر قرار گیرد. یکی از راه‌های غلبه بر این مشکل استفاده از تبدیل متغیر مناسب و انجام تحلیل میانگین‌ها به روش متداول است. راه حل دیگری که در این مقاله توسعه داده شده است، به طور مستقیم مشاهدات را از روش تحلیل میانگین‌ها مورد ارزیابی قرار می‌دهد.

کلمات کلیدی

آنالیز واریانس، تحلیل میانگین‌ها، مشاهدات غیر نرمال، تبدیل متغیر

Analysis of Means for Non-normal Data

R. Noorossana; J. Safari

ABSTRACT

Analysis of means (ANOM) is a graphical method for evaluating equality of several treatment means. Estimates of treatment means are plotted on a chart equipped with lower and upper decision limits to determine equality of the means. Normality of observations is an implicit assumption considered in the ANOM. When this assumption fails then the results can be affected. One way to overcome this problem is transform the non-normal data to normal data. Transformation of data is a common remedy suggested to be performed prior to the ANOM. This paper provides a method for conducting ANOM when the assumption of normality does not hold.

KEYWORDS

Analysis of variance (ANOVA), Analysis of means (ANOM), Non-normal data, Transformation.

۱- مقدمه

طراحی آزمایش‌ها در چند دهه اخیر به طور قابل ملاحظه‌ای در صنایع گوناگون برای ارتقای سطح کیفیت خدمات و محصولات استفاده شده است. یکی از آزمون‌های پرکاربرد این شاخصه از علم مهندسی، آزمون برابری میانگین جوامع می‌باشد. تحلیل واریانس[1] یا به اختصار ANOVA و تحلیل میانگین‌ها[2] یا به اختصار ANOM از جمله روش‌های اجرای این آزمون است. این دو روش در اصول و نتایج شبیه هم بوده؛ ولی در ماهیت کمی با هم اختلاف دارند. در زیر عمده تفاوت‌های این دو تکنیک تشریح شده است:

الف) پس از آنالیز واریانس فقط در مورد برابری و یا عدم برابری میانگین جوامع می‌توان اظهار نظر کرد، ولی نمی‌توان تشخیص داد که میانگین کدام جامعه با دیگر جوامع اختلاف دارد. برای کشف این موضوع از روش‌هایی

[i] استاد، دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران: rassoul@iust.ac.ir

[ii] دانشجوی دکترا، گروه مهندسی صنایع، دانشکده فنی دانشگاه آزاد اسلامی، واحد علوم و تحقیقات : jalalsafari@pideco.com.

مانند توکی[3]، فیشر[4]، بنفرونی[5] و غیره می‌بایست استفاده شود [۱]. در صورتی که روش تحلیل میانگین‌ها، جامعه‌ای را که میانگین آن با میانگین کل و سایر میانگین‌ها برابر نیست، مستقیماً نشان می‌دهد.

ب) درجه حساسیت تحلیل میانگین‌ها؛ برای شناسایی تغییرات غیر تصادفی یک سطح نسبت به آنالیز واریانس بیشتر است [۹].

ج) درک تحلیل میانگین‌ها به دلیل استفاده از نمودارها نسبت به آنالیز واریانس ساده‌تر است.

تحلیل میانگین‌ها در واقع حالت توسعه یافته نمودار کنترلی شوهارت[6] است [۱۰]. تحلیل میانگین‌ها اولین بار به وسیله Ott [3] ارائه و به وسیله Schilling [8] و Nelson [12] توسعه یافت. در مطالعات تکمیلی Schilling [9] با ارائه راه حلی توانست این تکنیک را برای مشاهدات غیر نرمال با تابع توزیع مشخص نظیر بینم، نمایی، پواسان و غیره بسط دهد؛ ولی در عمل، اکثر اوقات تابع توزیع مشاهدات معلوم نمی‌باشد. بنابراین، برای تحلیل میانگین‌های مشاهدات غیر نرمال؛ که تابع توزیع آنها مشخص نیست، چه باید کرد؟ در این مقاله، دو روش جدید برای تحلیل میانگین‌های مشاهدات غیر نرمال ارائه شده است که مشکل یاد شده را برطرف می‌کند.

۲- تحلیل میانگین ها

الگوریتم تحلیل میانگین‌های تک عاملی را می‌توان به صورت زیر تشریح کرد [۲]:

۱ـ محاسبه مقادیر $\overline{X}_i$ و $\overline{\overline{X}}$ که به ترتیب معرف میانگین مشاهدات در سطح i ام و میانگین کل مشاهدات است.

۲ـ محاسبه مقادیر R_i و $\overline{R}$ که به ترتیب معرف دامنه مشاهدات در سطح i ام و میانگین دامنه هاست.

۳- تخمین خطای آزمایش، σ_e، از رابطه (۲) با V درجه آزادی.

$$V = 0.9 \times k(r-1) \quad (۱)$$

در رابطه فوق، k تعداد سطوح و r تعداد مشاهدات در هر سطح را نشان می‌دهد.

$$\sigma_e = \overline{R} / d_2^* \quad (۲)$$

مقدار d_2^* باتوجه به مقادیر مفروض V و k از جدول (۱) تعیین می‌شود.

۴- محاسبه حدود تصمیم‌گیری پایین و بالا از رابطه(۳).

$$\overline{\overline{X}} \pm \left[\frac{h_\alpha \sigma_e}{\sqrt{r}}\right]\sqrt{\frac{k-1}{k}} \quad (۳)$$

مقدار h_α به ازای تعداد سطوح مورد مقایسه، k، درجه آزادی خطا، V، و درصد خطا، α، از جدول (۲) تعیین می‌شود.

۵- رسم نموداری مانند شکل (۱) که در آن حدود تصمیم گیری و میانگین سطوح نشان داده می‌شود. در صورتی که میانگین یکی از سطوح؛ که به صورت نقطه نشان داده شده است، در داخل حدود قرار نگیرد، نتیجه‌گیری می‌شود که عامل مورد نظر در این سطح دارای اثر معنی‌دار است و یا میانگین این سطح با میانگین کل برابر نیست.

۲-۱- تحلیل میانگین‌ها برای مشاهدات غیر نرمال[۹]

الگوریتم تحلیل میانگین‌ها برای مشاهدات غیر نرمال را؛ که توزیع معلوم دارند، می‌توان به صورت زیر تشریح کرد:

۱- محاسبه مقدار α' از رابطه (۴)، با فرض معلوم بودن مقدار خطا (α).

$$\frac{\alpha}{2k} = \alpha' \quad (۴)$$

۲- محاسبه تابع توزیع تجمعی($F_x(X)$).

۳- تعیین حد تصمیم گیری بالا[7] یا به اختصار UDL بر اساس کوچک‌ترین مقداری که رابطه $F_x(UDL) > 1-\alpha'$ را برقرار می‌سازد.

۴- تعیین حد تصمیم گیری پایین[8] یا به اختصار LDL بر اساس بزرگ‌ترین مقداری که رابطه $F_x(LDL) < \alpha'$ را برقرار می‌سازد.

جدول (۱): مقادیر d_2^* - فاکتوری برای تخمین انحراف استاندارد- با k سطح یا جامعه و V درجه آزادی

تعداد جوامع مورد مقایسه (k)	V	d_2^*	V	d_2^*	V	d_2^*	V	d_2^*	V	d_2^*	V	d_2^*
۱	۱	۱/۴۱	۲	۱/۹۱	۲/۹	۲/۲۴	۳/۸	۲/۴۸	۴/۷	۲/۶۷	۵/۵	۲/۸۳
۲	۱/۹	۱/۲۸	۳/۸	۱/۸۱	۵/۷	۲/۱۵	۷/۵	۲/۴	۹/۲	۲/۶	۱۰/۸	۲/۷۷
۳	۲/۸	۱/۲۳	۵/۷	۱/۷۷	۸/۴	۲/۱۲	۱۱/۱	۲/۳۸	۱۳/۶	۲/۵۸	۱۶	۲/۷۵
۴	۳/۷	۱/۲۱	۷/۵	۱/۷۵	۱۱/۲	۲/۱۱	۱۴/۷	۲/۳۷	۱۸/۱	۲/۵۷	۲۱/۳	۲/۷۴
۵	۴/۶	۱/۱۹	۹/۳	۱/۷۴	۱۳/۹	۲/۱	۱۸/۷	۲/۳۶	۲۲/۶	۲/۵۶	۲۶/۶	۲/۷۳
۶	۵/۵	۱/۱۸	۱۱/۱	۱/۷۳	۱۶/۶	۲/۰۹	۲۲	۲/۳۵	۲۷/۱	۲/۵۶	۳۱/۸	۲/۷۳
۷	۶/۴	۱/۱۷	۱۲/۹	۱/۷۳	۱۹/۴	۲/۰۹	۲۵/۶	۲/۳۵	۳۱/۵	۲/۵۵	۳۷/۱	۲/۷۲
۸	۷/۲	۱/۱۷	۱۴/۸	۱/۷۲	۲۲/۱	۲/۰۸	۲۹/۳	۲/۳۵	۳۶	۲/۵۵	۴۲/۴	۲/۷۲
۹	۸/۱	۱/۱۶	۱۶/۶	۱/۷۲	۲۴/۸	۲/۰۸	۳۲/۹	۲/۳۴	۴۰/۵	۲/۵۵	۴۷/۷	۲/۷۲
۱۰	۹	۱/۱۶	۱۸/۴	۱/۷۲	۲۷/۶	۲/۰۸	۳۶/۵	۲/۳۴	۴۴/۹	۲/۵۵	۵۲/۹	۲/۷۲
۱۱	۹/۹	۱/۱۶	۲۰/۲	۱/۷۱	۳۰/۳	۲/۰۸	۴۰/۱	۲/۳۴	۴۹/۴	۲/۵۵	۵۸/۲	۲/۷۲
۱۲	۱۰/۸	۱/۱۵	۲۲	۱/۷۱	۳۳	۲/۰۷	۴۳/۷	۲/۳۴	۵۳/۹	۲/۵۵	۶۳/۵	۲/۷۲
۱۳	۱۱/۶	۱/۱۵	۲۳/۹	۱/۷۱	۳۵/۷	۲/۰۷	۴۷/۴	۲/۳۴	۵۸/۴	۲/۵۵	۶۸/۸	۲/۷۱
۱۴	۱۲/۵	۱/۱۵	۲۵/۷	۱/۷۱	۳۸/۵	۲/۰۷	۵۱	۲/۳۴	۶۲/۸	۲/۵۴	۷۴	۲/۷۱
۱۵	۱۳/۴	۱/۱۵	۲۷/۵	۱/۷۱	۴۱/۲	۲/۰۷	۵۴/۶	۲/۳۴	۶۷/۳	۲/۵۴	۷۹/۳	۲/۷۱
۱۶	۱۴/۳	۱/۱۵	۲۹/۳	۱/۷۱	۴۳/۹	۲/۰۷	۵۸/۲	۲/۳۴	۷۱/۸	۲/۵۴	۸۴/۶	۲/۷۱
۱۷	۱۵/۲	۱/۱۵	۳۱/۱	۱/۷۱	۴۶/۷	۲/۰۷	۶۱/۸	۲/۳۴	۷۶/۲	۲/۵۴	۸۹/۸	۲/۷۱
۱۸	۱۶	۱/۱۵	۳۳	۱/۷۱	۴۹/۴	۲/۰۷	۶۵/۵	۲/۳۳	۸۰/۷	۲/۵۴	۹۵/۱	۲/۷۱
۱۹	۱۶/۹	۱/۱۴	۳۴/۸	۱/۷	۵۲/۲	۲/۰۷	۶۹/۱	۲/۳۳	۸۵/۲	۲/۵۴	۱۰۰/۴	۲/۷۱
۲۰	۱۷/۸	۱/۱۴	۳۶/۶	۱/۷	۵۴/۹	۲/۰۷	۷۲/۷	۲/۳۳	۸۹/۷	۲/۵۴	۱۰۵/۷	۲/۷۱
۲۵	۲۲/۲	۱/۱۴	۴۵/۷	۱/۷	۶۸/۵	۲/۰۷	--	--	--	--	--	--
۳۰	۲۶/۶	۱/۱۴	۵۴/۸	۱/۷	--	--	--	--	--	--	--	--
۵۰	۴۴/۲	۱/۱۳	--	--	--	--	--	--	--	--	--	--

۲-۲- تحلیل میانگین‌ها به روش تبدیل متغیر

یکی از مفروضات اصلی روش ارائه شده در بخش [۱-۲-§]، معلوم بودن تابع توزیع مشاهدات است. در صورت نقض این فرض دیگر نمی‌توان به نتایج حاصل اطمینان کرد. یکی از راه‌های غلبه بر این مشکل استفاده از تبدیل متغیر مناسب (تبدیل مشاهدات غیر نرمال به نرمال) و انجام تحلیل میانگین‌ها به روش متداول است.

اشاره به این نکته ضروری است که علاوه بر فرض نرمال بودن مشاهدات، مفروضات دیگری نظیر ثابت بودن پراکندگی جوامع و اثرات جمعی نیز وجود دارد. در صورت نقض این مفروضات نتایج تحلیل‌ها به شدت تحت تاثیر قرار خواهند گرفت. به منظور برقراری این مفروضات یا به عبارتی برای تبدیل مشاهدات غیر نرمال به نرمال، ثابت کردن پراکندگی و تبدیل اثرات ضربی به جمعی، به ترتیب استفاده از تبدیل متغیرهای $Z = \log y$, $Z \quad y^{\frac{1}{2}}$, $Z \quad y^{\frac{2}{3}}$ توصیه شده است؛ ولی این تبدیل متغیرها برقراری تمامی مفروضات بالا را به طور یکجا تضمین نمی‌کند. به عنوان مثال اگر مشاهدات نرمال نباشند و با استفاده از تبدیل متغیر $Z = y^{\frac{2}{3}}$ به نرمال تبدیل شوند، ثابت بودن پراکندگی مشاهدات تبدیل شده، ضرورت نخواهد داشت [۷]. یکی از راه‌های رفع این نقص تهیه تبدیل متغیر، استفاده از سیستم جانسون است. به منظور کسب اطلاعات بیشتر در خصوص مشکلات استفاده از تبدیل متغیرها به مطالعات Lewis [11] و همکارانش مراجعه شود.

جدول (۲): مقادیر h_α، فاکتوری برای محاسبه حدود تصمیم گیری با k جامعه مورد مقایسه و V درجه آزادی خطا و α درصد خطا

V	α	k													
		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۵	۲۰	۳۰	۴۰	۶۰
۳	0/05	۳/۱۸	۴/۸۶												
	0/01	۵/۸۴	۸/۵۴												
۴	0/05	۲/۷۸	۳/۹۷	۴/۳۲											
	0/01	۴/۶	۶/۲۵	۶/۷۵											
۵	0/05	۲/۵۷	۳/۵۴	۳/۸۱	۴/۰۴										
	0/01	۴/۰۳	۵/۲۴	۵/۶	۵/۸۹										
۶	0/05	۲/۴۵	۳/۲۸	۳/۵۲	۳/۷۱	۳/۸۷									
	0/01	۳/۷۱	۴/۷	۴/۹۹	۵/۲۱	۵/۴									
۷	0/05	۲/۳۶	۳/۱۲	۳/۳۴	۳/۵	۳/۶۴	۳/۷۶								
	0/01	۳/۵	۴/۳۴	۴/۶	۴/۷۹	۴/۹۵	۵/۰۹								
۸	0/05	۲/۳۱	۳/۰۱	۳/۲۱	۳/۳۵	۳/۴۸	۳/۵۹	۳/۶۸							
	0/01	۳/۳۶	۴/۱۳	۴/۳۴	۴/۵۱	۴/۶۴	۴/۷۷	۴/۸۷							
۹	0/05	۲/۲۶	۲/۹۴	۳/۱۲	۳/۲۵	۳/۳۶	۳/۴۷	۳/۵۵	۳/۶۲						
	0/01	۳/۲۵	۳/۹۶	۴/۱۵	۴/۳	۴/۴۳	۴/۵۴	۴/۶۲	۴/۷۱						
۱۰	0/05	۲/۲۳	۲/۸۷	۳/۰۴	۳/۱۸	۳/۲۸	۳/۳۷	۳/۴۵	۳/۵۲	۳/۵۸					
	0/01	۳/۱۷	۳/۸۲	۴/۰۱	۴/۱۵	۴/۲۷	۴/۳۷	۴/۴۵	۴/۵۳	۴/۵۹					
۱۵	0/05	۲/۱۳	۲/۶۹	۲/۸۴	۲/۹۵	۳/۰۳	۳/۱۱	۳/۱۸	۳/۲۴	۳/۲۹	۳/۴۹				
	0/01	۲/۹۵	۳/۴۸	۳/۶۳	۳/۷۳	۳/۸۲	۳/۹	۳/۹۷	۴/۰۲	۴/۰۸	۴/۳				
۲۰	0/05	۲/۰۹	۲/۶۱	۲/۷۵	۲/۸۵	۲/۹۲	۲/۹۹	۳/۰۶	۳/۱۱	۳/۱۵	۳/۳۳	۳/۴۷			
	0/01	۲/۸۵	۳/۳۳	۳/۴۶	۳/۵۶	۳/۶۴	۳/۷	۳/۷۵	۳/۸۱	۳/۸۵	۴/۰۳	۴/۱۶			
۳۰	0/05	۲/۰۴	۲/۵۴	۲/۶۶	۲/۷۵	۲/۸۳	۲/۸۸	۲/۹۴	۲/۹۹	۳/۰۴	۳/۱۹	۳/۳	۳/۴۶		
	0/01	۲/۷۵	۳/۱۸	۳/۳	۳/۳۹	۳/۴۶	۳/۵۲	۳/۵۷	۳/۶۱	۳/۶۵	۳/۸	۳/۹	۴/۰۵		
۴۰	0/05	۲/۰۲	۲/۵	۲/۶۲	۲/۷۱	۲/۷۸	۲/۸۴	۲/۸۹	۲/۹۴	۲/۹۷	۳/۱۳	۳/۲۳	۳/۳۸	۳/۴۷	
	0/01	۲/۷	۳/۱۲	۳/۲۳	۳/۳۱	۳/۳۷	۳/۴۳	۳/۴۷	۳/۵۱	۳/۵۵	۳/۷	۳/۷۹	۳/۹۲	۴/۰۲	
۶۰	0/05	۲	۲/۴۶	۲/۵۷	۲/۶۶	۲/۷۳	۲/۷۹	۲/۸۳	۲/۸۷	۲/۹۲	۳/۰۶	۳/۱۶	۳/۳	۳/۳۹	۳/۵۲
	0/01	۲/۶۶	۳/۰۶	۳/۱۶	۳/۲۳	۳/۳	۳/۳۵	۳/۳۹	۳/۴۳	۳/۴۶	۳/۵۹	۳/۶۸	۳/۸	۳/۹	۴/۰۱
۱۲۰	0/05	۱/۹۸	۲/۴۲	۲/۵۴	۲/۶۲	۲/۶۸	۲/۷۴	۲/۷۹	۲/۸۲	۲/۸۷	۲/۹۹	۳/۰۹	۳/۲۱	۳/۳۱	۳/۴۳
	0/01	۲/۶۲	۳	۳/۰۹	۳/۱۶	۳/۲۲	۳/۲۷	۳/۳۱	۳/۳۴	۳/۳۷	۳/۵	۳/۵۸	۳/۶۸	۳/۷۸	۳/۹
بینهایت	0/05	۱/۹۶	۲/۳۹	۲/۴۹	۲/۵۷	۲/۶۴	۲/۶۹	۲/۷۴	۲/۷۷	۲/۸	۲/۹۳	۳/۰۳	۳/۱۴	۳/۲۳	۳/۳۴
	0/01	۲/۵۸	۲/۹۴	۳/۰۳	۳/۰۹	۳/۱۴	۳/۱۹	۳/۲۳	۳/۲۶	۳/۲۹	۳/۴۱	۳/۴۸	۳/۵۹	۳/۶۷	۳/۷۶

از سیستم یا توزیع جانسون به دلیل انعطاف پذیری شکل‌های آن، می توان به عنوان یک توزیع مادر یاد کرد. این سیستم را می‌توان در مورد هر توزیعی استفاده کرد. بنابراین قبل از یافتن تبدیل متغیر مناسب، می‌بایستی فرض پیروی مشاهدات از توزیع جانسون را پذیرفت. این فرضی غالب است چرا که توزیع جانسون یک توزیع مادر است[۷].

ساختار کلی الگوریتم طراحی شده برای تبدیل مشاهدات غیر نرمال به نرمال را به صورت زیر می‌توان تشریح کرد:

۱- شناسایی یکی از سه خانواده جانسون (به نام‌های S_U ، S_B و S_L) که بهتر از دو خانوادهٔ دیگر توانایی مدل کردن مشاهدات را دارد.

۲- محاسبهٔ پارامترهای مجهول خانواده مناسب با استفاده از فرمول‌های جدول (۳).

۳- تبدیل مشاهدات غیر نرمال به نرمال با استفاده از تابع حاصل شده از قدم دوم.

جدول (۳) تابع توزیع احتمال این سه خانواده از توزیع جانسون را نشان می‌دهد. در این جدول، شرایط و بازه پارامترهای هر خانواده به همراه بازه‌ای که متغیرهای X (مشاهدات غیر نرمال) می‌توانند در آن قرار گیرند، نشان داده شده است. در صورت واقع نشدن مقدار متغیر X در بازه تعریف شده نتیجه گیری می‌شود که این مشاهدات را نمی‌توان به نرمال استاندارد تبدیل کرد [۷].

روش‌های مختلفی برای تخصیص یکی از سه خانواده جانسون به مشاهدات غیر نرمال وجود دارد که در زیر به آنها

اشاره شده است:

۱- روش گشتاور[۹] [۴]

۲- روش کمترین مربعات[۱۰] [۵]

۳- روش برآورد حداقل نرم L_P [۱۱] [۵]

۴- روش صدک[۱۲] [۱۳]

در ادامه، روش صدک به خاطر سادگی، بررسی می‌شود. این روش را در قدم‌های زیر می‌توان خلاصه کرد:

۱- تخصیص مقدار عددی به متغیر z (طبق مقایساتی که انجام شده بهترین مقدار z برابر ۰/۵۲۴ است).

۲- محاسبه مقدار تابع توزیع تجمعی نرمال استاندارد برای اعداد $3z$، z ، $-z$ و $-3z$ که به ترتیب با $F(3z)$، $F(z)$، $F(-z)$ و $F(-3z)$ نشان داده می‌شود.

۳- محاسبه مقادیر x_{-3z}، x_{-z}، x_z و x_{3z} که به ترتیب برابر $F(-3z)$، $F(-z)$، $F(z)$ و $F(3z)$ امین صدک مشاهدات است.

۴- محاسبه مقادیر متغیرهای n ، m ، p و QR با استفاده از روابط زیر:

$$n = x_{-z} - x_{-3z} \tag{۵}$$

$$m = x_{3z} - x_Z \tag{۶}$$

$$p = x_z - x_{-z} \tag{۷}$$

$$QR = \frac{m \times n}{p^2} \tag{۸}$$

۵- در صورتی که QR بزرگ‌تر، کوچک‌تر و یا مساوی یک باشد آنگاه خانواده مناسب برای مشاهدات به ترتیب S_U ، S_B و S_L خواهد بود.

۲-۳- تحلیل میانگین‌ها با تخمین تابع توزیع تجمعی

در بخش [§-۲-۱]، روش اجرای تحلیل میانگین‌های مشاهدات غیر نرمال با تابع توزیع معلوم تشریح شد. در این روش از تابع توزیع مشاهدات تنها برای محاسبه تابع توزیع تجمعی آنها استفاده می‌شود. به منظور رفع محدودیت معلوم بودن تابع توزیع مشاهدات می‌توان با استفاده از روش Burr [14] تابع توزیع تجمعی مشاهدات را بدون شناخت توزیع آنها، تخمین زد. ساختار کلی الگوریتم طراحی شده برای تحلیل میانگین‌های مشاهدات غیر نرمال با تابع توزیع نامعلوم را می‌توان به صورت زیر تشریح کرد:

۱- محاسبه میانگین کل و میانگین مشاهدات در هر سطح

۲- محاسبه حداکثر میانگین مشاهدات در هر سطح

۳- محاسبه حدود بالا (UDL_1) و پایین (LDL_1) تصمیم گیری با استفاده از روش ارائه شده در بخش [§-۲-۱]،

۴- مرتب سازی صعودی میانگین مشاهدات در هر سطح و میانگین محاسبه شده در گام دوم که به صورت $y_1, y_2, \dots, y_{k+1}$ نشان داده می‌شود.

۵- محاسبه تابع توزیع تجمعی تجربی $F_e(y_i)$ و تابع تجربی[۱۳] $G_e(y_i)$ برای هر y_i با استفاده از روابط زیر:

$$F_e(y_i) = {}^{i}\!/_{n} \tag{۹}$$

$$G_e(y_i) = \ln\left[\frac{F_e(y_i)}{1-F_e(y_i)}\right] \tag{۱۰}$$

در رابطه‌های فوق، i رتبه میانگین مشاهدات را نشان می‌دهد. باید به این نکته توجه داشت که محاسبه رابطه (۱۰) برای y_{k+1} میسر نخواهد بود، چرا که هیچگاه مخرج کسر نباید صفر شود.

۶- تخمین رابطه بین y_i و $G_e(y_i)$، به صورت یک تابع با درجه فرد و جایگزینی $G_s(x)$ (s معرف درجه تابع تجربی است) به جای $G_e(y_i)$. به عنوان مثال، اگر درجهٔ انتخاب شده برای تابع تجربی برابر ۵ باشد، آنگاه $G_s(x)$ به صورت زیر نشان داده می‌شود:

$$G_s(x) = a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x^1 + a_0 \tag{۱۱}$$

جدول (۳): توابع خانواده‌های جانسون

خانواده جانسون	بازه تغییرات متغیر X	تبدیل متغیر	بازه پارامترهای تبدیل متغیر	فرمول‌های محاسبه پارامترهای تبدیل متغیر
S_B	$\varepsilon < X < \varepsilon + \lambda$	$Z = \gamma + \eta \ln\left(\frac{X-\varepsilon}{\lambda+\varepsilon-X}\right)$	$\eta > 0$	$\eta = z\left\{\cosh^{-1}\left(\frac{1}{2}\left[\left(1+\frac{p}{m}\right)\left(1+\frac{p}{n}\right)\right]^{1/2}\right)\right\}^{-1}$
			$-\infty < \gamma < \infty$	$\gamma = \eta \sinh^{-1}\left\{\left(\frac{p}{n}-\frac{p}{m}\right)\left[\left(1+\frac{p}{m}\right)\left(1+\frac{p}{n}\right)-4\right]^{1/2} \times \left[2\left(\frac{p}{m}\times\frac{p}{n}-1\right)\right]^{-1}\right\}$
			$\lambda > 0$	$\lambda = p\left\{\left[\left(1+\frac{p}{m}\right)\left(1+\frac{p}{n}\right)-2\right]^2-4\right\}^{1/2}\left(\frac{p}{m}\times\frac{p}{n}-1\right)^{-1}$
			$-\infty < \varepsilon < \infty$	$\varepsilon = \frac{x_z + x_{-z}}{2} - \frac{\lambda}{2} + p\left(\frac{p}{n}-\frac{p}{m}\right)\left[2\left(\frac{p}{m}\times\frac{p}{n}-1\right)\right]^{-1}$
S_L	$X > \varepsilon$	$Z = \gamma + \eta \ln(X-\varepsilon)$	$\eta > 0$	$\eta = \frac{2z}{\ln(m/p)}$
			$-\infty < \gamma < \infty$	$\gamma = \eta \ln\left(\frac{m/p-1}{p(m/p)^{1/2}}\right)$
			$-\infty < \varepsilon < \infty$	$\varepsilon = \frac{x_z + x_{-z}}{2} - \frac{p}{2}\left(\frac{m/p+1}{m/p-1}\right)$
S_U	$-\infty < X < \infty$	$Z = \gamma + \eta \sinh^{-1}\left(\frac{X-\varepsilon}{\lambda}\right)$	$\eta > 0$	$\eta = 2z\left\{\cosh^{-1}\left[\frac{1}{2}\left(\frac{m}{p}+\frac{n}{p}\right)\right]\right\}^{-1}$
			$-\infty < \gamma < \infty$	$\gamma = \eta \sinh^{-1}\left\{\left(\frac{n}{p}-\frac{m}{p}\right)\left[2\left(\frac{m}{p}\times\frac{n}{p}-1\right)^{1/2}\right]^{-1}\right\}$
			$\lambda > 0$	$\lambda = 2p\left(\frac{m}{p}\times\frac{n}{p}-1\right)^{1/2}\left[\left(\frac{m}{p}+\frac{n}{p}-2\right)\left(\frac{m}{p}+\frac{n}{p}+2\right)^{1/2}\right]^{-1}$
			$-\infty < \varepsilon < \infty$	$\varepsilon = \frac{x_z + x_{-z}}{2} + p\left(\frac{n}{p}-\frac{m}{p}\right)\left[2\left(\frac{m}{p}+\frac{n}{p}-2\right)\right]^{-1}$

باید به این نکته توجه داشت که هر چه درجه تابع تجربی بزرگ‌تر باشد، تخمین تابع توزیع تجمعی دقیق‌تر خواهد بود.

۷- تخمین تابع توزیع تجمعی با استفاده از رابطه زیر:

$$F_s(x) = \left[1 + \exp(-G_s(x))\right]^{-1} \qquad (۱۲)$$

۸- محاسبه مقدار α' با استفاده از رابطه (۴).

۹- محاسبه حدود بالا (UDL) و پایین (LDL) تصمیم‌گیری با استفاده از روابط زیر:

$$\begin{aligned} &Min \quad Z = |UDL_1 - UDL| \\ &S.t \begin{cases} F_s(UDL) = 1 - \alpha' \\ UDL > \overline{\overline{X}} \\ s \in \{3,5,7,..., \} \end{cases} \end{aligned} \qquad (۱۳)$$

$$\begin{aligned} &Min \quad Z = |LDL_1 - LDL| \\ &S.t \begin{cases} F_s(LDL) = \alpha' \\ LDL < \overline{\overline{X}} \\ LDL \geq 0 \\ s \in \{3,5,7,..., \} \end{cases} \end{aligned} \qquad (۱۴)$$

۳- مثال موردی

در این بخش با استفاده از یک مثال عددی عملکرد روش‌های پیشنهادی بررسی می‌شود. بدین منظور از مشاهدات ارائه شده در جدول (۴) که نتایج حاصل از بازرسی چهار بازرس را نشان می‌دهد [۶]، برای ارزیابی برابری میانگین‌های جوامع با خطای ۱۰ درصد استفاده می‌شود.

۳-۱- تحلیل میانگین‌ها به روش متداول

در این قسمت به منظور صرفه جویی در زمان از نرم افزار Minitab14 کمک گرفته می‌شود. همان گونه از شکل (۱) پیداست، تفاوت معنی دار در میانگین جوامع وجود ندارد.

P-value آزمون نرمال بودن مشاهدات بسیار پایین‌تر از درصد خطای آزمون است. این حاکی از غیر نرمال بودن مشاهدات است. در نتیجه، دیگر نمی‌توان به نتایج حاصل اطمینان کرد. بنابراین در ادامه از روش‌های پیشنهادی برای تحلیل مشاهدات استفاده می‌شود.

جدول (۴): مشاهدات بازرسان به همراه تبدیل آنها

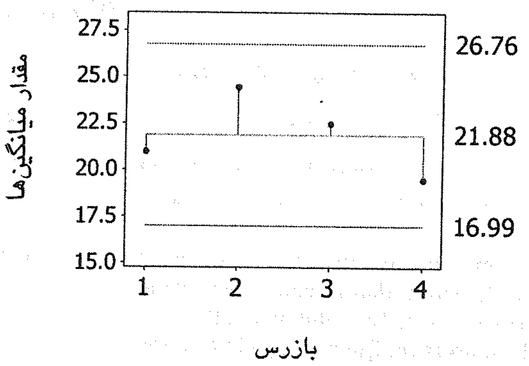
جامعه اول			جامعه دوم			جامعه سوم			جامعه چهارم		
شماره	مشاهده	عدد تبدیل شده	شماره	مشاهده	عدد تبدیل شده	شماره	مشاهده	عدد تبدیل شده	شماره	مشاهده	عدد تبدیل شده
۱	۱۹	-۰/۲۴۸۱۳	۷	۱۸	-۰/۵۲۴۰۰	۱۳	۴۰	۲/۲۹۵۲۰	۱۹	۱۸	-۰/۵۲۴۰۰
۲	۲۳	۰/۵۵۷۳۴	۸	۳۰	۱/۴۵۰۳۹	۱۴	۲۴	۰/۷۱۳۱۳	۲۰	۲۶	۰/۹۹۰۷۸
۳	۲۶	۰/۹۹۰۷۸	۹	۲۰	-۰/۰۱۰۱۳	۱۵	۲۲	۰/۳۸۷۲۲	۲۱	۱۹	-۰/۲۴۸۱۳
۴	۱۸	-۰/۵۲۴۰۰	۱۰	۲۰	-۰/۰۱۰۱۳	۱۶	۱۸	-۰/۵۲۴۰۰	۲۲	۱۷	-۰/۸۵۳۰۹
۵	۲۰	-۰/۰۱۰۱۳	۱۱	۳۲	۱/۶۴۶۶۹	۱۷	۱۶	-۱/۲۶۲۴۵	۲۳	۱۹	-۰/۲۴۸۱۳
۶	۲۰	-۰/۰۱۰۱۳	۱۲	۲۷	۱/۱۱۵۹۵	۱۸	۱۵	-۱/۸۰۶۹۹	۲۴	۱۸	-۰/۵۲۴۰۰

۳-۲- تحلیل میانگین‌ها به روش تبدیل متغیر

پس از اجرای گام‌های الگوریتم پیشنهادی در بخش [۲-۲-§]، تبدیل متغیر حاصل، به صورت زیر تعریف می‌شود:

$$Z = -0.98413 + 2.02995\sinh^{-1}\left(\frac{X - 14.30787}{9.57585}\right) \qquad (۱۵)$$

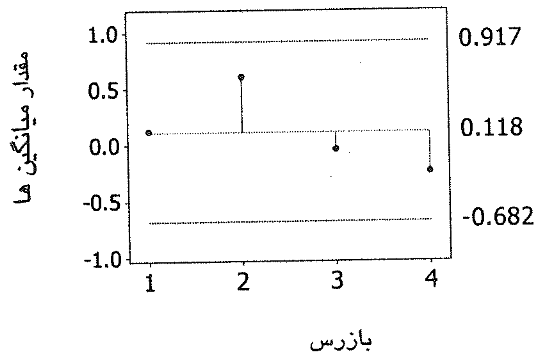
تمامی مراحل الگوریتم فوق را می‌توان با استفاده از برنامه کامپیوتری ضمیمه [۵-۱-§] اجرا کرد. خروجی این برنامه کامپیوتری؛ که همان مشاهدات نرمال می‌باشد، در جدول (۴) نشان داده شده است. پس از تحلیل میانگین‌های این مشاهدات؛ که در شکل (۲) نشان داده شده است، نتیجه‌گیری می‌شود که میانگین جوامع با هم برابر است. در پایان به منظور مقایسه روش‌های مختلف با یکدیگر می‌بایست حدود تصمیم‌گیری محاسبه شده را با استفاده از معکوس تابع جانسون به مقیاس واقعی تبدیل کرد. این حدود در جدول (۵) آورده شده است.



شکل (۱): نمودار تحلیل میانگین مشاهدات غیر نرمال

۳-۳- تحلیل میانگین‌ها با تخمین تابع توزیع تجمعی

برای تحلیل میانگین‌های مشاهدات غیر نرمال با تابع توزیع نامعلوم از الگوریتم ارائه شده در بخش [۲-۳-§] برای تعیین حدود تصمیم گیری استفاده می‌شود. مقادیر حاصل از اجرای این الگوریتم با استفاده از برنامه کامپیوتری ضمیمه [۵-۲-§] محاسبه و در جدول (۵) آورده شده است.



شکل (۲): نمودار تحلیل میانگین مشاهدات تبدیل شده

۴- نتیجه گیری

از روش‌های موجود برای تحلیل میانگین‌های مشاهدات غیر نرمال با توزیع نامعلوم نمی‌توان استفاده کرد؛ به همین منظور در این تحقیق دو روش جدید ارائه و بررسی شده است. یکی از روش‌ها، استفاده از تخمین تابع توزیع تجمعی در تحلیل میانگین‌هاست. این روش مشکلاتی دارد. یکی از مشکلات این روش وجود رابطه مستقیم بین تعداد سطوح مورد مقایسه و دقت در محاسبات حدود تصمیم گیری است. به عبارتی هر قدر تعداد سطوح مورد مقایسه بیشتر باشد، تخمین تابع توزیع تجمعی به واقعیت نزدیک‌تر شده و در نهایت حدود تصمیم گیری دقیق‌تر محاسبه می‌شود. در مورد روش دیگر؛ که استفاده از تبدیل متغیر جانسون است، بایستی به توانایی‌های این تبدیل متغیر اشاره کرد. این تبدیل متغیر هیچیک از نواقص توابع تبدیل دیگر را ندارد و می‌توان از آن به عنوان بهترین تبدیل متغیر مشاهدات غیر نرمال به نرمال یاد کرد.

۴-۱- مقایسه نتایج

حدود محاسبه شده از طریق سه روش درجدول (۵) به طور خلاصه ارائه شده است. با توجه به مقدار پایین P-value آزمون نرمال بودن مشاهدات و بر اساس نتایج جدول(۵)، اختلاف حدود محاسبه شده چندان قابل توجه نیست.

جدول (۵): مقایسه نتایج حاصل از روش‌های مختلف

حد پایین تصمیم گیری	حد بالای تصمیم گیری	نام روش حل	روش حل
۱۶/۹۹	۲۶/۷۶	تحلیل میانگین‌ها بدون دقت به غیر نرمال بودن مشاهدات	۱
۱۶/۳۹	۲۶/۸۳	تحلیل میانگین‌ها برای مشاهدات غیر نرمال با استفاده از تخمین تابع توزیع تجمعی	۲
۱۷/۴۹	۲۵/۴۵	تحلیل میانگین‌ها برای مشاهدات غیر نرمال با استفاده از تبدیل متغیر	۳

۵- ضمائم

۵-۱- ضمیمه (۱)

در این ضمیمه، برنامه کامپیوتری تهیه شده به زبان Maple 9.5 برای محاسبه تابع تبدیل مشاهدات غیر نرمال به نرمال استاندارد، از طریق تبدیل جانسون ارائه می‌شود.

```
> with(stats): with(describe): data:=[          ]:
z:=0.524:
>me:=mean(data):mad:=median(data):std:=evalf(standarddeviation(data)): con:=count(data): z1:=-3*z:z2:=-z:]z3:=z:z4:=3*z: with(statevalf);
p1:=convert(cdf[normald](z1),fraction):
p2:=convert(cdf[normald](z2),fraction):
p3:=convert(cdf[normald](z3),fraction):
p4:=convert(cdf[normald](z4),fraction):
> with(describe):
x1:=evalf(quantile[p1](data)):x2:=evalf(quantile[p2](data)):
>x3:=evalf(quantile[p3](data)):x4:=evalf(quantile[p4](data)): m:=x4-x3:p:=x3-x2:n:=x2-x1: qr:=m*n/(p^2):
> if qr>1 then au=(2*z)/(arccosh(0.5*(m/p+n/p))) elif qr<1 then ab=z/(arccosh(.5*sqrt((1+p/m)*(1+p/n)))) elif qr=1 then al=2*z/(ln(m/p)) fi: a:=rhs(%):
> if qr>1 then bu=a*arcsinh((n/p-m/p)/(2*sqrt(qr-1))) elif qr<1 then bb=a*arcsinh(((p/n-p/m)*sqrt((1+p/m)*(1+p/n)-4))/(2*(1/qr-1))) elif qr=1 then bl=a*ln((m/p-1)/(p*sqrt(m/p))) fi: b:=rhs(%):
> if qr>1 then cu=2*p*sqrt(qr-1)*1/((m/p+n/p-2)*sqrt(m/p+n/p+2)) elif qr<1 then cb=p*sqrt((((1+p/m)*(1+p/n)-2)^2-4)/(1/qr-1) elif qr=1 then cl=0 fi: c:=rhs(%):
> if qr>1 then du=(x2+x3)/2+p*(n/p-m/p)*1/(2*(m/p+n/p-2)) elif qr<1 then db=(x2+x3)/2-c/2+(p*(p/n-p/m))/(2/qr-2) elif qr=1 then dl=( x2+x3)/2-p/2*((m/p)+1)/((m/p)-1) fi: d:=rhs(%):
> if qr>1 then Z=b+a*arcsinh((v-d)/c) elif qr<1 then Z=b+a*ln((v-d)/(c+d-v)) elif qr=1 then Z=c+b+a*ln(v-d)
```

```
fi: equ:=rhs(%);f:=unapply(equ,v):
    > s1:=x->(x<d): sb:=x->is(x>(c+d)):
s1set:=select(s1,data): sbset= select(sb,data):
subset:=map(f,data):
    > if qr>1 then answer=acceptable elif qr<1 and
s1set=[] and sbset=[]  then answer=acceptable elif qr=1
and s1set=[] then answer=acceptable else
answer=nonacceptable fi;
```

۵-۲- ضمیمه (۲)

در این ضمیمه برنامه کامپیوتری تهیه شده به زبان Maple 9.5 به منظور محاسبه حدود تصمیم گیری بالا و پایین نمودار تحلیل میانگین‌ها از طریق روش تخمین تابع توزیع تجمعی ارائه می‌شود.

```
    > with(stats); with(transform);
data:=[18.852,18.948,19.776,21.388,22]; alfa:=0.1;
contlevel:=4; sor:=statsort(data);
    > with(describe); mm:=evalf(mean(sor));
con:=count(sor); with(transform); fe1:=[$1..(con-1)];
    > fe2:=apply[x->x/con](fe1); ge:=evalf(apply[x-
>ln(x/(1-x))](fe2)); yk:=sor[1..(con-1)]; with(stats);
with(fit);
    >gex:=fit[leastsquare[[x,y],y=c*x^3+f*x^2+g*x+t]]([
yk,ge]); alfaprim:=alfa/(2*contlevel); al:=-ln(1/alfaprim-
1);
    > au:=-al; ss:=con/contlevel; rrl:=rhs(gex)-al;
rru:=rhs(gex)-au; sol:=[evalf(solve(rrl,x))];
sou:=[evalf(solve(rru,x))];
    > with(describe); col:=count(sol); cou:=count(sou);
s:=1;
    > if type(sol[s],nonreal) or sol[s]>mm or sol[s]<0 then
ld[s]=0 else ld=sol[s] fi; l[s]:=rhs(%);
    > if type(sou[s],nonreal) or sou[s]<mm or sol[s]<0
then ud[s]=infinity else ud=sou[s] fi; u[s]:=rhs(%);s:=2;
    > if type(sol[s],nonreal) or sol[s]>mm or sol[s]<0 then
ld[s]=0 else ld=sol[s] fi; l[s]:=rhs(%);
    > if type(sou[s],nonreal) or sou[s]<mm or sol[s]<0
then ud[s]=infinity else ud=sou[s] fi; u[s]:=rhs(%);s:=3;
    > if type(sol[s],nonreal) or sol[s]>mm or sol[s]<0 then
ld[s]=0 else ld=sol[s] fi; l[s]:=rhs(%);
    > if type(sou[s],nonreal) or sou[s]<mm or sol[s]<0
then ud[s]=infinity else ud=sou[s] fi; u[s]:=rhs(%);
    >LDL[3]:=max(l[1],l[2],l[3]);
UDL[3]:=min(u[1],u[2],u[3]);
    > with(stats); with(fit);
gex:=fit[leastsquare[[x,y],y=a*x^5+b*x^4+c*x^3+f*x^2
+g*x+t]]([yk,ge]);
    > rrl:=rhs(gex)-al; rru:=rhs(gex)-au;
sol:=[evalf(solve(rrl,x))]; sou:=[evalf(solve(rru,x))];
with(describe);
    > col:=count(sol); cou:=count(sou); s:=1; if
type(sol[s],nonreal) or sol[s]>mm or sol[s]<0 then
ld[s]=0 else ld=sol[s] fi;
    > l[s]:=rhs(%);if type(sou[s],nonreal) or sou[s]<mm
or sol[s]<0 then ud[s]=infinity else ud=sou[s] fi;
u[s]:=rhs(%);
    > s:=2; if type(sol[s],nonreal) or sol[s]>mm or
sol[s]<0 then ld[s]=0 else ld=sol[s] fi; l[s]:=rhs(%);
    > if type(sou[s],nonreal) or sou[s]<mm or sol[s]<0
then ud[s]=infinity else ud=sou[s] fi; u[s]:=rhs(%);s:=3;
    > if type(sol[s],nonreal) or sol[s]>mm or sol[s]<0 then
ld[s]=0 else ld=sol[s] fi; l[s]:=rhs(%);
    > if type(sou[s],nonreal) or sou[s]<mm or sol[s]<0
then ud[s]=infinity else ud=sou[s] fi; u[s]:=rhs(%);s:=4;
    > if type(sol[s],nonreal) or sol[s]>mm or sol[s]<0 then
ld[s]=0 else ld=sol[s] fi; l[s]:=rhs(%);
    > if type(sou[s],nonreal) or sou[s]<mm or sol[s]<0
then ud[s]=infinity else ud=sou[s] fi; u[s]:=rhs(%);s:=5;
    > if type(sol[s],nonreal) or sol[s]>mm or sol[s]<0 then
ld[s]=0 else ld=sol[s] fi; l[s]:=rhs(%);
    > if type(sou[s],nonreal) or sou[s]<mm or sol[s]<0
then ud[s]=infinity else ud=sou[s] fi; u[s]:=rhs(%);
    >LDL[5]:=max(l[1],l[2],l[3],l[4],l[5]);
UDL[5]:=min(u[1],u[2],u[3],u[4],u[5]);
```

۶- مراجع

[۱] Douglas C. Montgomery, *Design and Analysis of Experiments*, 6nd Edition, John Wiley&Sons, 2005.

[۲] Harrison M. Wadsworth, Kenneth S. Stephens, A. Blanton Godfrey, *Modern Method for Quality Control and Improvement*, 2nd Edition, John Wiley & Sons, 2002.

[۳] Ott, E.R., Analysis of Means, Technical Report No.1, Rutgers University Statistics Center, 1958.

[٤] David J.Debrotoa, Robert S.Dittus, Stephen D.Roberts, James R.Wilson, James J.Swain, Sekhar Rekatraman, *Input Modeling with Johnson System of Distributions*, Proceeding of the Simulation Conference, pp. 165-173, 1989.

[٥] David J.Ddebrotoa, Robert S.Dittus, Stephen D.Roberts, James R.Wilson, James J.Swain, sekhar rekatraman, *Modeling input processes with Johnson Distributions*, Proceeding of the Simulation Conference, pp. 308-318, 1989.

[۶] Charles R. Hicks, Kenneth V. Turner, Fundamental Concepts in the *Design of Experiments*, 5nd Edition, Oxford University Press, 1999.

[٧] Chou, Polasky, Mason, *Transforming Non-normal Data to Normality in Statistical Process Control*, Journal of Quality Technology, vol .30, pp.133-141, 1998.

[٨] Edward G. Schilling, *A Systematic Approach to the Analysis of Means, Part 1, Analysis of Treatment Effects*, Journal of Quality Technology, vol. 5, pp. 93-108, 1973.

[٩] Edward G. Schilling, *A Systematic Approach to the Analysis of Means, Part 3, Analysis of Non-normal Data*, Journal of Quality Technology, vol.5, pp.156-159, 1973.

[١٠] Ellis R .Ott, *Analysis of Means a Graphical Procedure*, Journal of Quality Technology, vol. 15 pp.10-18 1983.

[١١] Lewis, Montgomery, Myers, *The Analysis of Designed Experiments with Non-normal Responses*, Quality Engineering, vol.12 pp. 225-243, 1999.

[١٢] Lloyd S. Nelson, *Factors for the Analysis of Means*, Journal of Quality Technology, vol. 6, , pp.175-181, 1974.

[١٣] Nicholas R. Farnum, *Using Johnson Curves to Describe Non-normal Process Data*, Quality Engineering, Vol.9, pp 329-336, 1997.

[١۴] Philippe Castagliola, *Evaluation of Non-normal Process Capability Indices Using Burrs Distribustions*, Quality Engineering, vol.8, pp 587-593, 1996.

۷- زیر نویس ها:

---

[1] Analysis of Variance (ANOVA)
[2] Analysis of means (ANOM)
[3] Tukey
[4] Fisher
[5] Benfroni
[6] Shewhart
[7] Upper decsion line (UDL)
[8] Lower decsion line (LDL)
[9] Moment matching
[10] Least squares
[11] Minimum Lp norm estimation
[12] Percentile matching
[13] Empirical function