بسم الله الرحمن الرحيم

دانشکده علوم ریاضی و کامپیوتر

پایان‌نامه کارشناسی ارشد

آمار ریاضی

عنوان

آزمون جایگشت

تدوین

امید حسینی تبار

استاد راهنما

عین‌اله پاشا

شهریور ۱۳۹۲

بسمه تعالی

مشخصات پایان‌نامه تحصیلی دانشجویان

دانشگاه خوارزمی

عنوان: آزمون جایگشت

نام نویسنده: امید حسینی‌تبار

استاد راهنما: عین‌اله پاشا

دانشکده: دانشکده علوم ریاضی و کامپیوتر	گروه: آمار ریاضی	رشته تحصیلی:
تاریخ تصویب: ۱۳۹۲/۰۶	تاریخ دفاع: ۱۳۹۲/۰۶	
مقطع تحصیلی: کارشناسی ارشد	تعداد صفحات: ؟؟	
امضای استاد راهنما:	تاریخ:	

اظهارنامه

عنوان پایان‌نامه : آزمون جایگشت

اینجانب امید حسینی‌تبار دانشجوی دوره کارشناسی ارشد دانشکده دانشکده علوم ریاضی و کامپیوتر دانشگاه خوارزمی نویسنده پایان‌نامه تحت راهنمایی عین‌اله پاشامتعهد می‌شوم:

آ. تحقیقات در این رساله توسط اینجانب انجام شده و از صحت و اصالت برخوردار است.

ب. در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.

ج. مطالب مندرج در این پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی به جایی ارائه نشده است.

د. کلیه حقوق این اثر متعلق به دانشگاه خوارزمی است و مقالات مستخرج با نام "دانشگاه خوارزمی" و یا " Kharazmy University" به چاپ خواهد رسید.

ه. حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی رساله تاثیرگذار بوده‌اند در مقالات مستخرج از آن رعایت شده است.

و. درکلیه مراحل انجام این رساله، در مواردی که از موجود زنده(یا بافت‌های آن‌ها) استفاده شده، ضوابط و اصول اخلاقی رعایت شده است.

ز. درکلیه مراحل انجام این رساله، در مواردی‌که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده، اصل رازداری، ضوابط و اصول اخلاقی انسانی رعایت شده است.

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- **کلیه حقوق این اثر و محصولات آن(مقالات مستخرج، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه خوارزمی است. این مطلب بایستی به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.**
- **استفاده از اطلاعات و نتایج این رساله بدون ذکر مرجع مجاز نیست.**

تقدیم به

الهه مهر و مهربانی

مادر عزیزم

سپاس‌گزاری...

سپاس خداوندگار حکیم را که با لطف بی‌کران خود، آدمی را زیور عقل آراست.

در آغاز وظیفه‌ی خود می‌دانم از زحمات بی‌دریغ استاد راهنمای خود، جناب آقای دکتر عین‌اله پاشا صمیمانه تشکر و قدردانی نمایم که از راهنمایی‌های ایشان در راستای پیشبرد پایان‌نامه حاصل فراوان بردم و همواره شاگرد مکتب علم و انسانیت و منش والای ایشان هستم.

همچنین لازم می‌دانم از اساتید فرهیخته، جناب آقای دکتر علی‌اکبر رحیم‌زاده ثانی و جناب آقای دکتر بردیا پناه بحق که داوری این پایان‌نامه را به عهده گرفتند با تمام وجود تشکر و قدردانی نمایم.

در پایان تشکر می‌کنم از پدر و مادر و خواهر عزیزم به پاس عاطفه سرشار و گرمای امیدبخش وجودشان، که در این سردترین روزگاران، بهترین پشتیبان من بودند.

امید حسینی تبار

شهریور ۱۳۹۲

چکیده

آزمون جایگشت نوعی آزمون ناپارامتری استکه اولین بار توسط فیشر در سال ۱۹۳۵ معرفی شد. در این آزمون، با بدترین حالت داده‌ها مواجه هستیم. یعنی توزیع داده‌ها را در اختیار نداریم و همچنین داده‌ها دارای حجم بسیار کمی هستند. P-مقدار در این آزمون به سه روش: دقیق، مونت کارلو و مجانبی محاسبه می‌شود.

واژه‌های کلیدی: آزمون جایگشت، آزمون دقیق، آزمون تصادفی، روش مونت کارلو

رده‌بندی موضوعی انجمن ریاضی ۲۰۱۰: $65C05, 11K45, 62E15, 05A05$

مقدمه

در آمار پارامتری ما فرض می‌کردیم که مشاهدات از توزیع‌هایی که صورت دقیق آن‌ها بر ما معلوم‌اند حاصل شده‌اند. در بسیاری از مسائل، مشاهدات از خانواده پارامتری از توزیع‌ها حاصل نشده‌اند. مسائلی که در آن توزیع‌های ممکن برای مشاهدات از یک خانواده پارامتری از توزیع‌ها نیستند مسائل ناپارامتری و روش‌های آماری که در چنین شرایطی به کار می‌روند، روش‌های ناپارامتری می‌نامیم.

تاریخ آمار ناپارامتری هم مانند آمار پارامتری به اوایل قرن هجدهم میلادی برمی‌گردد. اصطلاح ناپارامتری برای اولین بار در رساله دکتری یک آماردان به نام ولفویتز[1] در سال ۱۹۴۲ به‌کار رفته است. در یک استنباط ناپارامتری برای آزمون، روش‌ها مبتنی بر توابعی از مشاهدات نمونه‌ای هستند که متغیر تصادفی متناظر آن‌ها دارای توزیعی است که بستگی به تابع توزیع جامعه‌ای که نمونه از آن استخراج شده است ندارد. بنابراین، فرض‌های مربوط به جامعه مربوطه لازم نیستند. فرضی‌که اغلب لازم است، فقط این است که جامعه پیوسته باشد، در بعضی مواقع فرض‌هایی که بیشتر محدود کننده هستند مثل متقارن بودن جامعه مطرح می‌شوند، اما نه به میزانی‌که توزیع جامعه را دقیقاً بیان کنند. مسئله عملکرد قطعاً یک موضوع حائز اهمیت است. باکس و اندرسن[2] چنین بیان می‌کنند:

«برای اینکه احتیاجات آزمایشگر برآورده شوند، ملاک‌های آماری باید (۱) نسبت به تغییراتی که در عوامل خاص مورد آزمون پیش می‌آیند حساس باشند، (۲) نسبت به تغییرات اندازه‌ای مربوط به عوامل خارجی که در عمل محتملاً رخ می‌دهند حساس نباشند».

این خواص که به ترتیب **توان** و **استواری** نامیده می‌شوند به طور کلی به عنوان شرایط اولیه عملکرد خوب در آزمون فرض، مورد قبول هستند. آزمون‌های پارامتری به طریقی به دست می‌آیند که اولین نیاز بالا، برای یک تابع احتمال خاص مفروض، مثلاً با استفاده از روش نسبت درست‌نمایی برآورده شود. اما وقتی فرض‌ها برقرار نباشند، این چنین آزمون‌هایی حتی معتبر نیستند. بنابراین در آمار پارامتری استواری بسیار مورد توجه است. از طرف دیگر آزمون‌های ناپارامتری ذاتاً استوارند، زیرا فقط نیاز به فرض‌های خیلی کلی دارند. انتظار می‌رود که به واسطه این امر کاهشی در توان آزمون حاصل شود. بنابراین طبیعی به نظر می‌رسد که به استواری به عنوان یک ملاک عملکرد برای آزمون‌های پارامتری، و به توان به عنوان یک ملاک عملکرد برای آزمون‌های ناپارامتری نگاه کنیم. در مواردی‌که فرض‌های لازم برقرار باشند، بسیاری از آزمون‌های پارامتری به عنوان تواناترین آزمون شناخته می‌شوند. اما در مواردی‌که مطالعات مقایسه‌ای انجام می‌شوند، آزمون‌های ناپارامتری تقریباً همیشه به اندازه روش‌های پارامتری، بخصوص برای نمونه‌های کوچک، توانا هستند. و بنابراین هر موقع که شکی در مورد فرض‌ها وجود دارد، بیشتر مطلوب‌اند.

[1] Wolfowitz

[2] Box and Anderson

شاید برتری عمده آزمون‌های ناپارامتری در کلیت زیاد آن‌ها باشد. روش‌های ناپارامتری را عموماً می‌توان سریع و آسان به کار برد، زیرا شامل محاسبات ساده‌ای هستند. نظریه استنباط ناپارامتری به خواص آماره مورد استفاده در شیوه استقرایی مربوط می‌شود. بحث درباره این خواص مستلزم به دست آوردن توزیع نمونه‌ای آماره مناسب است، اما این کمتر از آمار پارامتری به ریاضیات پیشرفته نیاز دارد. در اکثر حالات آماره آزمون به صورت یک متغیر تصادفی گسسته است، که فقط تعداد متناهی از مقادیر را با احتمال غیر صفر می‌پذیرد، و توزیع نمونه‌ای آن‌را اغلب می‌توان به صورت شمارش و یا فرمول‌های ساده ترکیباتی به‌دست آورد. توزیع‌های مجانبی اغلب نرمال، مربع خی، و یا توابع مشهور دیگری هستند. در مورد داده‌هایی که به اصطلاح داده‌های «ناجور» نامیده می‌شوند، معمولاً روش‌های ناپارامتری خیلی مناسب‌تر از روش‌های پارامتری‌اند. در موارد بسیاری، به این‌که داده‌های اصلی موجود، به شکل مقدار واقعی باشند نیازی نیست. مثلاً اگر آزمون، مبتنی بر رتبه‌ها باشد، فقط رتبه‌ها مورد نیازند.

در بعضی مواقع در یک آزمون ناپارامتری ما با یکی از بدترین حالات برای یک آزمون مواجه می‌شویم. یعنی هم توزیع داده‌ها را در اختیار نداریم و هم داده‌ها دارای حجم بسیار کم می‌باشند. در این موارد به دلیل پایین بودن حجم نمونه استفاده از توزیع‌های مجانبی منجر به نتایج نادرست می‌شود. سوال این است که این آزمون‌ها را چگونه و با چه روش‌هایی انجام دهیم؟ این سوالی است که در این پایان‌نامه به دنبال جواب آن می‌باشیم.

برای اجرای این نوع آزمون‌ها فیشر روشی را ابداع کرد که به آزمون جایگشت یا آزمون تصادفی کردن فیشر معروف است. این آزمون را فیشر[1] پایه گزاری کرد و بعدها توسط آماردانان دیگری از جمله پیتمن توسعه داده شد. زمینه پیدایش این آزمون داستان جالبی دارد:

در سال ۱۹۲۰ رونالد فیشر خانمی را ملاقات کرد که ادعا می‌کرد می‌تواند با چشیدن تشخیص دهد که در محلول چای و شیر کدام یک ابتدا ریخته شده است. یعنی ابتدا شیر ریخته شده و سپس چای به آن اضافه شده است و یا ابتدا چای ریخته شده و سپس شیر به آن اضافه شده است. فیشر آزمایشی را برای تست صحت ادعای ایشان انجام داد. او ۸ فنجان را انتخاب کرد، در چهار فنجان ابتدا شیر و سپس چای و در چهار فنجان بعدی ابتدا چای و سپس شیر ریخت. فیشر از ایشان خواست همه آن‌ها را چشیده و نوع آن‌را بگوید. فیشر کل جایگشت‌های ۸ فنجان را آزمایش کرد. او $\binom{۸}{۴}$ حالت ممکن را بررسی کرد و در هر آزمایش تعداد حالات ممکن و تعداد جواب‌های درست را در دو ردیف به صورت جدول A یادداشت کرد. او فنجانی را که اول در آن شیر ریخته شده بود با صفر و فنجانی را که ابتدا چای ریخته شده بود با یک نشان داد. جدول A قسمتی از نتایج را نشان می دهد.

در این آزمایش هدف فیشر آزمون زیر بود:

$$\begin{cases} H_۰ : & \text{قدرت تشخیص و محلول مستقل‌اند} \\ H_۱ : & \text{قدرت تشخیص و محلول در ارتباط‌اند} \end{cases}$$

فارغ از نتیجه، این آزمایش، فیشر را بر آن داشت تا آزمونی را پایه گذاری کند و آن‌را آزمون جایگشت نامید. این آزمون را اولین بار فیشر در کنفرانسی در سال (۱۹۳۴) ارائه داد. به همین خاطر آن را آزمون تصادفی کردن فیشر می‌نامند. ممکن است این فکر ایجاد شودکه آزمون‌های جایگشت، پس از آزمون‌های رتبه‌ای و آزمون‌های مبتنی بر

[1] Fisher

جدول A: جایگشت‌ها و تعداد جواب‌های درست هر جایگشت

N	جایگشت	تعداد جواب‌های درست
۱	۱ ۱ ۱ ۱ ۰ ۰ ۰ ۰	۶
۲	۱ ۰ ۰ ۰ ۱ ۱ ۱ ۰	۴
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
۷۰	۰ ۰ ۰ ۰ ۱ ۱ ۱ ۱	۴

سایر امتیازها، سومین روش تولید آزمون‌های ناپارامتری است، ولی آزمون‌های جایگشت در واقع از نظر زمانی از این دو آزمون جلوتر است. پس شاید این سوال پیش آید که چرا آزمون جایگشت زیاد مورد توجه قرار نگرفته است؟ این آزمون برای حجم نمونه زیاد بسیار خسته کننده است، و به علت کمبود امکانات در دهه‌های پیشین، اجرای آزمون وقتی حجم نمونه زیاد بوده تقریباً غیر ممکن بوده است. به همین دلیل این آزمون در اوایل ظهور، مخالفان و منتقدان بسیاری داشت. آن‌ها اعتقاد داشتند برای حجم نمونه زیاد، اجرای آزمون غیر ممکن است. اما با پیدایش محاسبات مدرن آماری و نرم‌افزارهای کامپیوتری این اعتقاد کم کم رنگ باخت. در حقیقت پیشرفت علم آمار به اعتبار و اهمیت آزمون جایگشت افزود، تا جائی‌که هم اکنون در بسیاری از علوم مختلف از آن استفاده می‌شود.
آزمون جایگشت برای زوج‌های جور شده را فیشر (۱۹۳۵) مورد بحث قرار داده است. بعد از فیشر، آماردانان زیادی به این آزمون پرداختند. به عنوان مثال، آزمون جایگشت برای دو نمونه مستقل را پیتمن[1] (۱۹۳۷)، به موازات آزمون جایگشت برای همبستگی ارائه داده است. آزمون‌های جایگشت برای داده‌های چندمتغیره را چانگ[2] (۱۹۵۸) ارائه داد. سن[3] (۱۹۶۷) مقاله‌ای درباره آزمون‌های جایگشت برای چند نمونه ارائه داد. کلرو[4] (۱۹۶۹) بحث تقریب را در آزمون جایگشت ارائه داد. کتاب‌های جدیدی نیز به این بحث پرداخته‌اند، اگرستی[5] (۱۹۸۴،۱۹۹۰،۱۹۹۶) تحلیل داده‌های رسته ای را با استفاده از آزمون جایگشت بررسی کرد. گود[6] (۱۹۹۴) کتابی در مورد آزمون‌های جایگشت نوشت. سیدک و سن[7] (۱۹۹۹) آزمون‌های جایگشت را با استفاده از رتبه‌ها بررسی کردند.
در این پایان‌نامه آزمون‌های متعددی را با استفاده از جایگشت‌ها اجرا می‌کنیم. فصل اول کلیاتی از آزمون‌های ناپارامتری و چند موضوع مرتبط با مفاهیم پایان‌نامه را شامل می‌شود. فصل دوم را به آزمون‌های جایگشت درباره پارامترهای مرکزی برای یک، دو و چند نمونه اختصاص داده‌ایم. در فصل سوم به آزمون‌هایی درباره رابطه دو متغیر تصادفی پرداخته‌ایم و در نهایت در فصل چهارم آزمون‌های استقلال دو متغیر تصادفی را با استفاده از جایگشت‌ها

[1] Pitman
[2] Chung
[3] Sen
[4] Cleroux
[5] Agresti
[6] Good
[7] Sidak and Sen

بررسی کرده‌ایم. جواب اکثر مثال‌ها با استفاده از نرم‌افزارهای SPSS و SAS به‌دست آمده‌اند. چون خروجی این دو نرم افزار برای آزمون‌های یکسان، همانند می‌باشند بنابراین در پایان مثال‌ها به نوع نرم افزار اشاره نشده است. تا جائی‌که امکان استفاده از SPSS فراهم بوده است از نرم افزار SAS استفاده نشده است. فقط در بعضی مثال‌ها که SPSS قابلیت حل آن‌ها را نداشته است از SAS استفاده کرده‌ایم.

مقاله اصلی:

Philip T. Reiss, M. Henry H. Stevens, Zarrar Shehzad Eva Petkova, and Michael P. Milham, June 2010, On Distance-Based Permutation Tests for Between-Group Comparisons, Department of Child and Adolescent Psychiatry, New York University, New York 10016, U.S.A.

مقاله فرعی:

Kenneth A. Rose, Eric P. Smith, 2008, Statistical assessment of model goodness-of-fit using permutation tests 1,2, Department of Statistics, Virginia Polytechnic Insititute and State University, Blacksburg, VA24061, U.S.A.

فهرست مطالب

چکیده آ

مقدمه ب

۱ کلیات ۱

۱.۱ آزمون علامت برای یک نمونه ۱

۲.۱ آزمون علامت با نمونه‌ای از دوتایی‌ها ۲

۳.۱ آزمون رتبه‌ای-علامت‌دار ویلکاکسون ۳

۱.۳.۱ استفاده از آماره‌های رتبه علامت‌دار ویلکاکسون برای آزمون تقارن ۴

۴.۱ آزمون جمعی-رتبه‌ای ویلکاکسون ۴

۵.۱ آزمون من-ویتنی ۵

۱.۵.۱ داده‌های برابر ۵

۶.۱ آزمون میانه ۶

۷.۱ آزمون کروسکال-والیس ۸

۸.۱ آزمون تصادفی شده ۹

۹.۱ تحلیل واریانس ۹

۱۰.۱ ضریب همبستگی پیرسون و رتبه‌ای اسپرمن ۱۱

۱۱.۱ استواری ۱۱

۱۲.۱ سازگاری ۱۲

۱۳.۱ کارایی پیتمن ۱۲

۱۴.۱ P-مقدار ۱۳

۲ آزمون‌های جایگشت برای پارامترهای مرکزی ۱۴

۱.۲ مقدمه ۱۴

۲.۲ آزمون‌های جایگشت برای یک نمونه ۱۵

۱.۲.۲ آزمون جایگشت برای یک نمونه با استفاده از داده‌های خام ۱۵

۲.۲.۲ آزمون جایگشت برای یک نمونه با استفاده از آماره آزمون رتبه-علامت‌دار ویلکاکسون ۱۹

۳.۲.۲ آزمون جایگشت با استفاده از آماره آزمون علامت ۲۴
۴.۲.۲ برآورد P–مقدار در حالت یک نمونه با استفاده روش مونت‌کارلو ۲۶
۳.۲ آزمون‌های جایگشت برای دو نمونه مستقل ۲۷
۱.۳.۲ آزمون جایگشت برای دو نمونه مستقل با استفاده از داده‌های خام ۲۷
۲.۳.۲ آزمون جایگشت برای دو نمونه مستقل با استفاده از آزمون میانه ۳۲
۳.۳.۲ آزمون جایگشت برای دو نمونه مستقل با استفاده از آماره جمعی–رتبه‌ای ویلکاکسون ۳۵
۴.۳.۲ آزمون برابری میانگین‌های دو جامعه مستقل با استفاده از روش مونت‌کارلو ۳۸
۵.۳.۲ آزمون جایگشت برای نمونه‌های بزرگ (محاسبه تقریبی P–مقدار) ۴۱
۴.۲ آزمون‌های جایگشت برای دو نمونه وابسته ۴۳
۱.۴.۲ آزمون جایگشت برای دو نمونه وابسته با استفاده از داده‌های خام ۴۴
۲.۴.۲ آزمون جایگشت برای دو نمونه وابسته با استفاده از آماره جمعی– رتبه‌ای ویلکاکسون ۴۸
۳.۴.۲ آزمون جایگشت برای دو نمونه جفت شده با استفاده از روش مونت‌کارلو ۵۱
۵.۲ آزمون‌های جایگشت برای k نمونه مستقل ۵۱
۱.۵.۲ آزمون جایگشت برای k نمونه مستقل با استفاده از داده‌های خام ۵۲
۲.۵.۲ آزمون جایگشت برای k نمونه مستقل با استفاده از آزمون میانه ۵۴
۳.۵.۲ آزمون جایگشت برای k نمونه مستقل با استفاده از آماره آزمون کروسکال ـ والیس . ۵۷
۴.۵.۲ آزمون جایگشت مبتنی بر آماره آزمون کروسکال–والیس وقتی گره وجود دارد ۶۰
۵.۵.۲ آزمون جایگشت برای k نمونه مستقل با استفاده از آماره آزمون جانخیر–ترپسترا . . ۶۱
۶.۲ آزمون جایگشت برای k نمونه وابسته ۶۳
۱.۶.۲ آزمون جایگشت برای k نمونه وابسته با استفاده از آماره فریدمن ۶۳
۲.۶.۲ اجرای آزمون با استفاده از روش مونت‌کارلو ۶۶
۳.۶.۲ آزمون جایگشت برای پاسخ‌های دوتایی ۶۶
۷.۲ خلاصه فصل ۲ . . . . ۶۸

۳ آزمون‌های جایگشت برای همبستگی بین دو متغیر ۷۲
۱.۳ مقدمه ۷۲
۲.۳ آزمون جایگشت با استفاده از آزمون میانه ۷۲
۳.۳ آزمون جایگشت مبتنی بر آماره پیرسون ۷۵
۴.۳ آزمون جایگشت مبتنی بر آماره اسپیرمن ۷۷
۵.۳ خلاصه فصل ۳ . . . . ۸۱

۴ آزمون‌های جایگشت برای داده‌های رسته‌ای ۸۲
۱.۴ مقدمه ۸۲
۲.۴ رسته‌های با خصیصه اسمی ۸۳
۱.۲.۴ آزمون دقیق فیشر ۸۳

۲.۲.۴ آزمون جایگشت با استفاده از آماره خی دوی پیرسون ۸۴
۳.۲.۴ آزمون جایگشت با استفاده از آماره نسبت درستنمایی ۸۶
۴.۲.۴ آزمون جایگشت با استفاده از روش مونت کارلو ۸۷
۳.۴ رسته‌های سطری اسمی و رسته‌های ستونی مرتب شده(ترتیبی) ۸۸
۴.۴ رسته‌های سطری و ستونی مرتب شده (ترتیبی) ۹۰
۵.۴ خلاصه فصل ۴ ۹۱

مراجع ۹۳

واژه‌نامه فارسی به انگلیسی ۹۴

واژه‌نامه انگلیسی به فارسی ۹۷

نمایه ۱۰۰

فصل ۱

کلیات

۱.۱ آزمون علامت برای یک نمونه

نمونه تصادفی $X_۱, X_۲, \ldots, X_N$ از جامعه F_X با میانه نامعلوم M استخراج می‌شود. که در آن F_X، در نزدیکی M پیوسته فرض شده است. به عبارت دیگر فرض بر این است که مشاهدات مستقل‌اند و $P(X = M) = ۰$. فرضی که قرار است آزمون شود مربوط به مقدار میانه جامعه می‌باشد یعنی: $H_۰ : M = M_۰$. در این آزمون، ما تنها فرض پیوسته بودن توزیع در نزدیکی M را در نظر می‌گیریم و هیچ فرض دیگری در مورد توزیع نداریم. چون فرض وجود میانگین را نداریم آزمون فرض درباره میانگین نخواهیم داشت اما هر توزیع پیوسته‌ای دارای میانه است. فرض مقابل، آزمون یک‌طرفه یا دوطرفه درباره مقدار M است. مثلاً $H_۱ : M \neq M_۰$ برای هر توزیع که رابطه $P(X = M) = ۰$ را برقرار می کند. بنابر تعریف میانه داریم:

$$P(X > M) = P(X < M) = ۰/۵ \; .$$

چون در این‌جا فرض ما بیان می‌کند که $M_۰$، آن مقدار X است که مساحت زیر توزیع فراوانی را به دو قسمت مساوی تقسیم می‌کند، نمایش نمادی هم‌ارز $H_۰$ به صورت زیر است:

$$H_۰ : \theta = P(X > M_۰) = P(X < M_۰) \; .$$

اگر داده‌های نمونه با آن مقدار مفروض میانه سازگار باشند، به‌طور متوسط نیمی از مشاهدات نمونه بعد از $M_۰$، و نیمی دیگر قبل از $M_۰$ قرار خواهند گرفت. بدین ترتیب تعداد مشاهدات بعد از $M_۰$ را که با K نشان می‌دهیم، می‌توان برای آزمون کردن فرض $H_۰$ به‌کار بریم. وقتی‌که مشاهدات بدین طریق دوحالتی شده باشند، صرف‌نظر از توزیع F_X، مجموعه‌ای از N متغیر تصادفی مستقل از جامعه برنولی باپارامتری به‌صورت $\theta = P(X > M_۰)$ را تشکیل می‌دهند. در این‌صورت توزیع نمونه‌ای متغیر تصادفی K، توزیع احتمال دوجمله‌ای با پارامتر θ است، که اگر

فرض صفر درست باشد $\theta = ۰/۵$ است. چون K در واقع تعداد علامت‌های مثبت در بین N تفاضل $X_i - M_۰$، $i = ۱, ۲, \ldots, N$ است، آزمون ناپارامتری مبتنی بر K، به آزمون معمولی علامت موسوم است.
ناحیه رد مناسب به فرض مقابل بستگی دارد. فرض مقابل یک‌طرفه مورد نظر ممکن، این است که مقدار درست میانه از مقدار فرض شده آن زیادتر باشد، که هم‌ارز با این حکم است که احتمال یک علامت مثبت بیش از یک علامت منفی است. به طور نمادی می‌نویسیم: $H_۱ : M > M_۰$ یا $H_۱ : \theta = P(X > M_۰) > P(X < M_۰)$.
اگر تعداد تفاضل‌های مثبت زیادتر باشند، نمونه منعکس کننده حالت فوق خواهد بود. بنابراین ناحیه رد برای آزمونی در سطح α به صورت، $K \geq K_\alpha$ برای $K \epsilon R$ است. که در آن K_α کوچکترین عدد صحیحی انتخاب می‌شود که در رابطه $\sum_{K=K_\alpha}^{N} \binom{N}{K} ۰/۵^N \leq \alpha$ صدق کند. همچنین برای آزمون یک‌طرفه با فرض مقابل:

$$H_۱ : M < M_۰ \text{ یا } H_۱ : \theta = P(X > M_۰) < P(X < M_۰) .$$

ناحیه رد برای آزمونی با سطح α برابر است با: $K \leq K'_\alpha$ برای $K \epsilon R$. که در آن K'_α بزرگترین عدد صحیحی است که در رابطه $\sum_{K=۰}^{K'_\alpha} \binom{N}{K} ۰/۵^N \leq \alpha$ صدق می‌کند. اگر فرض مقابل دوطرفه باشد:

$$H_۱ : M \neq M_۰ \text{ یا } H_۱ : \theta = P(X > M_۰) \neq P(X < M_۰) .$$

ناحیه رد باید متشکل از مقادیر خیلی بزرگ یا خیلی کوچک K باشد. بدین ترتیب فرض صفر را وقتی رد می‌کنیم که $K \geq K_{\alpha/۲}$ یا $K \leq K'_{\alpha/۲}$، که در آن $K_{\alpha/۲}$ و $K'_{\alpha/۲}$ به‌ترتیب، کوچک‌ترین و بزرگ‌ترین اعداد صحیحی هستند که در روابط $\sum_{K=K_{\frac{\alpha}{۲}}}^{N} \binom{N}{K} ۰/۵^N \leq \frac{\alpha}{۲}$ و $\sum_{K=۰}^{K'_{\frac{\alpha}{۲}}} \binom{N}{K} ۰/۵^N \leq \frac{\alpha}{۲}$ صدق می‌کنند [۳، ۲.۶].

۲.۱ آزمون علامت با نمونه‌ای از دوتایی‌ها

شیوه آزمون علامت تک نمونه‌ای برای آزمون فرض درباره M، در مورد داده‌های نمونه زوجی نیز قابل استفاده است. از یک نمونه تصادفی N زوجی:

$$(X_۱, Y_۱), (X_۲, Y_۲), \ldots, (X_N, Y_N)$$

N تفاضل $D_i = Y_i - X_i$ به‌دست می‌آید. اگر جامعه تفاضل‌ها در نزدیکی M پیوسته فرض شود به‌طوری‌که، $P(D = M) = ۰$ و اگر θ را به‌صورت $\theta = P(D > M_۰)$ تعریف کنیم، با قرار دادن D_i ها به جای X_i ها، همان شیوه قبلی در اینجا نیز معتبر است. فرض صفر را به‌صورت $H_۰ : M_D = M_۰$ تعریف می‌کنیم. D_i های بزرگ‌تر از $M_۰$ را با B نشان می‌دهیم، اگر B خیلی بزرگ باشد فرض صفر را رد می‌کنیم. آماره B دارای توزیع $B(N, \frac{۱}{۲})$ است. توجه می‌کنیم که این آزمون، آزمونی است برای میانه تفاضل‌ها، که لزوماً همان تفاضل دو میانه M_Y و M_X نیست.

۳.۱ آزمون رتبه‌ای-علامت‌دار ویلکاکسون

نمونه تصادفی $X_۱, X_۲, \ldots, X_N$ را از توزیع پیوسته و متقارن F_X با میانه مجهول M را در نظر می‌گیریم. تحت فرض $H_۰ : M = M_۰$، تحت فرض صفر تفاضل‌های $X_i - M_۰ = D_i$ به‌طور متقارن حول صفر توزیع شده‌اند. در آزمون علامت تنها چیزی‌که مورد استفاده قرار می‌گرفت نشانه D_i بود. در آزمون رتبه‌ای-علامت‌دار ویلکاکسون[1]، نه تنها نشانه D_i بلکه رتبه قدرمطلق آن‌ها یعنی رتبه $|D_i|$ مورد استفاده قرار می‌گیرد. در این آزمون فرض می‌کنیم D دارای توزیع پیوسته و متقارن نسبت به c باشد. می‌خواهیم مثلاً آزمون یک‌طرفه زیر را انجام دهیم:

$$\begin{cases} H_۰ : \quad c = ۰ \, , \\ H_۱ : \quad c > ۰ \, . \end{cases}$$

برای این‌کار نخست مشاهدات $D_۱, D_۲, \ldots, D_n$ را در نظر می‌گیریم. رتبه‌های $|D_۱|, |D_۲|, \ldots, |D_n|$ را به ترتیب با $R_۱, R_۲, \ldots, R_n$ نشان می‌دهیم. واضح است که این رتبه‌ها درست یک جایگشت برای $۱, ۲, \ldots, n$ می‌باشند. حال تابع نشانگر:

$$I_{(x)} = \begin{cases} -۱ & x \leq ۰ \, , \\ ۱ & x > ۰ \, . \end{cases}$$

را در نظر می‌گیریم و فرض می‌کنیم که برای $i = ۱, ۲, \ldots, n$، $U_i = I_{(D_i)}$. چون تحت فرض صفر توزیع D_i پیوسته و متقارن حول صفر می‌باشد داریم:

$$P(U_i = -۱) = P(D_i < ۰) = \frac{۱}{۲} \, ,$$
$$P(U_i = ۱) = P(D_i > ۰) = \frac{۱}{۲} \, .$$

متغیرهای تصادفی $U_۱, U_۲, \ldots, U_n$ هم‌توزیع و مستقل می‌باشند. آماره $W = \sum_{i=۱}^{n} U_i R_i$ به آماره رتبه‌ای-نشانه‌ای ویلکاکسون شهرت دارد. اگر W خیلی بزرگ شود معلوم می‌شود که تعداد زیادی از یافته‌های نمونه تصادفی $D_۱, D_۲, \ldots, D_n$ در طرف راست صفر می‌باشند. و با صفر فاصله زیادی دارند. پس D نمی‌تواند متقارن باشد. یعنی فرض صفر را باید رد کرد. بنابراین ناحیه بحرانی به‌صورت $W \geq k$ است. در این آزمون چون فرض تقارن توزیع را داریم، اگر فرض شود که میانگین توزیع موجود باشد پس میانه و میانگین تعویض‌پذیر هستند. یعنی آزمون درباره میانه، همان آزمون درباره میانگین می‌باشد [۲، ۲.۹].

[1] Wilcoxon

۱.۳.۱ استفاده از آماره‌های رتبه علامت‌دار ویلکاکسون برای آزمون تقارن

آماره‌های رتبه علامت‌دار ویلکاکسون را می‌توان برای آزمون تقارن نیز در نظر گرفت، در صورتی‌که تنها فرض منظور شده این باشد که نمونه تصادفی از توزیع پیوسته استخراج شده باشد. اگر فرض صفر بیان کند که جامعه با میانه $M_۰$ متقارن است، آنگاه تحت فرض صفر توزیع آماره آزمون دقیقاً مانند آزمون درباره میانگین یا میانه است. اگر مثلاً با استفاده از آزمون دوطرفه، فرض صفر رد شود نتیجه‌گیری می‌کنیم که جامعه نامتقارن است.

۴.۱ آزمون جمعی-رتبه‌ای ویلکاکسون

دو نمونه تصادفی مستقل $X_۱, X_۲, \ldots, X_m$ و $Y_۱, Y_۲, \ldots, Y_n$ را از متغیرهای تصادفی پیوسته $X \sim F_X$ و $Y \sim G_Y$ با $G(y) = F(y-c)$ در نظر می‌گیریم. به عبارت دیگر این دو نمونه تصادفی یا هم‌توزیع هستند و یا اگر اختلافی داشته باشند اختلاف آن‌ها در اندازه گرایش به مرکز آن‌هاست. پارامتر مجهول c را پارامتر تغییر مبدأ می‌نامند. می‌خواهیم یکی از آزمون‌های زیر را انجام دهیم:

$$(a)\begin{cases} H_۰: & c = ۰ \\ H_۱: & c < ۰ \end{cases} \qquad (b)\begin{cases} H_۰: & c = ۰ \\ H_۱: & c > ۰ \end{cases} \qquad (c)\begin{cases} H_۰: & c = ۰ \\ H_۱: & c \neq ۰ \end{cases}$$

ما تنها در مورد آزمون (b) سخن می‌گوییم. دو آزمون دیگر نیز همانند آزمون (b) هستند. تحت فرض صفر نمونه ترکیبی $X_۱, X_۲, \ldots, X_m$ و $Y_۱, Y_۲, \ldots, Y_n$ را می‌توان به چشم یک نمونه $N = n + m$ تایی نگاه کرد. فرض می‌کنیم در این نمونه ترکیبی R_i رتبه X_i و S_j رتبه Y_j باشد. آنگاه (۱.۱) یک جایگشت برای $(۱, ۲, \ldots, N)$ می‌باشد.

$$Q = (S_۱, S_۲, \ldots, S_n \text{ و } R_۱, R_۲, \ldots, R_m) \; . \tag{۱.۱}$$

واضح است که تحت فرض صفر نمونه ترکیبی بالا یک نمونه تصادفی N تایی از توزیع F_X می‌باشند. آماره‌های $W_R = \sum_{i=۱}^{m} R_i$ و $W_S = \sum_{j=۱}^{n} S_j$ آماره‌های جمعی-رتبه‌ای ویلکاکسون هستند. ناحیه بحرانی آزمون $\begin{cases} H_۰: & c = ۰ \; , \\ H_۱: & c > ۰ \; , \end{cases}$ را به‌دست می‌آوریم. می‌گوییم هرگاه فرض $H_۱ \; : \; c > ۰$ درست باشد، آنگاه طبق خاصیت یکنوایی تابع توزیع برای هر عدد مثبت z داریم:

$$\begin{aligned} G(z) &= F(z-c) \leq F(z) \; , \\ P(Y \leq z) &\leq P(X \leq z) \; , \\ P(Y > z) &\geq P(X > z) \; . \end{aligned}$$

نامساوی آخر حاکی از این است که هر Y_i شانس آن‌را دارد که از هر X_i بزرگ‌تر باشد. به عبارت دیگر هر S_j شانس آن‌را دارد که از هر R_i بزرگ‌تر باشد. بنابراین هرگاه $W_S = \sum_{j=1}^{n} S_j$ خیلی بزرگ شود فرض صفر را رد می‌کنیم. پس ناحیه بحرانی به‌صورت $\{W_S \geq k\}$ می‌باشد. برای این آماره داریم:

۱. $$E(W_S) = \frac{n(m+n+1)}{2}\ ,\ \mathrm{V}(W_S) = \frac{mn(m+n+1)}{12}$$

۲. W_s نسبت به $E(W_S)$ متقارن است.

۳. $W_R - \frac{m(m+1)}{2}$ و $W_S - \frac{n(n+1)}{2}$ هم‌توزیع هستند [۲، ۱.۷].

۱.۵ آزمون من-ویتنی

دو نمونه تصادفی $X_1, X_2, \dots, X_m$ و $Y_1, Y_2, \dots, Y_n$ را که در آزمون ویلکاکسون داشتیم، بار دیگر در نظر می‌گیریم. تمام دوتایی‌های (X_i, Y_j) را که تعداد آن‌ها mn می‌باشد تشکیل می‌دهیم. با فرض پیوستگی و صعودی بودن F_X، در هر جفت داریم $X_i < Y_j$ یا $X_i > Y_j$.
حال W_{xy} و W_{yx} را به‌صورت زیر تعریف می‌کنیم:
W_{xy} برابر با تعداد (X_i, Y_j) ها با فرض $X_i < Y_j$ و W_{yx} برابر با تعداد (X_i, Y_j) ها با فرض $X_i > Y_j$ می‌باشد. دو آماره شمارشی W_{xy} و W_{yx} را که مجموع آن‌ها برابر با mn است، آماره‌های من-ویتنی[1] می‌نامند. واضح است که در آزمون $\begin{cases} H_0: & c = 0\ , \\ H_1: & c > 0\ , \end{cases}$ فرض صفر را موقعی رد می‌کنیم که W_{xy} خیلی بزرگ شود. یعنی تعداد زیادی از Y_jها بزرگتر از X_i ها شوند. هر آزمونی را با ناحیه بحرانی $\{W_{xy} \geq k\}$ آزمون من-ویتنی می‌گویند.

۱.۵.۱ داده‌های برابر

آزمون‌های ویلکاکسون و من-ویتنی بر این فرض استوار بودند که مقادیر مشاهده شده Y_i و X_i متمایز باشند. آزمون علامت نیز بر این استوار بود که که توزیع در اطراف میانه پیوسته باشد. از آن‌جایی‌که اندازه‌گیری‌ها با دقت محدودی اندازه‌گیری می‌شوند، امکان وجود داده‌های برابر (در آزمون ویلکاکسون و من-ویتنی) یا برابری مشاهدات با میانه (در آزمون علامت) وجود دارد. پس حالت داده‌های برابر و یا برابری داده‌ها با میانه ممکن است رخ دهند و رخ می‌دهند. طرز عمل در این حالات را برای یکی از آزمون‌ها شرح می‌دهیم، برای بقیه با استدلالی مشابه عمل می‌کنیم.
مثلاً اگر در آزمون من-ویتنی متوجه شویم زوج‌های برابر وجود دارند، در این حالت یک روش این است که آزمون را دوبار اجرا کنیم. در آزمون اول برای هر یک از برابری‌های $X_i = Y_i$ فرض می‌کنیم $X_i < Y_i$ و در آزمون دوم فرض می‌کنیم $X_i > Y_i$. اگر مساحت دمی هر دو آزمون تقریباً برابر باشند، آنگاه این برابری داده‌ها اهمیت چندانی ندارد. شیوه معمولی که در این حالت از آن پیروی می‌شود این است که از تفاضل‌های صفر صرف‌نظر می‌کنیم و مقدار N را به همان اندازه کاهش می‌دهیم و آزمون را با داده‌های کاهش یافته اجرا می‌کنیم. لذا استنباط‌ها، روی تفاضل‌های مخالف صفر مشاهده شده، شرطی است. اگر مساحت دمی اختلاف قابل ملاحظه‌ای داشته باشد، آنگاه این برابری

[1] Mann-Whitney

داده‌ها ممکن است به‌طور جدی روی استنباط‌ها تاثیر بگذارند. این نشان می‌دهد که ممکن است داده‌های ما کافی نباشند.
اگر در این حالت نخواهیم در مورد داده‌ها تجدید نظر کنیم و انجام آزمون با این داده‌ها ضروری باشد، می‌توان نیمی از صفرها را به‌عنوان علامت مثبت و نیمی دیگر را به عنوان علامت منفی در نظر گرفت، همچنین می‌توان به‌طور تصادفی علامت‌های مثبت و منفی را به صفر نسبت داد. همچنین می‌توانیم از یک روش اکیداً محافظه‌کارانه استفاده کنیم، مثلاً به تمام صفرها آن علامتی را نسبت دهیم که کمترین اثر را در رد H_0 دارد [۱، ۸.۰۹].

قضیه ۱.۵.۱. آزمون من–ویتنی و آزمون جمعی–رتبه‌ای ویلکاکسون معادل می‌باشند، یعنی با آزمون اول H_0 را رد می‌کنیم اگر و تنها اگر با آزمون دوم همین‌کار را انجام دهیم [۲، ۱.۰۸].

۶.۱ آزمون میانه

در این آزمون می‌خواهیم بیازماییم که متغیرهای مستقل $X_1, X_2, \ldots, X_k$ هم‌توزیع هستند (فرض H_0) در برابر این‌که این متغیرها دارای توزیع‌های گوناگون هستند. ما در این آزمون در مورد توزیع‌ها هیچ فرضی را در نظر نمی‌گیریم. این کلیت در اکثر مواقع کارایی آزمون را کاهش میدهد. تحت فرض صفر، همانندی جامعه‌ها، یک نمونه تصادفی به حجم $N = \sum_{i=1}^{k} n_i$ از جامعه مشترک داریم. M میانه کل نمونه‌های ادغام شده، برآوردی برای میانه این جامعه مشترک است. بنابراین یک مشاهده از هرکدام از K نمونه، همان‌قدر که محتمل است بزرگ‌تر از M باشد، همان‌قدر نیز احتمال دارد که کوچک‌تر از آن باشد. بنابراین مجموعه N مشاهده فرض صفر را تاییدخواهد کرد اگر برای هر یک از K نمونه حدود نیمی از مشاهدات آن نمونه کوچک‌تر از میانه نمونه‌ای کل باشند.
میانه نمونه‌ای کل به‌صورت مشاهده‌ای در نمونه ادغام شده مرتب است. اگر N فرد باشد با رتبه $\frac{N+1}{2}$ تعریف می‌شود و اگر N زوج باشد به‌صورت هر عددی بین دو مشاهده با رتبه‌های $\frac{N}{2}$ و $\frac{N+2}{2}$ تعریف می‌شود. در این‌صورت برای هر نمونه به‌طور جداگانه، مشاهدات برحسب اینکه کوچک‌تر از M باشند یا نه، دو حالتی هستند. متغیر تصادفی U_i را به‌عنوان تعداد مشاهداتی از نمونه شماره i ام تعریف می‌کنیم که کمتر از M باشند و فرض می‌نماییم که t تعداد کل مشاهداتی را نشان دهد که کمتر از M هستند. بنابراین با توجه به تعریف M داریم:

$$t = \sum_{i=1}^{k} u_i = \begin{cases} \frac{N}{2} & \text{اگر } N \text{ زوج}, \\ \\ \frac{N-1}{2} & \text{اگر } N \text{ فرد}. \end{cases}$$

تحت فرض صفر، تمام $\binom{N}{t}$ مجموعه ممکن از t مشاهده برای بودن در رسته کمتر از M، $t = 0, 1, 2, \ldots, N$ ، متساوی‌الاحتمال هستند، و تعداد موارد دوحالتی کردن با این برآمد نمونه‌ای خاص، مساوی $\prod_{i=1}^{k} \binom{n_i}{u_i}$ است.