بسم الله الرحمن الرحیم

دانشگاه پیام نور

دانشکده علوم

گروه آمار

عنوان پایان نامه

آماره ی یو و کاربردهای آن

برای دریافت درجه کارشناسی ارشد

در رشته آمار ریاضی

مولف

صدیقه همتی

استاد راهنما

دکتر نرگس عباسی

ماه و سال انتشار

دی ماه ۱۳۸۷

دانشگاه پیام نور

بسمه تعالی

تصویب پایان‌نامه

پایان‌نامه تحت عنوان:

آماره‌ی یو و کاربردهای آن

که توسط خانم **صدیقه همتی** در مرکز شیراز تهیه و به هیأت داوران ارائه گردیده است مورد تأیید می‌باشد. تاریخ دفاع: **۱۳۸۷/۱۰/۱۰** نمره: **۱۸/۷۵** درجه ارزشیابی: **عالی**

اعضای هیأت داوران	مرتبه علمی	امضاء
۱- استاد راهنما: **سرکار خانم دکتر نرگس عباسی**	دانشیار	
۲- استاد مشاور: **سرکار خانم دکتر مینا توحیدی**	استادیار	
۳- استاد داور: **جناب آقای دکتر عبدالرضا بازرگان لاری**	استادیار	
٤- نماینده تحصیلات تکمیلی: **جناب آقای دکتر محمد بهرامی**	استادیار	

تقدیم به پدر عزیزم

که اولین آموزگار زندگیم است. او که دست‌هایش بوی تلاش و امید را می‌دهند و چشمانش امید را در دلم می‌نشاند و نور ایمانش همواره آرامش را برایم به ارمغان می‌آورد .

تقدیم به مادر صبورم

هم او که دریچه لطف و محبت را بر دیدگانم گشود و درس آزادگی و درست زیستن را به من آموخت. هم او که دریایی از صفاست و دستانش همیشه بوی بهار و جاودانگی را می‌دهند.

تقدیم به خواهر مهربانم

گل همیشه بهار زندگیم. او که لبخندش سرشار از احساس و دلش پر از امید است.

تقدیم به برادران خوبم

دوستان عزیز زندگیم، وجودشان برایم آرامش و نگاه مهربانشان سرشار از محبت است.

حمد و سپاس خداوندی را که گل ستایش از نسیم عنایت او بر چمن زبان می‌روید و مرغ نیایش به هوای رحمت بی‌نهایت او از آشیان دهان پر می‌گشاید و از باغ لطف و رضایت او دانه امید می‌روید.

حمد خدایی که اول همه‌ی آثار هستی اوست و قبل از او اولی نبوده و آخر است بی‌آن که پس از او آخری باشد، خدایی که دیده‌ی بندگان از دیدنش قاصر و اندیشه و فهم وصف‌کنندگان از وصفش عاجز است. به دست قدرتش آفریدگان را ایجاد کرد و آنان را بر اساس اراده‌ی خود صورت بخشید، آن‌گاه همه را در راه اراده خود راهی نمود، و در مسیر محبت و عشق به خود برانگیخت.

خداوندا همیشه یاور و پناه بندگانت باش و راه درست و مسیر حق را به آنان بنمای و دل‌هایشان را از نور ایمان به خود روشن کن و تواضع و خدمت صادقانه به خلق را به آنان بیاموز. **الهی آمین**

اینک که این رساله به انجام رسیده، ضمن سپاس‌گزاری از ایزد سبحان از همه‌ی استادان و صاحب‌نظران ارجمند به ویژه استاد گرانقدر سرکار خانم دکتر نرگس عباسی که با تعهد، خلوص نیت و دقت نظر فراوان هدایت و راهنمایی این رساله را برعهده داشتند و تجربه سال‌ها تحقیق و تفحص علمی‌شان را در اختیار اینجانب قرار دادند، صمیمانه قدردان و سپاس‌گزارم. سرافرازی، تندرستی و توفیق روزافزون ایشان را از خداوند متعال خواستارم.

از سرکار خانم دکتر مینا توحیدی که مشاوره رساله اینجانب را به عهده داشتند و از محضرشان علم آموخته‌ام، بی‌نهایت سپاس‌گزار و ممنونم و عزت و عافیت ایشان را از خداوند منان خواهانم.

از جناب آقای دکتر بازرگان لاری که قبول زحمت فرموده و داوری این رساله را بر عهده گرفتند، قدردان و متشکرم.

از ناظر تحصیلات تکمیلی جناب آقای دکتر محمد بهرامی که نظارت جلسه‌ی دفاعیه را بر عهده گرفتند کمال تقدیر و تشکر را دارم.

از کلیه اساتیدی که از محضرشان علم آموخته‌ام کمال تقدیر و تشکر را دارم.

همچنین از دوستان خوبم خانم‌ها: فاطمه مزارعی، سیده نسیبه موسوی، مریم شبانه، فریبا گودرزی، مریم معشوری، فاطمه حبیبی، هدی کاظمی، عصمت ناهیدی، اسما عموزاده، الهه قربانی، محبوبه بحرینی، مریم مرادی‌نسب، ایمان مرزوکی، مهناز رستمی، آمنه ثابت، صفیه آتش‌روز، فرزانه افشار، مریم موحدی‌نیا، فاطمه قوتی، دلارا سیاهپوش آقایان ایمان حفیظی، ایمان بهارلو، محمود قزلباش، هدایت حقیقت، فاتحی و طباطبایی صمیمانه تشکرو قدردانی می‌کنم و عافیت و سرفرازیشان را از خداوند بزرگ خواهانم.

چکیده

در بحث برآوردیابی نقطه‌ای، توجه به برآوردگرهای نااریب با کمترین واریانس همواره بخش مهمی را تشکیل می‌دهد. با پیدایش نظریه‌ی آماره‌ی یو این ویژگی در برآوردگرهایی از نوع آماره‌ی یو دیده شد. در این پایان‌نامه به مطالعه‌ی گسترده‌ای از آماره‌های یو پرداخته و مباحثی از قبیل برآوردیابی گشتاورها، گشتاورهای مرکزی، کاربرد آن‌ها در تعیین آماره‌های آزمون، تعیین توزیع حدی را در حیطه‌ی تحقیق‌مان قرار می‌دهیم. این پایان‌نامه در چهار فصل تنظیم شده است. فصل اول حاوی مقدمات و مثال‌هایی از آماره‌های یو است. فصل دوم نگاهی متمرکز بر روی روابط بین آماره‌های یو و گشتاورهای نمونه‌ای داریم. در فصل سوم و چهارم با هدف تعیین توزیع حدی آماره‌های یو، محاسبات تعیین واریانس آماره‌های یو ارائه می‌گردد.

فهرست

مقدمه

با شروع از یک برآوردگر نااریب که در آن از کمترین مشاهدات استفاده می‌شود، آماره‌ی یو[1] تشکیل می‌گردد. مبنای تئوری آماره‌ی یو اولین بار توسط هافدینگ[2] در سال (۱۹۴۸) گسترش داده شد. شرح جزئیات موضوع کلی در دنکر[3] (۱۹۸۵) و لی[4] (۱۹۹۰) توضیح داده شده است. قابل توجه است که در فراسر[5] (۱۹۵۷) فصل ۶ و (۱۹۵۲) و (۱۹۵۳) و سرفلینگ[6] (۱۹۸۰) فصل ۵ و لهمن[7] (۱۹۹۹) فصل ۶ به تفصیل به این موضوع پرداخته شده است.

.ویژگی‌های خاصی که در این گروه از برآوردگرها ظاهر می‌شود منجر به تغییر در معیارهای شناسایی توزیع‌ها، آماره‌های آزمون‌های پارامتری و ناپارامتری شده است. هدف اصلی که این پایان‌نامه دنبال می‌کند ارائه‌ی برآوردگرهایی به صورت فرمول‌هایی که قابلیت کاربرد به صورت مستقیم را دارند است. که مهمترین آن‌ها گشتاورهای مرکزی توزیع‌ها در نظر گرفته شده است. با یافتن فرمول‌های این گروه از برآوردگرها، توزیع حدی آن‌ها نیز قابلیت محاسبه را پیدا می‌کند. گروهی از خانواده‌ها که منجر به آماره‌های یویی که توزیع‌های تباهیده دارند مورد مطالعه قرار می‌گیرند.

این پایان‌نامه در چهار فصل تنظیم شده است. در فصل اول آماره‌ی یو در حالت یک نمونه‌ای و دو نمونه‌ای معرفی می‌شود و کاربردهای و مثال‌هایی از آماره‌های یو بیان می‌گردد. در فصل دوم نگاهی متمرکز برروی روابط بین آماره‌های یو و گشتاورهای نمونه‌ای داریم. در این فصل گشتاور مرکزی مرتبه‌ی rام را محاسبه و آماره‌ی یو مناسب برای آن ارائه شده است، سپس نشان داده‌ایم که در حالت حدی آماره‌های یوی برآورد گشتاورهای مرکزی به‌سمت گشتاورهای مرکزی نمونه‌ای میل می‌کنند. در فصل سوم

1. U-Statistics
2. W.Hoeffding
3. M.Denker
4. A.J.Lee
5. Fraser
6. Scrfling
7. Lehman

واریانس آماره‌ی یو در حالت یک نمونه‌ای و دو نمونه‌ای ارائه شده است. در فصل چهارم با استفاده از مباحث مطرح شده در فصل سوم توزیع حدی آماره‌های یو ارائه می‌گردد.

فصل اول

آماره‌های یو

۱-۱ مقدمه

در این فصل ابتدا آماره‌ی یو را معرفی و برای آشنایی بیشتر مثال‌هایی ارائه می‌شود، سپس آماره‌های یو برای برآوردیابی پارامترهای دو جامعه مورد بررسی قرار می‌گیرد، در پایان فصل نیز کاربردهایی از آماره‌ی یو مطرح می‌گردد.

۲-۱ آماره‌ی یو

با شروع از یک برآوردگر نااریب که در آن از کمترین مشاهدات استفاده می‌شود، آماره‌ی یو تشکیل می‌گردد.

فرض کنید $\mathscr{P}$ یک خانواده از اندازه‌ی احتمال بر روی یک فضای دلخواه قابل اندازه‌گیری باشد. تئوری کلی یک راه‌کار برای مسائل ناپارامتری ارائه می‌دهد. یعنی ما با خانواده‌ی خاصی روبرو نیستیم به عبارت دیگر $\mathscr{P}$ از یک خانواده‌ی بزرگ توزیع‌ها گرفته می‌شود فقط ممکن است با این محدودیت که پیوسته باشند یا گشتاورها موجود باشد، مواجه باشیم.

برای $P \in \mathscr{P}$، $\theta(P)$ را به عنوان تابعی حقیقی تعریف می‌کنیم لازم است بدانیم که این یک پارامتر قابل برآورد است. (گاهی اوقات پارامتر منظم نیز نامیده می‌شود).

هالموس[1] (۱۹۴۶) تعریف زیر را برای برآورد پذیری پارامتر $\theta(P)$ ارائه کرد.

1. Halmos

تعریف (۱)

$\theta(P)$ یک پارامتر برآورد پذیر نسبت به اندازه‌ی $\mathscr{P}$ است، اگر به ازای حداقل m متغیر تصادفی مستقل و هم توزیع (مثلا از توزیع $P \in \mathscr{P}$) تابع حقیقی اندازه پذیر $h(x_1,\ldots,x_m)$ وجود داشته باشد به‌طوری که به ازای $P \in \mathscr{P}$

$$E(h(X_1,\ldots,X_m))=\theta(P). \qquad (1)$$

در تعریف فوق m درجه‌ی $\theta(P)$ معرفی می‌شود.

نکته‌ای که باید به آن توجه کنیم این است که در این‌جا تابع h، یک تابع متقارن فرض می‌شود. در غیر این صورت اگر f یک برآوردگر نااریب نامتقارن برای $\theta(P)$ باشد، آنگاه متوسط f که برای همه‌ی جایگشت‌های متغیرها به‌کار برده شده محاسبه می‌گردد، آن‌را تبدیل به برآوردگری نااریب و متقارن می‌سازد.

$$h(x_1,\ldots,x_m)=\frac{1}{m!}\sum_{\pi\in\Pi_m} f(x_{\pi_1},\ldots,x_{\pi_m}) \qquad (2)$$

در این‌جا Π_m مجموعه‌ای متشکل از کلیه‌ی جایگشت‌های m متغیر تصادفی در نظر گرفته شده و جمع بر روی Π_m برای همه‌ی جایگشت‌های موجود بسته می‌شود. پس بدون اینکه از کلیت مسئله کم شود h را متقارن فرض می‌کنیم، مگر در موارد خاص که ذکر خواهد شد.

در تعریف (۲) مفهوم هسته و آماره‌ی یو ارائه می‌شود.

تعریف (۲)

برای یک تابع حقیقی اندازه‌پذیر $h(x_1,\ldots,x_m)$ و برای نمونه‌ی تصادفی $X_1,\ldots,X_n$ $(n \ge m)$ از توزیع $\mathscr{P}$ آماره‌ی یو با هسته‌ی h به صورت زیر تعریف می‌شود.

$$U_n = U_n(h) = \frac{(n-m)!}{n!}\sum_{P_{m,n}} h(X_{i_1},\ldots,X_{i_m}) \qquad (3)$$

در این‌جا جمع برروی $P_{m,n}$ (مجموعه‌ی همه‌ی $\frac{(n-m)!}{n!}$ جایگشت بسته می‌شود.) و $(i_1,\ldots,i_m)$ از $(1,2,\ldots,n)$ انتخاب می‌شود.

اگر هسته‌ی h متقارن باشد آماره‌ی یو به شکل زیر ساخته می‌شود

$$U_n = U_n(h) = \frac{1}{\binom{n}{m}} \sum_{C_{n,m}} h(X_{i_1}, \ldots, X_{i_m}) \tag{۴}$$

در این‌جا جمع بر روی مجموعه‌ی $C_{n,m}$ (مجموعه‌ی همه‌ی ترکیبات m از n عدد صحیح) به‌طوری‌که $i_1 < i_2 < \cdots < i_m$ از $(1,2,\ldots,n)$ انتخاب می‌شود.

یکی از ویژگی‌های بارز آماره‌ی یو این است که در صورتی که $\theta(P) = E_P(h(X_1,\ldots,X_m))$ برای همه $P \in \mathcal{P}$ وجود داشته باشد، آنگاه اگر $\mathcal{P}$ به اندازه‌ی کافی بزرگ باشد U_n بهترین برآوردگر نااریب پارامتر $\theta(P)$ است لی (۱۹۹۰)

برای همه‌ی توزیع‌های P اگر $\theta(P)$ برآوردپذیر باشد، آنگاه آماره‌های ترتیبی، یک آماره‌ی کافی و کامل برای $P \in \mathcal{P}$ است و با توجه به اینکه U_n یک تابع متقارن و نااریب از نمونه $X_1,\ldots,X_n$ می‌باشد پس تابعی از آماره‌های ترتیبی است. بنابراین با توجه به قضیه هادگیز[1]- لهمن بهترین برآورد مورد انتظار خواهد بود. به این معنی که هیچ برآورد نااریب دیگری بر پایه‌ی $X_1,\ldots,X_n$ برای $\theta(P)$ وجود ندارد به‌طوری‌که واریانس آن از واریانس U_n کوچکتر باشد. به عبارتی آماره‌های یو $UMVUE$ برای $\theta(P)$ هستند.

مثال ۱

فرض کنید $\mathcal{P}$ مجموعه‌ی توزیع‌های حقیقی با میانگین متناهی باشد. میانگین $\mu = \mu(P) = \int x dP(x)$ یک پارامتر برآوردپذیر با درجه ۱ است زیرا $f(x_1) = x_1$ یک برآورد نااریب برای میانگین است

$$E(X_1) = \mu$$

پس $h(x_1) = x_1$ یک هسته برای میانگین تعریف می‌شود. بر طبق محاسبه‌ی آماره‌ی یو، میانگین نمونه‌ای به صورت زیر حاصل می‌گردد.

1. Hodges

$$U_n = \frac{1}{\binom{n}{1}} \sum_{i=1}^{n} X_i = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X_n}.$$

مثال ۲

$\mathscr{P}$ مجموعه‌ی توزیع‌های حقیقی با گشتاور kام متناهی است. گشتاور kام

$$\mu_k = \int x^k \, dP(x)$$

یک پارامتر برآوردپذیر با درجه ۱ است زیرا $f(x_1) = x_1^k$ یک برآورد نااریب برای μ^k است.

$$E(X_1^k) = \mu_k$$

پس $h(x_1) = x_1^k$ یک هسته‌ی متقارن برای μ_k می‌باشد. بنابراین آماره‌ی یو به صورت زیر تشکیل می‌شود.

$$U_n = \frac{1}{\binom{n}{1}} \sum_{i=1}^{n} X_i^k = \frac{1}{n} \sum_{i=1}^{n} X_i^k.$$

مثال ۳

فرض کنید این‌بار $\theta(P) = \mu^2$ پارامتر مورد نظر باشد. $f(x_1, x_2) = x_1 x_2$ یک برآورد نااریب برای μ^2 است.

$$E(X_1 X_2) = \mu^2.$$

بنابراین $h(x_1, x_2) = x_1 x_2$ یک هسته‌ی متقارن با درجه‌ی ۲ برای μ^2 می‌باشد. آماره‌ی یو برابر است با

$$\begin{aligned} U_n &= \frac{1}{\binom{n}{2}} \sum_{i<j} X_i X_j \\ &= \frac{2}{n(n-1)} \sum_{i<j} X_i X_j. \\ &= \frac{1}{n(n-1)} \left[\left(\sum_{i=1}^{n} X_i \right)^2 - \sum_{i=1}^{n} X_i^2 \right]. \end{aligned} \tag{5}$$

مثال ٤

اگر $\mathscr{P}$ مجموعه‌ی همه‌ی توزیع‌های حقیقی با گشتاور مرتبه‌ی دوم متناهی باشد (گشتاور مرتبه‌ی دوم موجود باشد) سپس $\sigma^2 = \mu_2 - \mu^2$ نیز یک پارامتربرآوردپذیر با درجه‌ی ۲ است.

می‌توانیم μ_2 را با استفاده از X_1^2 برآورد کنیم و μ^2 را با استفاده از X_1X_2 برآورد کنیم. پس $f(x_1,x_2)=x_1^2 - x_1x_2$ یک برآورد نااریب برای σ^2 است

$$E\,(X_1^2 - X_1X_2) = \sigma^2 \qquad (6)$$

همان‌طور که ملاحظه می‌شود هسته‌ی $f\,(x_1,x_2)=x_1^2 - x_1x_2$ نسبت به شناسه‌های x_1 و x_2 متقارن نیست. با توجه به فرمول (۲) هسته‌ی متقارن به صورت زیر محاسبه می‌شود.

$$\begin{aligned} h(x_1,x_2) &= \frac{1}{2!}\big(f\,(x_1,x_2)+f\,(x_2,x_1)\big) \\ &= \frac{x_1^2 - 2x_1x_2 + x_2^2}{2} \\ &= \frac{(x_1-x_2)^2}{2} \qquad (7) \end{aligned}$$

و لذا آماره‌ی یو برابر است با

$$\begin{aligned} U_n &= \frac{1}{\binom{n}{2}}\sum_{i<j}\frac{(X_i-X_j)^2}{2} \\ &= \frac{2}{n(n-1)}\sum_{i<j}\frac{(X_i-X_j)^2}{2} \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2 \\ &= S_x^2. \qquad (8) \end{aligned}$$

این یک برآورد نااریب برای واریانس جامعه است. می‌توان نشان داد که هر ترکیب خطی از پارامترهای برآوردپذیر نیز قابل برآورد است. بنابراین با جایگزینی آماره‌های یوی متناظر می‌توان آماره‌های یو برای همه‌ی گشتاورها و کومولنت‌ها تشکیل داد.

در تعریف آماره‌ی یو برای پارامتر قابل برآورد، هیچ محدودیتی بر روی بعد فضایی که توزیع روی آن ساخته می‌شود وجود ندارد. بنابراین به ازای هر $P \in \mathscr{P}$ می‌توان برای یک توزیع بر روی صفحه یا یک فضای d بعدی تعریف شده بر طبق مشاهدات، یک بردار پارامتری را در نظر گرفته، و در صورت داشتن ویژگی برآوردپذیری، آماره‌ی یو را برای آن تشکیل داد.

می‌توان از آماره‌ی یو برای برآورد کوواریانس و گشتاورهای متقاطع جزئی بالاتر استفاده کرد.

۱-۳ آماره‌های یو برای دو جامعه

کاربرد مهم آماره‌ی یو برای پارامترهایی است که از دو جامعه گرفته شده‌اند ، این موضوع توسط لهمن (۱۹۵۱) مورد بحث قرار گرفته است. اساس این نظریه برای حالت دو نمونه‌ای است که در این قسمت به توضیح آن می‌پردازیم. در این حالت $\mathscr{P}$ یک خانواده از جفت اندازه‌ی احتمال (F,G) می‌باشد. توجه کنید که نمونه‌ی تصادفی $X_1,\ldots,X_{n_1}$ مستقل با توزیع $F(x)$ و $Y_1,\ldots,Y_{n_2}$ مستقل و دارای توزیع $G(y)$ هستند. فرض کنید تابع $h\,(x_1,\ldots,x_{m_1},y_1,\ldots,y_{m_2})$ هسته و $\mathscr{P}$ مجموعه‌ی تمام جفت اندازه‌های احتمال باشد، به‌طوری‌که عبارت زیر

$$\theta=\theta(F,G)=E_{F_1,F_2}[\,h(X_1,\ldots,X_{m_1},Y_1,\ldots,Y_{m_2})\,] \qquad (۹)$$

متناهی باشد.

همانند بخش ۱-۲، بدون اینکه از کلیت مسئله کم شود فرض می‌کنیم که h تحت فرض استقلال جایگشت‌های $x_1,\ldots,x_{m_1}$ و $y_1,\ldots,y_{m_2}$ متقارن است. منظور از استقلال آن است که اندیس‌های x و y به‌طور جداگانه تغییر می‌کنند. در این حالت درجه‌ی پارامتر به صورت (m_1,m_2) نشان داده می‌شود.

پس از تعریف هسته و درجه‌ی پارامتر، حال آماره‌ی یو را در حالت دو نمونه‌ای تعریف می‌کنیم

$$U_{n_1 n_2}=U(h)=\frac{1}{\binom{n_1}{m_1}\binom{n_2}{m_2}}\sum h(X_{i_1},\ldots,X_{i_{m_1}},Y_{j_1},\ldots,Y_{j_{m_2}}) \qquad (۱۰)$$

به طوری‌که جمع بر روی کلیه‌ی حالات که تعداد آن $\binom{n_1}{m_1}\binom{n_2}{m_2}$ است، بسته می‌شود به طوری‌که $1 \le i_1 \le i_2 \le \ldots \le i_{m_1} \le n_1$ و $1 \le j_1 \le j_2 \le \ldots \le j_{m_2} \le n_2$. همانند قبل آماره‌ی یو یک برآورد نااریب برای پارامتر θ است.

۱-۴ آزمون‌های مقایسه‌ی دو توزیع

در مبحث آزمون سازی، آزمون‌های مقایسه‌ی دو توزیع بر پایه‌ی آماره‌ی یو وجود دارد که فرض صفر آن تساوی دو توزیع می‌باشد، $H_0 : F = G$. آماره‌ی آزمون با تغییر شرایط و اطلاعات توزیع‌های مطرح شده در فرضیه، تغییر می‌یابد. در زیر به بررسی حالت‌های مختلف می‌پردازیم.

۱-۴-۱ مقایسه میانگین‌ها در حالت دو نمونه‌ای

فرض کنید F و G توابع توزیع پیوسته با واریانس متناهی باشند. $\theta = E(X) - E(Y)$ یک پارامتر برآورد پذیر با درجه‌ی $(m_1, m_2) = (1,1)$ است که هسته‌ی آن $h(x_1, y_1) = x_1 - y_1$ می‌باشد. با توجه به تعریف، آماره‌ی یو به صورت زیر ساخته می‌شود

$$U_{n_1,n_2} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (X_i - Y_j) = \overline{X}_{n_1} - \overline{Y}_{n_2} \qquad (11)$$

۱-۴-۲ آزمون دو نمونه‌ای من ویتنی[1] و ویلکاکسون[2]

با توجه به تعریف آزمون دو نمونه‌ای من ویتنی(۱۹۴۷) و ویلکاکسون(۱۹۴۵) این آزمون بر پایه‌ی آماره‌ی یو به‌صورت زیر در نظر گرفته می‌شود. فرض کنید F و G توابع توزیع پیوسته باشند. $\theta = P(Y < X)$ یک پارامتر برآوردپذیر با درجه‌ی $(m_1, m_2) = (1,1)$ است که هسته‌ی آن $h(x, y) = I(y < x)$ می‌باشد. با توجه به تعریف، آماره‌ی یو به صورت زیر ساخته می‌شود

1. Mann-Whitney
2. Wilcoxon

$$U_{n_1,n_2}=\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} h(X_i,Y_j)=\frac{W}{n_1 n_2} \qquad (۱۲)$$

به طوری‌که W تعداد جفت‌های (X_i,Y_j) می‌باشد که در آن‌ها $X_i > Y_j$ است. تحت فرض صفر $F=G$، $\theta=\frac{1}{2}$ می‌باشد. لازم به ذکر است که از آماره‌ی U_{n_1,n_2} فقط زمانی می‌توان استفاده کرد که فرض مقابل آزمون $P_{\{F,G\}}(X>Y)\neq\frac{1}{2}$ باشد.

۲-۴-۲ یک روش متفاوت برای آزمون مقایسه‌ی دو توزیع

فرض کنید F و G توابع توزیع پیوسته هستند آنگاه:

$$\theta=P(X_1<Y\,,X_2<Y)+P(Y_1<X\,,Y_2<X)$$

یک پارامتر برآوردپذیر با هسته‌ی

$$h(x_1,x_2,y_1,y_2)=I(x_1<y_1\,,x_2<y_1)+I(y_1<x_1\,,y_2<x_1)$$

تعریف می‌شود. هسته‌ی متقارن شده شامل چهار جمله می‌شود،

$$\begin{aligned}h^*(x_1,x_2,y_1,y_2)&=I(x_1<y_1\,,x_2<y_1)+I(y_1<x_1\,,y_2<x_1)\\&+I(x_1<y_2\,,x_2<y_2)+I(y_1<x_2\,,y_2<x_2),\end{aligned}$$

امید ریاضی آن برابر است با:

$$\begin{aligned}\theta&=P(X_1<Y\,,X_2<Y)+P(Y_1<X\,,Y_2<X)\\&=\frac{2}{3}+\frac{1}{2}\int(F(x)-G(y))^2d(F(x)+G(y))\end{aligned} \qquad (۱۳)$$

(اثبات برابری (۱۳) در ضمیمه‌ی الف آمده است.)

آماره‌ی یو به‌صورت زیر ساخته می‌شود

$$U_{n_1,n_2}=\frac{1}{\binom{n_1}{2}\binom{n_2}{2}}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}h^*(X_{i_1},X_{i_2},Y_{j_1},Y_{j_2})$$

در این جا فرض $F=G$ معادل با $\theta=\frac{2}{3}$ است. در این آزمون فرض H_0 را رد می‌شود، اگر آماره‌ی یو خیلی بزرگ باشد.