بسمه تعالی

دانشگاه شهید بهشتی

دانشکده علوم ریاضی - گروه آمار

پایان نامه جهت دریافت درجه کارشناسی ارشد آمار

موضوع:

نظریه همبستگی کلاسیک

و

همبستگی ناهمگن و خم های همبستگی

استاد راهنما:

آقای دکتر سیامک نوربلوچی

استاد مشاور:

آقای دکتر محمد رضا مشکانی

استاد داور:

آقای دکتر عبدالرحیم شهلایی

نگارش:

محمدرضا ادراکی

شهریور ۱۳۷۷

تقدیم به :

خانواده عزیزم که همواره مرا در تمام مراحل زندگی یاری کرده‌اند.

بسمه تعالی

تاریخ ....
شماره ....
پیوست..................

دانشگاه شهید بهشتی

صورتجلسه دفاع از پایان نامه

:::::::::::::::::::::::::::::::

جلسه هیئت داوران ارزیابی پایان نامه آقای/~~خانم~~ محمدرضا ادراکی

به شناسنامه شماره ۱۰۵ صادره از مشهد متولد ۱۳۵۱/۱/۱۵

دانشجوی دوره کارشناسی ارشد ناپیوسته رشته آمار ــ محض

با عنوان نظریه همبستگی کلاسیک و نظریه همبستگی ناهمگن وخمهای همبستگی

به راهنمائی دکتر سیامک نوربلوچی طبق دعوت قبلی در تاریخ ۷۵/۶/۳۱

تشکیل گردید و براساس رای هیأت داوری وباعنایت به ماده ۴ آئین

نامه کارشناسی ارشد مورخ ۶۸/۸/۲۸ پایان نامه مزبور بانمره هیجده تمام

و درجه عالی مورد تصویب قرارگرفت.

۱ـ آقای دکتر سیامک نوربلوچی استاد راهنما

۲ـ " " محمدرضا مشکانی استاد مشاور

۳ـ " " عبدالرحیم شهلایی داور

۴ـ

۵ـ

عنوان **صفحه**

فصل اول : مقدمه و کلیات

عنوان **صفحه**

❋ فصل دوم : توزیع نرمال دو متغیره و برآورد درستنمایی ماکزیمم ضریب همبستگی

عنوان **صفحه**

فصل سوّم: تابع توزیع ضرایب همبستگی نمونه

عنوان **صفحه**

عنوان **صفحه**

پیشگفتار:

با این که بیش از یک قرن از معرفی مفهوم همبستگی می‌گذرد و در طی این مدت معیارهای زیادی برای اندازه‌گیری شدت ارتباط بین دو متغیر ارائه شده است ، ضریب همبستگی گشتاور حاصلضرب پیرسونی ، هنوز جایگاه خود را حفظ کرده است . ضریب همبستگی پیرسون ، دربسیاری اززمینه‌ها، برای آزمایشهای مبتنی بر مشاهدات ، یک ابزار قوی آماری ، برای اندازه‌گیری شدت رابطه خطی بین دو متغیر، محسوب می‌شود . دراین تحقیق تکامل نظریه همبستگی را بطور اعّم و ضریب همبستگی پیرسون را بطور اخّص (در توزیع نرمال دو متغیره) ، مورد بحث قرار داده‌ایم . نهایتاً ، خم همبستگی ، که همبستگی ناهمگن را بر حسب واریانس تبیین شده بوسیله رگرسیون برای هر مقدار متغیر مستقل ، موضعاً محاسبه می‌کند را معرفی کرده‌ایم .

فصل اوّل را به تکامل نظریه همبستگی و تاریخچه آن اختصاص داده‌ایم .

منابع اصلی این فصل پیرسون (۱۹۹۰)،بیرشک (۱۳۷۵) وعمیدی (۱۳۶۸) می‌باشد ، که جهت کامل کردن رساله از آنها استفاده کرده‌ایم .

در فصل دوّم برآورد درستنمایی ماکسیمم ضریب همبستگی از جامعه‌ای نرمال دو متغیره ، را بدست آورده‌ایم و تفسیری هندسی از آن ، به عنوان تابعی از زاویه بین دو بردار مشاهده ، ارائه داده‌ایم . مطالب این فصل بیشتر جهت یادآوری هستند ، منبع اصلی این فصل اندرسون (۱۹۸۴) می‌باشد.

در فصل سوّم توزیع ضریب همبستگی پیرسون را بدست آورده‌ایم و نشان داده‌ایم که ضریب همبستگی پیرسون ، برآوردگری اریب برای ρ، می‌باشد . همچنین برای ضریب همبستگی جامعه ، ρ، نظریه آزمون فرضیه های کلاسیک را ارایه کرده ایم .

فصل چهارم را به استنباط بیزی برای ضریب همبستگی در جامعه نرمال اختصاص داده‌ایم و توزیع تقریبی و دقیق ضریب همبستگی رابا فرض پیشین مرجع و پیشین جفریز بدست آورده‌ایم .

در فصل پنجم ، خم همبستگی را معرفی کرده‌ایم و برآورد نزدیکترین k همسایه رابرای ، ۴۰۰۰ زوج مشاهده ازیک مدل نظری (مثال ۱.۵) ، و سپس برای مجموعه داده‌های مدل تبدیل یافته محاسبه کرده ایم . منبع اصلی این فصل داکسام (۱۹۹۴) می‌باشد.

« فصل اوّل »

مقدمه وکلیات

۱.۱. پیدایش و تکامل نظریه همبستگی

گالتون[۱](۱۸۸۹)، از نخستین کسانی است که کارهایی در باره تعریف مفاهیمی چون همبستگی[۲] و برگشت[۳] (رگرسیون) انجام داده است و شیوه‌هایی برای اندازه‌گیری تاثیرات وراثت عرضه کرده و ابزارهایی برای این منظور فراهم آورده است . در واقع گالتون پدیده آماری رگرسیون را در حدود سال ۱۸۷۵ ، در جریان آزمایشهایی که با بذر نخود برای تعیین قانون وراثت در طول و وزن انجام می‌داد، کشف کرد.وی با استفاده از ۱۰۰ بذر ماده ، از هفت رده که از روی اندازه‌های قطر آنها معین کرده بود و اعداد بدست آمده از هر رده از زاده‌های جدید، متوجه شد که میانگین قطر بذرهای زاده هر رده از بذرهای ماده ، تقریباً بر روی یک خط راست واقع می‌شوند. علاوه بر این ، میانگین قطر زاده‌های رده‌های ماده های بزرگتر، کوچکتر از میانگین قطر بذرهای ماده ، و میانگین قطر زاده‌های کوچکتر از میانگین قطر رده ماده بزرگتر بوده اند ، و این امر ازگرایش اندازه‌های « میانگین » قطر زاده‌ها به « برگشتن » به اندازه‌ای می نمود که شاید بتوان آن را میانگین قطر اجدادی نامید . وی ابتدا این

۱-*Galton , F* ۲-*Correlation*

۳-*Regression*

پدیده را « عکس [1]» و سپس آن را « رگرسیون » نامید .

وی با استفاده از شیوه‌هایی که برای اندازه‌گیری تاثیر وراثت بکار می برد ، در مساله اندازه‌گیری درجه وابستگی [2] میان اندازه ی دو اندام متفاوت یک فرد، به مفهومی از « ضریب هم رابطگی » رسید که اندازه‌ای از درجه بستگی میان دو ویژگی بود، وتشخیص داد که r اندازه « عکس » یا «رگرسیون» ،مضربی از این گونه هم رابطگی [3]یا همبستگی است که کاربرد وسیعی دارد .به سبب پیچیدگی ریاضیات مورد نیاز برای مطالعه همبستگی و رگرسیون ،گالتون نتوانست صورت بندی ریاضیاتی این مفاهیم را بیابد .اما پیرسون[4] از پس این کار برآمد .

در ۱۸۹۲ پیرسون درباره تغییرات و همبستگی ، سخنرانی کرد و سخنانش در « مذاکرات فلسفی [5]» (۱۹۰۲)انتشار یافت .

در آن زمان وی مشغول یافتن نظریه عمومی همبستگی نرمال برای سه ، چهار و n متغیر بود ، و آرام آرام به سوی نظریه کلی همبستگی چوله [6] و رگرسیون غیرخطی پیش می‌رفت که تا سال ۱۹۰۵ انتشار نیافت .

همان گونه که مفهومها و شیوه‌های همبستگی و رگرسیون را گالتون معرفی کرد و روشهای مقدماتی حسابی و ترسیمی را (با استفاده از بعضی میانه‌ها و چارکهای داده‌هایی که در دست بود) برای بدست آوردن ضریب رگرسیون یا «شاخص همبستگی » ، r، بر اساس یک نمونه طرح‌ریزی کرد ، پیرسون ،

۱-*Reversion*

۲-*association*

۳-*Co_ relatio*

۴-*pearson,k*

۵-*philosophical Transactions*

۶-*skew correlation*

بوسیله آنچه امروز اصطلاح روش درستنمایی ماکسیمم [1] نامیده می‌شود ، نشان داد که «بهترین برآورد ضریب همبستگی» (ρ) برای یک توزیع نرمال دو متغیری توسط ضریب همبستگی حاصل ضرب گشتاورِ نمونه‌ای

$$r=\frac{\Sigma x\,y}{NS_x\,S_y}=\frac{\Sigma x\,y}{\sqrt{\Sigma x^{2}\Sigma y^{2}}}$$

بدست می‌آید. اصطلاح « ضریب همبستگی » را ظاهراً ا ج ورث در ۱۸۹۲ بکاربرده است ، امّا مقدار r، که با معادله بالا داده شده به « ضریب همبستگی پیرسون » معروف است . به عنوان مثال می‌توان به مقاله « رگرسیون ، وراثت ، تولید مثل ایلخی (یا آمیزش بی حق انتخاب [*panmixia*])» (۱۸۹۶)که نخستین مقاله اساسی پیرسون درباره نظریه همبستگی و کاربرد آن در نظریه وراثت بود مراجعه نمود. در مقاله فوق پیرسون نشان داده است که چگونه ممکن است « بهترین مقدار » r را به نحوی مناسب از انحرافهای معیار نمونه S_y و S_x و یکی از هر دو مقدار S_{x-y} یا S_{x+y} برآورد کرد و از محاسبه گشتاور حاصلضرب نمونه ($\frac{\Sigma xy}{N}$) پرهیز نمود .

وی در بحثی که در حالت سه متغیری پیش کشید، آنچه که امروزه ضرایب همبستگی «چندگانه» و «جزئی» نامیده می‌شود را به صراحت بر حسب سه ضریب همبستگی r_{12} و r_{13} و r_{23} بیان کرد.فرمول ا ج ورث (۱۸۹۲) برای توزیع نرمال سه متغیری را بصورتی بهتر و با نمایش ماتریسی از نو بیان کرد و آنرا به حالت همبستگی نرمال p متغیره بسط داد و به صورتی درآورد که محاسبات را در

۱-*maximum likelihood*

حیطه توانایی کسانی قرارداد که فاقد ورزیدگی در ریاضیات پیشرفته هستند .

در مقاله فوق بیان کرده است هرگاه توزیع مشترک تعدادی از متغیرهای X_p ,.., $X_۲$, $X_۱$،

($p\geq ۲$) توزیع نرمال چند متغیری باشد، آنگاه ضرایب همبستگی جامعه یعنی ρ_{ij}

($i \neq j$ و $i,j=۱,۲,...,p$)، درجات همبستگی میان این خصلتها را در جامعه نشان می‌دهند . متغیرهای

X_i و X_j مستقل از یکدیگرند اگر و فقط اگر $\rho_{ij}=۰$ ،وکاملاً به یکدیگر وابسته‌اند اگر و فقط اگر $\rho_{ij}=\pm ۱$.

نظریه عمومی همبستگی چندگانه و جزئی را ج .آدنی یول ، دستیار ریاضی او ، در دو مقاله در ۱۸۹۷ ادامه داد . یول اولین کسی بود که برای ضرایب همبستگی جزئی ، که خودش برای آنها اصطلاح «ضرائب همبستگی خالص » را بکار می‌برده است ، صورت ریاضی بیان نمود.

بر آنچه پیرسون ضرایب رگرسیون دوگانه نامیده بود، یول نام رگرسیونهای خالص گذاشت ،که اکنون آنها را ضرایب رگرسیون جزئی می‌نامند.

منشا اصطلاحات « همبستگی چندگانه » و « همبستگی جزئی » مقاله‌ای است که پیرسون ،با الیس لی نوشت و در ۱۸۹۷ در انجمن سلطنتی قرائت کرد .

پیرسون سعی فراوانی نمود تا روشهای همبستگی را به داده‌های دومتغیری که بر حسب هر خصلت به دو طبقه ، یا بیشتر، طبقه‌بندی شده بودند تعمیم دهد . وی (در ۱۹۰۰) در مقاله « همبستگی خصلتهایی که از جنبه کمیّتی قابل اندازه‌گیری نیستند» ضریب r_t که موسوم به « ضریب چهار دسته‌ای[۱]» است را معرفی کرد .

۱-*tetrachoric*

استودنت (و.س . گوست [۱])،(۱۹۰۷)، استنباط کرده بود که در نمونه های تصادفی با اندازه n،از توزیع نرمال دو متغیری ، توزیع r، به صورت متقارن در حول صفر توزیع می‌شود . استودنت در مقاله دیگری با عنوان ، «خطاهای احتمالی میانگین» که در ۱۹۰۸ منتشر ساخت ، توزیع S^2 را در نمونه‌های تصادفی از توزیع نرمال دو متغیری بدست آورد . این کشفها جوانه های علمی و تجربه‌ای بود که «استودنت» در آزمایشگاه زیست سنجی پیرسون در لندن بدست آورده‌بود . امّا پیرسون و همکارانش هیچ علاقه‌ای به بسط نظریه آماری و شیوه‌های خاص تحلیل نتایج آزمایشهای کوچک نمونه ای نداشتند .

در ۱۹۱۴ دستنویس مقاله‌ای به پیرسون رسید که درآن ر.ا.فیشر[۲] توزیع نمونه‌ای r را در نمونه‌های با هر اندازه $n \geq ۲$ و از جامعه نرمال دو متغیری با هر درجه همبستگی $۱+ \geq \rho \leq ۱-$ بدست آورده بود. پیرسون از این نوشته با شور و شوق استقبال کرد و به فیشر «از ته دل برای توفیق در بدست آوردن صورت فعلی توزیع r» شادباش گفت و اظهار داشت که « هرگاه تحلیل ، چنان که بسیار محتمل است ، درست باشد، وی از انتشار مقاله در مجله بیومتریکا بسیار مسرور خواهد شد.» یک هفته بعد به فیشر نوشت : « اکنون من مقاله شما را کاملاً خوانده‌ام و فکر می‌کنم که از پیشرفتی بسیار ممتاز نشان داشته باشد... از منتشر ساختن آن بسیار مسرور خواهم شد ... مقاله در شماره آینده (مه ۱۹۱۵) منتشر خواهد شد ... ای کاش وقت داشتید که صفحات آخر را کمی بسط دهید ... دوست دارم شاهد کوششی باشم برای تعیین آنکه به ازای چه مقدارهای ρ می‌توانیم فرض کنیم که توزیع عملاً نرمال است ».

۱-*william sealy Gosset*

۲-*Ronald A. Fisher*