بسم الله الرحمن الرحيم

دانشگاه فردوسی مشهد

پایان نامه

« جهت دریافت درجه کارشناسی ارشد آمار ریاضی »

موضوع :

مقایسه توابع توان آزمون رگرسیون چند متغیره خطی و

تحلیل واریانس بدون اثر متقابل

استاد راهنما : آقای دکتر حسن صادقی
استاد مشاور : آقای دکتر علی مشکانی
استاد داور : آقای دکتر غلامحسین شاهکار

نگارش:

علیرضا قدسی

بهمن - ۷۶

تقدیم به :

ساحت مقدس ثامن الحج حضرت رضا (سلام الله علیه)

که در جوار بارگاه ملکوتیش

توفیق کسب علم و دانش یافتم .

و تقدیم به :

مادر و پدر بزرگوارم

که با تحمل مشکلات فراوان و فراهم نمودن شرایط

مطلوب ، مرا در طی نمودن مقاطع تحصیلی یاری نمودند .

تشکّر و قدردانی :

شکر و سپاس خداوند متعال را که موفّقیتهایم نتیجهٔ لطف و رحمت بی پایان اوست .

در اینجا از کلیه کسانی که مرا در تهیه این مقوله یاری نموده اند تشکر می نمایم . و برای تمام آنها از خداوند متعال طلب مغفرت و قبولی طاعات دراین ماه مبارک را می نمایم .

از جناب آقای دکتر **حسن صادقی** استاد راهنما ی اینجانب که این مجموعه را مرهون زحمات ایشان می دانم که با صبر و حوصله با مسائل برخورد نمودند تشکر نمایم . و نیز از جناب آقای دکتر **علی مشکانی** که مشاوره پایان نامه را تقبل نمودند و از جناب آقای دکتر **غلامحسین شاهکار** به عنوان استاد داور که با حوصله تمام رساله را مطالعه و نکات لازم را تذکر داده اند ، سپاسگزارم .

لازم است که از سرکار خانم حسینی (منشی گروه آمار) که در تایپ مطالب زحمتهای فراوان کشیده اند قدردانی و تشکر نمایم .

برخود می دانم که ارادت قلبی خود را نسبت به جناب آقای دکتر **علی رضا فتوحی** که در دورهٔ کارشناسی مشوق اینجانب بودند و قبولی در مقطع کارشناسی ارشد را مدیون ایشان می باشم ، ابراز نمایم .

از مسئولین کتابخانه آقای اتحاد و آقای داود نژاد و از دوستان عزیز آقای رخشانی و آقای قاسمیان نیز قدردانی می نمایم .

از زحمات بی دریغ پدر و مادر عزیزم که تاکنون برایم کشیده اند و از خواهر و برادرانم و همچنین از آقای دکتر **محمد احمد آبادی** ، دایی گرامی به خاطر

دلسوزی هایشان نسبت به اینجانب متشکر و ممنونم .

از محبت ها و کمک های بیدریغ یگانه یار و همراه راستین زندگیم و خانواده ایشان ممنون و سپاسگزارم .

« خدایا کمکم کن که فقط برای رضای تو گام بردارم و همیشه پیروی پیامبر (ص) و ائمه اطهار (ع) باشم . »

علیرضا قدسی

عنوان صفحه

فصل سوم : رگرسیون چند متغیره خطی ۴۱

۳-۱ مقدمه ۴۲

۳-۲ فرضهای آماری ۴۳

۳-۳ برآورد پارامترها ۴۴

۳-۴ آماره آزمون ۴۵

۳-۵ تابع توان آزمون ۴۶

۳-۶ بررسی چند مثال ۴۶

۳-۷ چند جمله ایهای متعامد ۵۱

فصل چهارم :تحلیل واریانس بدون اثر متقابل ۵۴

۴-۱ مقدمه ۵۵

۴-۲ تحلیل واریانس تک عاملی ۵۶

۴-۳ تحلیل واریانس دو عاملی ۶۰

۴-۴ تحلیل واریانس چند عاملی ۶۳

۴-۵ تناظر بین تحلیل واریانس و رگرسیون ۶۶

عنوان صفحه

فصل پنجم: مقایسه توابع توان آزمونهای رگرسیون چند متغیره خطی و تحلیل واریانس بدون اثر متقابل ۸۰

۵-۱ مقدمه: بیان معیاری برای مقایسه و چگونگی استفاده از آن ۸۱

۵-۲ طرح یک فاکتوری ۸۴

۵-۳ طرح دو فاکتوری ۸۶

۵-۴ طرح سه فاکتوری ۸۷

جداول مربوط به بخش (۵-۲) ۹۰

جداول مربوط به بخش (۵-۳) ۹۴

جداول مربوط به بخش (۵-۴) ۱۱۰

ضمیمه ....

جدول ضرایب چند جمله ایهای متعامد ۱۲۸

جداول مربوط به محاسبه مقادیر تابع توان آزمونها در مدلهای خطی کلی ۱۲۹

منابع ۱۷۰

پیشگفتار

در تحلیل رگرسیونی ما قادریم که بین یک متغیر مورد نظر به عنوان متغیر وابسته (پاسخ) و یک یا چند متغیر مستقل تعیینی (پیش بین) ، رابطه ای را به تجربه معین کرده و از آن استفاده نمائیم . علاوه بر این می توانیم با استفاده از آزمون فرضهای خطی به این نتیجه برسیم که کدام یک از متغیرهای مستقل روی متغیر وابسته تأثیر دارند .

می دانیم که آنالیز واریانس ، معادل رگرسیون با متغیرهای ظاهری است . لذا در اینجا نیز با استفاده از فرضهای خطی می توان تأثیر متغیرهای مستقل (عوامل مورد بررسی) را روی متغیر پاسخ آزمون نمود . دراین حالت اگر عوامل مورد بررسی کمّی باشند ، مقادیر این عوامل در مدل شرکت ندارند .

اگر بتوان مقادیر این عوامل را در مدل وارد نمود با استفاده از مدل جدید می توان مقادیر متغیر پاسخ را به ازای مقادیر عاملها ، تخمین زد .

دراین پایان نامه ابتدا روش فرموله کردن مسائل آنالیز واریانس با عوامل کمّی به مدل رگرسیون با شرکت مقادیر عاملها گفته می شود . سپس معیاری معرفی می شود که به کمک آن بتوان توان آزمونها را در دو حالت رگرسیون چند متغیره خطی و آنالیز واریانس بدون اثر متقابل با یکدیگر مقایسه نموده تا با استفاده از آن معین کنیم که در چه وضعیتی بهتر است از چه روشی استفاده شود .

قسمت اول کار نظری است که با استفاده از روشها و محاسبات آماری لازم و مراجعه به مقاله ها و کتب مربوط انجام می پذیرد . و قسمت دوم عملی است که با تعیین

معیار بدست آمده و با استفاده از رایانه می توان شرایط مختلف را بررسی نمود .

لذا ابتدا در فصل اول تعاریف و قضایای مربوط به فرضهای درجهٔ دوم را بیان کرده ایم .

در فصل دوم مدلهای خطی کلی را معرفی و قضیه های لازم را آورده ایم . و سپس فرمول محاسبه مقدار توان آزمون فرضهای خطی کلی را بیان نموده ایم .

در فصل سوم رگرسیون چند متغیره خطی که حالت خاصی از مدلهای خطی کلی است را معرفی نموده ایم .

در فصل چهارم ، تحلیل واریانس بدون اثر متقابل را مورد بحث قرار داده و چگونگی استفاده از متغیرهای ظاهری برای تبدیل مدل تحلیل واریانس به مدل رگرسیونی را بیان کرده ایم . آنگاه روش فرموله نمودن این مدل در حالتی که عوامل کمی می باشند را به مدل رگرسیون با شرکت مقادیر عوامل مورد بررسی شرح داده ایم و تابع توان آزمون را در هر حالت معرفی کرده ایم .

و بالاخره در فصل آخر معیار گفته شده را معرفی و از آن برای اهداف گفته شده استفاده شده است . آنگاه با استفاده از نرم افزار *SAS* ، برنامه ها و جداولی ارائه داده ایم که با استفاده از آنها بتوان در عمل ، یکی از روشهای گفته شده را انتخاب نمود .

قدسی

۷۶/۱۰/۲۲

فصل اول

فصل اول (۱)

فرمهای درجهٔ دوم و توزیعهای غیر مرکزی χ^2 و F

۱-۱ تعریف :

هر تابع درجهٔ دوم همگنی از n متغیر $X_1, \ldots, X_n$ را یک فرم درجه دوم ازاین متغیر ها گوئیم. چنین فرمی به صورت زیر نوشته می شود که در آن α_{ij} ها ضرایب عددیند :

$$Q=\sum_{i=1}^{n}\sum_{i=1}^{n}\alpha_{ij}X_iX_j$$

و یا به صورت ماتریسی داریم :

$$Q=X'AX$$

که در آن $X=(X_1,\ldots,X_n)'$ یک بردار ستونی از متغیرها و

$$(i,j=1,2,\ldots,n)\quad A=\left(\frac{\alpha_{ij}+\alpha_{ji}}{2}\right)$$

و یک ماتریس متقارن می باشد که ماتریس فرم درجه دوم نام دارد .

۱-۲ انواع فرمهای درجهٔ دوم : (۲)

فرمهای درجه دوم نسبت به مقادیر ی که در ازای مقادیر مختلف بردار X پیدا می کنند ، به صورت زیر دسته بندی می شوند .

(الف) معین مثبت : اگر برای تمام Xهای غیر صفر ، $X'AX$ یک عدد مثبت باشد ، آنگاه فرم درجه دوم $X'AX$ را " معین مثبت " یا pd (۳) گویند .

---

۱ - در سراسر پایان نامه جاهای که منبع ذکرنشده است از [1] استفاده شده است .
وکل رساله بر مبنای [12] پایه ریزی شده است .

2 - [2]

3 - Positive definite

(ب) **معین منفی** : اگر $X'AX$ - معین مثبت باشد آنگاه فرم درجهٔ دوم را "معین منفی" یا nd (۱) می نامند .

(ج) **نیمه معین مثبت** : اگر فرم درجه دوم به ازای بعضی از مقادیر X، مثبت و حداقل برای یک X غیر صفر مساوی صفر باشد ، آن را "نیمه معین مثبت" یا psd (۲) گویند .

دراین حالت می نویسیم : $X'AX \geq 0$.

(د) **نیمه معین منفی** : اگر فرم درجه دوم به ازای بعضی از مقادیر X، منفی و حداقل برای یک X غیر صفر مساوی صفر باشد ، آن را " نیمه معین منفی " یا nsd (۳) می نامند . و

می نویسیم : $-X'AX \geq 0$

۱-۲-۱ نتیجه :

یک ماتریس را "معین مثبت" گوئیم هرگاه فرم درجه دوم آن معین مثبت باشد . و "نیمه معین مثبت" گوئیم اگر فرم درجه دوم آن نیمه معین مثبت باشد .

۱-۳ امید ریاضی و واریانس فرمهای درجهٔ دوم

۱-۳-۱ قضیه (۴) : اگر $X \sim N_n(\mu, V)$ آنگاه برای متغیر تصادفی $X'AX$ داریم :

(الف) $$E(X'AX) = tr(AV) + \mu' A \mu$$

(این رابطه هنگامی که توزیع ، نرمال هم نباشد برقرار است .)

(ب) کمولانت rام متغیر تصادفی $X'AX$ به صورت زیر است :

---

1 - negative definite

2 - Positive semidefinite

3 - negative semidefinite

۴ - اثبات قضیه در کتاب Searle (۱۹۷۱) صفحهٔ ۵۵ آمده است .

$$K_r(X'AX)=2^{r-1}\ (r-1)!\ [tr(AV)^r+r\mu'A(VA)^{r-1}\mu] \quad , \quad r\geq 1$$

(ج) $cov(X,X'AX)=2VA\mu$

۱-۳-۱-۱ **نتیجه** : اگر $\mu=0$ آنگاه $cov(X,X'AX)=0$.

۱-۳-۱-۲ **نتیجه** : $V(X'AX)=K_2(X'AX)=2tr(AV)^2+4\mu'AVA\mu$

۱-۴: استقلال فرمهای درجهٔ دوم [1]

۱-۴-۱ **قضیه** : هرگاه $Y\sim N_n(\mu,V)$ آنگاه فرم درجه دوم $Y'AY$ و فرم خطی BY که در آن B ماتریسی از مرتبه mxn است مستقل خواهند بود اگر و تنها اگر $BVA=0$.

۱-۴-۲ **قضیه** : هرگاه $Y\sim N_n(\mu,V)$ آنگاه فرمهای درجهٔ دوم $Y'AY$ و $Y'BY$ مستقل از یکدیگرند اگر و تنها اگر $BVA=0$ یا $AVB=0$

۱-۵ توزیعهای ناشی از فرمهای درجهٔ دوم

۱-۵-۱ توزیع χ^2 مرکزی و غیر مرکزی

(الف) توزیع χ^2 مرکزی

فرض کنید $X\sim N_n(0,I)$ ، آنگاه متغیر تصادفی $U=\sum_{i=1}^{n}X_i^2=X'X$ دارای توزیع χ^2 ی مرکزی با n درجهٔ آزادی است و می نویسیم $U\sim\chi_n^2$ تابع چگالی چنین متغیری به صورت زیر است .

$$f_U(u)=\frac{u^{\frac{n}{2}-1}e^{-\frac{u}{2}}}{\Gamma(\frac{n}{2})\ 2^{\frac{n}{2}}} \qquad ; \quad u>0$$

۱ - اثبات قضایا در کتاب Searle (۱۹۷۱) صفحهٔ ۵۹ آمده است .

۱-۵-۱-۱ **نتیجه**: با استفاده از قضیه ۱-۳-۱ به سادگی داریم

$$E(\chi_n^2) = n \quad , \quad Var(\chi_n^2) = 2n$$

(ب) توزیع χ^2 غیر مرکزی.

فرض کنید $X \sim N_n(\mu, I)$ آنگاه تابع چگالی متغیر تصادفی $U=X'X$ به شکل زیر می باشد:

$$f_U(u) = \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \cdot \frac{u^{\frac{n}{2}+j-1} e^{-\frac{u}{2}}}{\Gamma(\frac{n}{2}+j)\ 2^{\frac{n}{2}+j}} \quad ; \quad u>0 \qquad (۱-۵)$$

$$= \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \cdot \chi^2_{n+2j}$$

که بصورت مجموع وزنی تعداد نامتناهی از توابع چگالی χ^2 مرکزی با $n+2j$ درجهٔ آزادی است. وزنها تابع چگالی پواسن با پارامتر λ می باشند که در آن $\lambda = \frac{1}{2}\mu'\mu$ است.

این توزیع را توزیع "χ^2 غیر مرکزی" با n درجه آزادی و پارامتر غیر مرکزی $\lambda = \frac{1}{2}\mu'\mu$ می نامیم و می نویسیم:

$$U \sim \chi^{2'}_{n\,,\,\lambda}$$

برای اثبات رابطه (۱) به صورت زیر عمل می کنیم: [۱]

تبدیل متعامد $Y=CX$ را در نظر بگیرید و فرض کنید $X \sim N_n(\eta, I)$.

فرض کنید $E(y)=C\mu=\eta$. با توجه به اینکه $C'C=I$ آنگاه $Y \sim N_n(\eta, I)$.

C را می توان طوری انتخاب کرد که $\lambda = \frac{1}{2}\mu'\mu$ و $\eta' = (0,0,\ldots,0,\sqrt{2\lambda})$

حال فرضهای درجه دوم $Y'AY$ و $Y'BY$ را در نظر بگیرید که A و B ماتریسهای مربع از

۱- منبع [9] .

مرتبه nxn و به صورت زیر تعریف شده باشند.

$$A = (O \ \vdots \ b) \quad ; \quad O=(0)_{nx(m-1)} \ , \ b'_{1xn}=(0,0,....,1)$$

$$B = \left(\begin{array}{c:c} I_{n-1} & 0 \\ \hline 0 & 0 \end{array} \right) \qquad ; \quad 0_{1x(n-1)}$$

اما چون: $AB=0$ بنا به قضیه ۱-۴-۲، $Y'AY$ و $Y'BY$ مستقل از یکدیگرند.

توجه کنید که
$$Y'BY = \sum_{j=1}^{n-1} Y_j^2 \quad , \quad Y'AY = Y_n^2$$

می دانیم که
$$W = \sum_{j=1}^{n-1} Y_j^2 \sim \chi_{n-1}^2$$

ابتدا توزیع $V=Y_n^2$ را بدست می آوریم.

پس
$$Y_n \sim N(\sqrt{2\lambda} \ , \ 1)$$

$$f_{Y_n}(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\sqrt{2\lambda})^2}{2}}$$

اما داریم
$$Y_n = \pm\sqrt{V} \quad \Rightarrow \quad \frac{dy_n}{dv} = \pm\frac{1}{2\sqrt{v}}$$

و لذا

$$f_V(v) = \frac{1}{\sqrt{2\pi} \ . \ 2\sqrt{v}} \left\{ e^{\frac{-(\sqrt{v}-\sqrt{2\lambda})^2}{2}} + e^{-\frac{(-\sqrt{v}-\sqrt{2\lambda})^2}{2}} \right\}$$

$$= \frac{v^{-\frac{1}{2}}}{2\sqrt{2\pi}} \left\{ e^{-\frac{1}{2}(\sqrt{v}-\sqrt{2\lambda})^2} + e^{-\frac{1}{2}(\sqrt{v}+\sqrt{2\lambda})^2} \right\}$$

$$= \frac{v^{-\frac{1}{2}}}{2^{\frac{3}{2}}\Gamma(\frac{1}{2})} e^{-\frac{1}{2}(v+2\lambda)} . \left\{ e^{\sqrt{2v\lambda}} + e^{-\sqrt{2v\lambda}} \right\}$$