به نام خدا

**تکلیف شماره 5 و 6 محاسبات آماری (دانشگاه صنعتی اصفهان) مهلت تحویل: جمعه 21 دی 97 ساعت 24**

**آزمون صفر بودن زیرمجموعه ای از ظرایب رگرسیونی:**

فرض کنید مدل شامل X1 تا X4 است و علاقمندیم به آزمون صفر بودن همزمان ضرایب متغیرهای X2 و X4:

H0: 2=4=0 H1: o.w.

برای اینکار دو مدل زیر را لازم است برازش دهیم

Yi = 0 + 1 Xi1 + 2 Xi2 + 3 Xi3 + 4 Xi4 + i (Full model)

Yi = 0 + 1 Xi1 + 3 Xi3 + i (Reduced model)

و سپس به صورت زیر عمل کنیم:

F = [(SSEr – SSEf )/r] / MSEf تعداد ضرایب رگرسیونی مدل کامل p: تفاضل پارامترهای دو مدل r:

If F>F(1-)(r,n-p) then reject H0

کد مربوط به این آزمون

> full = lm(Y ~ X1 + X2 + X3 + X4) # Full model

> reduced = lm(Y ~ X1 + X3) # Reduced model

> anova(reduced, full) # Compare the models

سپس، آماره F و پی – مقدار مربوط به این دستور مدنظر قرار می گیرند.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

داده ھای auto.txt را درنظر بگیرید.

1. الف) با فرض نرمال بودن توزيع قيمت خودرو به ازای ھريک از سطوح متغير استایل بدنه ماشین (Style)، آزمون کنید که آیا تغییر در استایل بدنه ماشین بر متوسط قیمت خودرو تاثیرگذار است؟
2. با فرض نرمال بودن توزيع قيمت خودرو به ازای ھريک از سطوح متغير استایل بدنه ماشین (Style)، با توجه به این مساله که می‌دانیم ممکن است متغیرهای height و horsepower بر قیمت تاثیرگذار باشند، آزمون کنید که آیا تغییر در استایل بدنه ماشین بر متوسط قیمت خودرو تاثیرگذار است؟
3. اگر تاثیرگذاری متغیرهای height و horsepower بر قیمت خودرو، به استایل خودرو بستگی داشته باشد، آزمون قبل را چگونه انجام می‌دهید.

\* در هر حالت، آزمون فرض و معادله مدل مورد بررسی را بنویسید.

2- الف) مدل های رگرسیونی میان متغیرهای horsepower و قیمت خودرو به تفکیک برای ماشین های دو در و چهار در برازش دهید.

ب) آزمون کنید که آیا عرض از مبدا دو خط ارتباط دهنده متغیرهای توضیحی و پاسخ در جمعیت‌های ماشین‌های دو در و چهار در برابر هستند؟

پ) آزمون کنید که آیا شیب این دو خط برابر هستند؟

ت) آزمون کنید که آیا این دو خط بر هم منطبق هستند؟

\*راهنمایی: برای حل سوال 2، لازم است متغیر نشانگر تعداد در و در برخی قسمت های سوال، اثر متقابل آن با horsepower درنظر گرفته شود.

1. متغیرهای horepower ، length ،highway\_mpg ،city\_mpg ،height ،fueltype را به عنوان متغيرھای توضيحی و قيمت خودرو را به عنوان متغير پاسخ درنظربگیرید. به سوالات زير پاسخ دھيد:

**انتخاب متغير:**

 الف- ابتدا مدلی با تمام متغيرھا برازش دھيد. تفسير عرض از مبدا و ضرایب fueltype و horsepower را بیان کنید.

ب- آزمون معناداری کلی مدل را انجام دھيد و آن را معنا کنيد.

پ- آزمون معناداری ضرايب رگرسيون جزيی را انجام دھيد و آن را معنا کنيد.

ت- توضيح دھيد چرا بايد تا جايی که امکان دارد، متغيرھايی که حضورشان در مدل معنادار نيست را حذف کرد؟

ث- ھمه متغيرھا را وارد مدل کنيد. آزمون کنيد که آيا می توان ھمزمان متغيرھای length و height را از مدل حذف کرد؟

ج- با استفاده از روش stepwise مدل مناسبی را برازش دهید. مراحل اجرای اين الکوريتم را بر اساس خروجی ھا توضيح دھيد.

**پيش بینی: (مدل قسمت ج را درنظر بگيرید)**

چ- پيش بينی نقطه ای و فاصله ای برای قيمت ماشينی با خصوصيات زير ارايه دھيد:

fueltype=gas, height=50, city\_mpg=22, highway\_mpg=28, length=200, horsepower=170

ح- برآورد نقطه ای و فاصله ای برای تخمين متوسط قيمت ماشين ھايی با الگوی قسمت چ برای متغيرھای توضيحی، ارايه دھيد.

**بررسی فرضيات مدل: (مدل قسمت ج را درنظر بگيرید)**

خ- آيا فرض نرمال بودن توزيع برقرار است؟ توضيح دھيد که اگر اين فرض برقرار نباشد، چه مشکلی پيش خواھد آمد.

همچنین، تابعی بنویسید که با اجرای آن، تمامی آماره ها و نمودارهای مربوط به بررسی فرض نرمال بودن توزیع خطا انجام شود. (مشابه تمرین کلاسی)

د-بررسی کنيد که آيا فرض ثبات واريانس برقرار است؟ توضيح دھيد که اگر به مساله عدم ثبات واريانس توجه نکنيم، چه مشکلی پيش می آيد؟

ذ- وزن مناسب را با توجه به قسمت د به دست آوريد و رگرسيون وزنی بزنيد با رسم نمودار مناسب بررسی کنيد که آيا مشکل عدم ثبات واريانس با وزن انتخابی شما برطرف شده است؟

**بررسی تاثيرگذاری مشاھدات: (مدل قسمت ج را درنظر بگيرید)**

ر- داده ھای دورافتاده، اھرم گون و تاثيرگذار را مشخص کنيد. توضيح دھيد ھر يک چه تاثيری بر مدل می گذارند.

ز- با استفاده از رسم شکل، نقاط اھرم گون، دورافتاده و برخی از نقاط تاثيرگذار را مشخص کنيد.

**بررسی مشکل ھمخطی چندگانه: (مدل قسمت ج را درنظر بگيرید)**

ژ- بررسی کنيد که آيا مشکل ھمخطی چندگانه وجود دارد يا خير؟ توضيح دھيد ھم خطی چندگانه چه مشکلی را برای مدل ايجاد

می کند.

موفق باشید.