

به نام خدا

مهلت تحویل: روز قبل از امتحان پایان ترم

تکلیف شماره ۴ محاسبات آماری (دانشگاه صنعتی اصفهان)

آزمون صفر بودن زیرمجموعه ای از ضرایب رگرسیونی:

فرض کنید مدل شامل X_1 تا X_4 است و علاقمندیم به آزمون صفر بودن همزمان ضرایب متغیرهای X_2 و X_4 :

$$H_0: \beta_2 = \beta_4 = 0 \quad H_1: \text{o.w.}$$

برای اینکار دو مدل زیر را لازم است برآزش دهیم

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i \quad (\text{Full model})$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_3 X_{i3} + \varepsilon_i \quad (\text{Reduced model})$$

و سپس به صورت زیر عمل کنیم:

$$F = [(SSE_r - SSE_f) / r] / MSE_f$$

r: تعداد ضرایب رگرسیونی مدل کامل p: تفاوت پارامترهای دو مدل

If $F > F_{(1-\alpha)}(r, n-p)$ then reject H_0

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.

۱- آزمون استقلال در تکلیف قبل را انجام دهید.

۲- داده های auto.txt را در نظر بگیرید. متغیرهای 'height', 'highway_mpg', 'length', 'horsepower' و 'fueltype' را به عنوان متغیرهای توضیحی و قیمت خودرو را به عنوان متغیر پاسخ در نظر بگیرید. به سوالات زیر پاسخ دهید:

انتخاب متغیر:

الف - ابتدا مدلی با تمام متغیرها برآزش دهید. تفسیر عرض از مبدا و ضرایب 'fueltype' و 'horsepower' را بیان کنید.

ب - آزمون معناداری کلی مدل را انجام دهید و آن را معنا کنید.

پ - آزمون معناداری ضرایب رگرسیون جزئی را انجام دهید و آن را معنا کنید.

ت - توضیح دهید چرا باید تا جایی که امکان دارد، متغیرهایی که حضورشان در مدل معنادار نیست را حذف کرد؟

ث - همه متغیرها را وارد مدل کنید. آزمون کنید که آیا می توان همزمان متغیرهای 'length' و 'height' را از مدل حذف کرد؟

ج - با استفاده از روش stepwise مدل مناسبی را برآزش دهید. مراحل اجرای این الگوریتم را بر اساس خروجی ها توضیح دهید.

پیش بینی: (مدل قسمت ج را در نظر بگیرید)

چ - پیش بینی نقطه ای و فاصله ای برای قیمت ماشینی با خصوصیات زیر ارائه دهید:

$$\text{fueltype}=\text{gas}, \text{height}=100, \text{highway_mpg}=24, \text{length}=250, \text{horsepower}=170$$

ح - برآورد نقطه ای و فاصله ای برای تخمین متوسط قیمت ماشین هایی با الگوی قسمت چ برای متغیرهای توضیحی، ارائه دهید.

بررسی فرضیات مدل: (مدل قسمت ج را در نظر بگیرید)

خ - آیا فرض نرمال بودن توزیع برقرار است؟ توضیح دهید که اگر این فرض برقرار نباشد، چه مشکلی پیش خواهد آمد.

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

د- بررسی کنید که آیا فرض ثبات واریانس برقرار است؟ توضیح دهید که اگر به مساله عدم ثبات واریانس توجه نکنیم، چه مشکلی پیش می آید؟

ذ-وزن مناسب را با توجه به قسمت د به دست آورید و رگرسیون وزنی بزنیید با رسم نمودار مناسب بررسی کنید که آیا مشکل عدم ثبات واریانس با وزن انتخابی شما برطرف شده است؟

بررسی تاثیرگذاری مشاهدات: (مدل قسمت ج را در نظر بگیرید)

ر-داده های دورافتاده، اهرم گون و تاثیرگذار را مشخص کنید. توضیح دهید هر یک چه تاثیری بر مدل می گذارند. ز-با استفاده از رسم شکل، نقاط اهرم گون، دورافتاده و برخی از نقاط تاثیرگذار را مشخص کنید.

بررسی مشکل همخطی چندگانه: (مدل قسمت ج را در نظر بگیرید)

ژ-بررسی کنید که آیا مشکل همخطی چندگانه وجود دارد یا خیر؟ توضیح دهید هم خطی چندگانه چه مشکلی را برای مدل ایجاد می کند.

موفق باشید.