دانشکده‌ی علوم ریاضی

آمار و کاربردها		۲۶ دی ۱۳۹۸
	تمرین : سری ۳	
مدرّس: دکتر محسن شریفی تبار		مهلت تحویل ۱۰ بهمن

علیرضا درویشی.. ۹۶۱۰۹۶۷۴

مسأله ۱

الف)

$$Y_i = \alpha + \beta X_i = \alpha + \beta(x_i + \bar{X}) = (\alpha + \beta\bar{X}) + \beta x_i = \alpha' + \beta x_i$$
$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(y_i + \bar{Y} - \bar{Y})}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2} - \frac{\sum x_i \bar{Y}}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2}$$
$$w_i = \frac{x_i}{\sum x_i^2} \Rightarrow b = \sum w_i Y_i$$
$$\Rightarrow E[b] = \sum w_i E[Y_i] = \sum w_i(\alpha' + \beta x_i) = \sum w_i\alpha' + \beta w_i x_i = \alpha' \sum w_i + \beta \sum w_i x_i =$$
$$\beta \frac{\sum x_i x_i}{x_i^2} = \beta \frac{\sum x_i^2}{\sum x_i^2} = \beta$$
$$a = \bar{Y} - b\bar{X}$$
$$\Rightarrow E[a] = E[\bar{Y}] - E[b]\bar{X} = \frac{1}{n}\sum E[Y_i] - \beta\bar{X} = \frac{1}{n}\sum(\alpha' + \beta x_i) - \beta\bar{X} = \alpha' - \beta\bar{X} = \alpha$$

ب)

$$Cov(a,b) = E[ab] - E[a]E[b] = E[(\bar{Y} - b\bar{X})b] - E[\bar{Y} - b\bar{X}]E[b] =$$
$$E[b\bar{Y}] - \bar{X}E[b^2] - E[b]E[\bar{Y}] + \bar{X}E^2[b] = E[b\bar{Y}] - E[b]E[\bar{Y}] - \bar{X}Var[b]$$
$$Var[b] = Var[\sum w_i Y_i] = \sum w_i^2\sigma^2 = \sigma^2 \sum \frac{x_i^2}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2}$$
$$E[Y_j b] = E[Y_j \sum w_i Y_i] = E[\sum_{i\neq j} w_i Y_i Y_j] + E[w_j Y_j^2] = E[Y_j]\sum_{i\neq j} w_i Y_i + w_j E[Y_j^2] =$$
$$E[Y_j]\sum_{i\neq j} w_i Y_i + w_j E[Y_j^2] - w_j E^2[Y_j] + w_j E^2[Y_j] = E[Y_j]\sum w_i E[Y_i] + w_j\sigma^2 =$$
$$\beta E[Y_j] + w_j\sigma^2 \Rightarrow E[\bar{Y}b] = \frac{1}{n}\sum E[Y_j b] = \bar{Y}\beta$$
$$\Rightarrow Cov(a,b) = \bar{Y}\beta - \bar{Y}\beta - \bar{X}\frac{\sigma^2}{\sum x_i^2} = -\frac{\sigma^2\bar{X}}{\sum x_i^2}$$

مسأله ۲

الف)

$$y'_{ij} = \beta_0 + \beta_1 x_i + e'_i \ : jth\ data\ in\ x_i$$
$$y_i = \frac{\sum_j y'_{ij}}{n_i} \Rightarrow E[y_i] = \beta_0 + \beta_1 x_i\ ,\ Var[y_i] = \frac{1}{n_i^2}\sum Var[y'] = \frac{1}{n_i}\sigma^2 = \rho_i^2\sigma^2 \Rightarrow n_i = \frac{1}{\rho_i^2}$$

ب) شرط واریانس ثابت و میانگین صفرِ انحراف از خط با توجه به قسمت قبلی برقرار است شرط خطی بودن به وضوح از صورت معادله برقرار است شرط استقلال انحراف از خط هم با توجه به قسمت قبلی برقرار است.

ج)

$$s=\sum(z_i-b_0u_i-b_1v_i)^2$$
$$\Rightarrow \frac{ds}{db_0}=0\Rightarrow \sum u_i(z_i-b_0u_i-b_1v_i)=0$$
$$\frac{ds}{db_1}=0\Rightarrow \sum v_i(z_i-b_0u_i-b_1v_i)=0$$
$$\Rightarrow b_0=\frac{\sum v_i^2\sum z_iu_i-\sum z_iv_i\sum u_iv_i}{(\sum u_i^2)(\sum v_i^2)-(\sum u_iv_i)^2}, b_1=\frac{\sum u_i^2\sum z_iv_i-\sum z_iu_i\sum u_iv_i}{(\sum u_i^2)(\sum v_i^2)-(\sum u_iv_i)^2}$$

د)

$$s=\sum(z_i-b_0u_i-b_1v_i)^2=\sum(\rho_i^{-1}y_i-b_0\rho_i^{-1}-b_1\rho_i^{-1}x_i)^2=\sum(y_i-b_0-b_1x_i)^2\rho_i^{-2}$$

ه)

$$b_0=\sum w_iz_i, w_i=\frac{u_i\sum v_j^2-v_i\sum u_jv_j}{(\sum u_j^2)(\sum v_j^2)-(\sum u_jv_j)^2}\Rightarrow Var(b_0)=\sum w_i^2Var(z_i)=\sigma^2\sum w_i^2$$
$$\sum w_i^2=\sum\frac{u_i^2(\sum v_j^2)^2+v_i^2(\sum u_jv_j)^2-2v_iu_i(\sum v_j^2)(\sum u_jv_j)}{((\sum u_j^2)(\sum v_j^2)-(\sum u_jv_j)^2)^2}=\frac{\sum v_j^2((\sum u_j^2)(\sum v_j^2)-(\sum u_jv_j)^2)}{((\sum u_j^2)(\sum v_j^2)-(\sum u_jv_j)^2)^2}=$$
$$\frac{\sum v_j^2}{(\sum u_j^2)(\sum v_j^2)-(\sum u_jv_j)^2}$$
$$\Rightarrow Var(b_0)=\frac{\sigma^2\sum v_j^2}{(\sum u_j^2)(\sum v_j^2)-(\sum u_jv_j)^2}$$

به طریق مشابه:

$$Var(b_1)=\frac{\sigma^2\sum u_j^2}{(\sum u_j^2)(\sum v_j^2)-(\sum u_jv_j)^2}$$

مسأله ۳

از توزیع F استفاده می کنیم.
فرض صفر این است که مدل اول از مدل دوم بهتر فیت نشده است.

$$F=\frac{\frac{S_1-S_2}{(n-k_1)-(n-k_2)}}{\frac{S_2}{n-k_2}}=\frac{\frac{795-783}{(200-3)-(200-5)}}{\frac{783}{200-5}}=1.4$$
$$F_{\alpha=0.05}(2,195)=3.04223>1.4\Rightarrow \text{فرض صفر را رد نمیکنیم}$$

پس مدل اول بهتر از مدل دوم فیت نشده یا به عبارتی دلایل آماری کافی برای رد این فرض که مدل اول بهتر از مدل دوم فیت نشده وجود ندارد.

مسأله ۴

الف)

امکان دارد که چند جفت از این ۵ متغیر رابطه ای خطی با هم دیگر داشته باشند طوری که درواقع اگر رگرسیون چند گانه ای برای این متغیر ها بررسی می شد، یکی از این متغیرها تاثیر مستقیم نداشته باشند.

یا برای مثال یکی از متغیرهایی که تاثیرش رد شد، به دو متغیر دیکر بستگی داشته باشد و تاثیر دو متغیر دیگر در این متغیر نتیجه داده باشد که شیب رگرسیون ساده برابر صفر است ولی اگر رگرسیون چندگانه استفاده می شد، شیب صفر نمی شد.

ب)

بهتر است رابطه ی رگرسیون چندگانه ای با ۲۰ متغیر مورد بررسی فیت کند و بررسی کند که شیب کدام یک از متغیرها برابر صفر می شود. همچنین رابطه ی علت معلولی بین ۱۰۰ متغیر بررسی شود و متغیر های کاملا یکسان از مدل حذف شوند.

مسأله ۵

الف)

$$b=\frac{\sum x_i y_i}{\sum x_i^2},\ SE=\frac{s}{\sqrt{\sum x_i^2}} \Rightarrow b=800, SE=\frac{7300}{30}=243.3,\ t_{0.025,df=48}=2.01$$
$$\beta=b\pm t_{0.025}SE=800\pm 489.1$$

ب)

بله چون صفر در بازه ی اطمینان شیب نیست و به عبارتی دیگر $p-vlaue$ برای شیب از 0.05 کمتر است.

ج)

$$Y_0=(a+bX_0)\pm t_{0.025}s\sqrt{1+\frac{1}{n}+\frac{(X-X_0)^2}{\sum x^2}}$$
$$Y_0=2000\pm 2.01\times 7300\times\sqrt{1+\frac{1}{50}+\frac{1}{900}}=2000\pm 14827$$

د)

بله. البته در صورتی که فرض کنیم تمام پارامتر های دیگر که امکان دارد تاثیر بگذارند(برای مثال رنگ پوست یا شغل پدر یا تعداد فرزند یا ..) را ثابت فرض کنیم و فقط تاثیر متغیر مورد بررسی را درنظر بگیریم.