

به نام یزدان پاک

آموزش هدفمند درس به درس سرفصل های مصوب

دگرسیون

تهیه کننده

سارا قنبری

آموزش هدفمند درس به درس سرفصل‌های رگسیون

تهیه‌کننده : سارا قنبری
ناشر : پردازش
چاپ اول : ۱۴۰۳
تعداد : ۱۰۰
حروفچینی : پردازش
چاپ و صحافی: پردازش
قیمت : ۳۰۰۰۰ تومان

حق چاپ محفوظ و متعلق به نشر پردازش است.

عنوان و پدیدآور : آموزش هدفمند درس به درس سرفصل‌های رگسیون
تهیه‌کننده: سارا قنبری
مشخصات نشر : تهران: پردازش، ۱۴۰۳
مشخصات ظاهری : ۵۴ ص.
شابک : ۹۷۸-۶۰۰-۲۳۳-۷۳۲-۰
وضعیت فهرست نویسی:

نشر پردازش

تهران: میدان انقلاب، خیابان فروردین، خیابان شهدای ژاندارمری، روبه‌روی اداره پست، پلاک ۱۳۲ و ۱۳۰ ساختمان پردازش

تلفن : ۶۶۹۷۵۵۶۶ (۱۰۰ خط) دورنگار: ۶۶۴۰۷۴۰۸

www.Pardazesh.org

سایت برتر کشور در چهارمین و پنجمین همایش ملی تجارت الکترونیک

فهرست مطالب

فصل اول «رگرسیون خطی ساده».....	۷
فصل دوم «استنباط در ارتباط با مدل رگرسیون خطی ساده».....	۱۶
فصل سوم «ضریب همبستگی».....	۱۸
فصل چهارم «مدل رگرسیون خطی چندگانه».....	۲۱
فصل پنجم «استنباط در ارتباط با مدل رگرسیونی چندگانه».....	۲۶
فصل ششم «ضریب همبستگی».....	۲۹
«پاسخنامه سوال های طبقه بندی شده رگرسیون».....	۳۱
پاسخنامه فصل دوم «استنباط در ارتباط با مدل رگرسیون خطی ساده».....	۳۸
پاسخنامه فصل سوم «ضریب همبستگی».....	۴۰
پاسخنامه فصل چهارم «مدل رگرسیون خطی چندگانه».....	۴۴
پاسخنامه فصل پنجم «استنباط در ارتباط با مدل رگرسیونی چندگانه».....	۵۰
پاسخنامه فصل ششم «ضریب همبستگی».....	۵۳

فهرست آخرین سرفصل‌های مصوب رگرسیون

رگرسیون براساس آخرین سرفصل‌های مصوب وزارت علوم:

سرفصل درس:

۱- رگرسیون خطی ساده

۱-۱- توصیف مدل

۲-۱- فرضیه‌های مدل

۳-۱- روش‌های برآورد پارامترهای مدل

۴-۱- خواص برآوردگرهای حداقل مربعات

۲- استنباط در ارتباط با مدل رگرسیون خطی ساده

۱-۲- استنباط آماری در ارتباط با شیب خط رگرسیونی

۲-۲- استنباط آماری در ارتباط با عرض از مبدأ خط رگرسیونی

۳-۲- جدول آنالیز واریانس برای مدل رگرسیونی خطی ساده

۳- ضریب همبستگی خطی

۳-۱- ضریب همبستگی به عنوان اندازه ارتباط

۳-۲- آزمون فرض

۴- مدل رگرسیون خطی چندگانه

۴-۱- توصیف مدل

۴-۲- فرض‌های مدل

۴-۳- روش‌های برآورد پارامترهای مدل

۵- استنباط در ارتباط با مدل رگرسیونی چندگانه

۵-۱- فاصله اطمینان

۵-۲- استنباط در ارتباط با خط رگرسیون

۵-۳- جدول آنالیز واریانس

۵-۴- آزمون فر در مدل رگرسیون چندگانه

۶- ضریب همبستگی

۶-۱- ماتریس همبستگی

۶-۲- ضریب همبستگی چندگانه

۶-۳- ضریب همبستگی جزئی

۶-۴- آزمون معنی داری ضریب همبستگی جزئی

فصل اول: رگرسیون خطی ساده

۱ - ۱ توصیف مدل

۱. اگر $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ که در آن $E(\varepsilon_i) = 0$ و $\text{Var}(\varepsilon_i) = \sigma^2 f(x_i)$ ، به طوری که f تابعی غیر همانی از x_i باشد، برای تبدیل این مدل به یک مدل وزنی با واریانس ثابت وزن مناسب و مدل تبدیل یافته کدام است؟ (۸۸ تمار)

(۱) $\frac{1}{f(x_i)}$ ، رگرسیون چندگانه بدون عرض از مبدا است. (۲) $\frac{1}{\sqrt{f(x_i)}}$ ، رگرسیون چندگانه بدون عرض از مبدا است.

(۳) $\frac{1}{\sqrt{f(x_i)}}$ ، رگرسیون خطی ساده با عرض از مبدا غیر صفر است. (۴) $\frac{1}{\sqrt{f(x_i)}}$ ، رگرسیون خطی ساده بدون عرض از مبدا است.

۲. تغییر واحد اندازه گیری متغیر مستقل در کدام یک از موارد زیر منتج به تغییری نمی شود؟ (۹۱ تمار)

(۱) برآورد عرض از مبدا (۲) ضریب تعیین (۳) برآورد ضریب زاویه (۴) مجموع مربعات رگرسیونی

۳. اگر در مدل رگرسیونی یک متغیر مستقل کیفی دارای ۵ رسته باشد، تعداد متغیرهای کمکی مورد نیاز برای نشان دادن این رسته ها کدام است؟ (۹۱ تمار)

(۱) ۱ (۲) ۴ (۳) ۵ (۴) ۶

۴. ضریب رگرسیون β_1 را در مدل $y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$ و ضریب رگرسیون مدل β_1^* را در $y_i = \beta_0^* + \beta_1^* x_{i1} + \beta_2^* x_{i2} + \varepsilon_i$ برآورد می کنیم، اگر $\hat{\beta}_1 = \hat{\beta}_1^*$ ، گزینه صحیح کدام است؟ (۹۳ تمار)

(۱) $\sum x_{i1} = 0$ ، $\sum x_{i1} x_{i2} = 0$ (۲) $\sum x_{i2} = 0$ ، $\sum x_{i1} x_{i2} = 0$ (۳) $\sum x_{i1} = 0$ ، $\sum x_{i1} x_{i2} = 0$ (۴) $\sum x_{i1} = \sum x_{i2} = 0$ ، $\sum x_{i1} x_{i2} = 0$

۵. فرض کنید مشاهدات Y_1, Y_2, Y_3, Y_4 به صورت زیر داده شده است، که در آن β_1, β_2 پارامترهای مدل و مستقل از هم هستند، برآورد پارامترهای (β_1, β_2) به روش کمترین توان دوم خطا کدام است؟ (۹۳ تمار)

$y_1 = \beta_1 + \beta_2 + \varepsilon_1$
 $y_2 = \beta_1 + \varepsilon_2$
 $y_3 = \beta_2 + \varepsilon_3$
 $y_4 = \beta_1 - \beta_2 + \varepsilon_4$

(۱) $\left(\frac{y_1 + y_2 + y_4}{3}, \frac{y_1 - y_2 - y_4}{3} \right)$ (۲) $\left(\frac{2y_1 + y_2 + y_4}{4}, \frac{2y_1 - y_2 - y_4}{4} \right)$
(۳) $\left(\frac{y_1 + y_2 + y_3 + y_4}{4}, \frac{y_1 + y_2 - y_3 - y_4}{4} \right)$ (۴) $\left(\frac{y_1 + y_2 + y_4}{4}, \frac{y_1 + y_3 + y_4}{3} \right)$

۶. اگر به جای مدل رگرسیون خطی واقعی $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ ، مدل $Y = \beta_0 + \beta_1 x_1 + \varepsilon$ برازش داده شده باشد و $\hat{\beta}_1$ برآوردگر پارامتر β_1 به روش کمترین توان های دوم خطا در مدل برازش داده شده باشد، آنگاه کدام گزینه درست است؟ (۹۳ تمار)

(۱) $\hat{\beta}_1$ برآوردگر نااریب β_1 است.

(۲) $\hat{\beta}_1$ برآوردگر نااریب β_1 می شود اگر x_1 و x_2 نا همبسته باشند.

(۳) واریانس $\hat{\beta}_1$ بیشتر از واریانس برآوردگر کمترین توان های دوم خطاهای β_1 براساس مدل واقعی است.

(۴) اگر x_1 و x_2 نا همبسته باشند، واریانس $\hat{\beta}_1$ بیشتر از واریانس برآوردگر کمترین توان های دوم خطای براساس مدل واقعی است.



۲-۱ فرضیه های مدل

(تأمل ۸۸)

۷. کدام یک از روابط زیر قابل تبدیل به یک مدل رگرسیون خطی نیست؟

$$Y_i = \beta_0 + \log(\beta_1 X_{i1}) + \beta_2 x_{i2} + \varepsilon_i \quad (۱)$$

$$Y_i = \frac{\alpha\beta}{\alpha \sin^2 x + \beta \cos^2 x + \varepsilon} \quad (۲)$$

$$Y_i = \beta_1 + (1 + \beta_1) e^{-\beta_2(x_i - 2)} + \varepsilon_i \quad (۴)$$

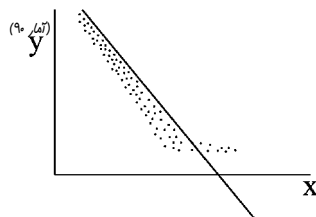
$$Y_i = \frac{1}{1 + e^{-\beta_1 - \beta_2 x_i + \varepsilon_i}} \quad (۳)$$

(تأمل ۹۰)

۸. اگر در مدل رگرسیونی، خطاها همگن نباشند، کدام گزینه صحیح است؟

- (۱) برآورد کمترین توان‌های دوم معمولی BLUE است.
 (۲) برآورد کمترین توان‌های دوم معمولی اریب است.
 (۳) برآورد کمترین توان‌های دوم معمولی نااریب اما خطی نیست.
 (۴) برآورد کمترین توان‌های دوم وزنی BLUE است.

۹. در یک مدل رگرسیونی، متغیر وابسته به صورت درصد می‌باشد و خط رگرسیونی مطابق شکل زیر به داده‌ها برازش داده شده است. شیب و عرض از مبدأ معنی‌دار و ضریب تعیین برابر ۰/۷۹ بوده است. یک آماردان با مشاهده نتایج اعلام نمود



که با این برازش موافق نمی‌باشد. کدام گزینه دلیل اصلی این نظر می‌باشد؟

- (۱) شیب خط برازش شده منفی است.
 (۲) ممکن است بتوان مدل مناسب‌تری برازش داد.
 (۳) برخی از مقادیر پیش‌بینی شده قابل قبول نمی‌باشد.
 (۴) ضریب تعیین به اندازه کافی بزرگ نمی‌باشد.

۱۰. در مدل رگرسیون خطی ساده $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ ، با فرض اینکه ε_i دارای توزیع نرمال باشد، مقدار

(تأمل ۹۲)

$P(Y_i > \beta_0 + \beta_1 x_i)$ کدام است؟

$$P(\hat{y}_i > \hat{\beta}_0 + \hat{\beta}_1 x_i) \quad (۲)$$

$$P(y_i > \hat{\beta}_0 + \hat{\beta}_1 x_i) \quad (۳)$$

$$۰/۹۷۵ \quad (۴)$$

۱۱. در مدل رگرسیون خطی ساده $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ ، که در آن $i = 1, \dots, n$ ، $\text{var}(\varepsilon_i) = \sigma^2$ ، $E(\varepsilon_i) = 0$ و $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ به‌ازای هر $i \neq j$ ، مانده‌های مدل (e_i) در کدام یک از شرایط زیر صدق می‌کنند؟

(تأمل ۹۷)

- (۱) واریانس ثابت دارند و همبسته‌اند.
 (۲) واریانس ثابت دارند و ناهمبسته‌اند.
 (۳) واریانس ثابت ندارند و همبسته‌اند.
 (۴) واریانس ثابت ندارند و ناهمبسته‌اند.

۳-۱ روش های برآورد پارامترهای مدل

۱۲. در مدل رگرسیونی $Y_i = \beta X_i + \varepsilon_i$ با فرض اینکه ε_i ها دارای توزیع گاما با پارامترهای $(1, 1)$ باشند، برآورد به روش

(تأمل ۸۸)

حداقل مربعات β کدام است؟

$$\frac{\sum X_i Y_i}{\sum X_i^2} \quad (۴)$$

$$\frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2} \quad (۳)$$

$$\text{Min}_{1 \leq i \leq n} \frac{Y_i}{X_i} \quad (۲)$$

$$\text{Min}_{1 \leq i \leq n} \frac{X_i}{Y_i} \quad (۱)$$



$$\frac{y_{x+2}}{y_x} = \frac{e^{1/5} e^{(x+2)/5}}{e^{x/5}} = e^1$$

(پاسخ آمار ۹۰)

۲۸. گزینه (۱)

$$\hat{B}_1 \sim N(\beta_1, \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}} \right)) \quad \bar{X} = 0$$

(پاسخ آمار ۹۰)

می‌دانیم که:

۲۹. گزینه (۱)

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{XX}} = \frac{\sigma^2}{55} \quad \text{Var}(\hat{\beta}_2) > \text{Var}(\hat{\beta}_1) \Rightarrow \text{SE}(\hat{\beta}_2) > \text{SE}(\hat{\beta}_1)$$

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{n} = \frac{\sigma^2}{10}$$

(پاسخ آمار ۹۰)

۳۰. گزینه (۱)

$$R^2 = \rho^2(x, y) = (0.8)^2 = 0.64$$

$$\Rightarrow R^2 = \frac{SSR}{SST} \Rightarrow SSR = 1/4 \Rightarrow SSE = 3/4 \Rightarrow \hat{\sigma}^2 = \text{MSE} = \frac{SSE}{n-2} = 0.75$$

(پاسخ آمار ۹۰)

۳۱. گزینه (۲)

$$\begin{aligned} \hat{\beta} &= \hat{\delta} & \hat{\alpha} &= \bar{y} - \hat{\beta} \bar{X} \\ & & &= \hat{\gamma} = \hat{\beta} \bar{X} \end{aligned}$$

(پاسخ آمار ۹۰)

می‌دانیم که:

(پاسخ آمار ۹۱)

۳۲. گزینه (۲)

۳۳. گزینه (۱)

$$X = a + by \Rightarrow \text{Cov}(X, Y) = 1$$

if $b > 0$

$$\text{Cov}(X, Y) = -1$$

if $b < 0$

$$\Rightarrow ab < 0$$

۳۴. گزینه (۳)

$$\text{Var}(\hat{\beta}_1 + \hat{\beta}_2) = \text{Var}(\bar{y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1) = \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2) + 2\text{Cov}(\hat{\beta}_1 + \hat{\beta}_2)$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}} \right) + \frac{\sigma^2}{S_{XX}} + 2 \left(\frac{\text{Cov}(\bar{y} - \hat{\beta}_1 \bar{X}) \hat{\beta}_1}{-2\bar{X} \text{Var}(\hat{\beta}_1)} \right) = \sigma^2 \left(\frac{1}{n} + \frac{(\bar{X} - 1)^2}{S_{XX}} \right)$$

(پاسخ آمار ۹۱)

۳۵. گزینه (۳)



با توجه به اینکه $\hat{b}_1$ یک برآوردگر حداقل مربعات است پس در نااریب بودن آن شکی نیست و با توجه به اینکه توزیع y_i ها پواسن است، پس دارای کمترین واریانس نمی‌باشد.

(پاسخ آمار ۹۲)

۳۶. گزینه ۲)

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \beta_1 \Rightarrow y'_i + \beta_1 x_i + \varepsilon_i \quad \hat{\beta}_1 = \frac{\sum x_i y'_i}{\sum x_i^2}$$

$$\begin{aligned} \sum x_i e_i &= \sum x_i (y'_i - \hat{y}'_i) = \sum x_i y'_i - \sum x_i (\hat{\beta}_1 x_i) = \sum x_i y'_i - \sum x_i^2 \frac{\sum x_i y'_i}{\sum x_i^2} \\ &= \sum x_i y'_i - \sum x_i y'_i = 0 \end{aligned}$$

۳۷. گزینه ۲)

$$\begin{aligned} \frac{(x_i - \bar{x})^2}{\sum x_i^2} = c_i &\Rightarrow \hat{\beta}_1 = \sum c_i y_i \quad E(\hat{\beta}_1) = \sum c_i E(y_i) \\ &= \sum c_i (\hat{\beta}_0 + \hat{\beta}_1 x_i) = \hat{\beta}_0 \sum c_i + \hat{\beta}_1 \sum c_i x_i \\ &= \frac{\sum (x_i - \bar{x})}{\sum x_i^2} \hat{\beta}_0 + \frac{\sum (x_i - \bar{x})^2}{\sum x_i^2} \hat{\beta}_1 \\ &= \hat{\beta}_1 \frac{\sum (x_i - \bar{x})^2}{\sum x_i^2} \end{aligned}$$

(پاسخ آمار ۹۲)

۳۸. گزینه ۲)

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{X} \quad \hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \\ \text{cov}(\hat{\beta}_0, \bar{y}) &= \text{cov}(\bar{y}, \bar{y}) - \bar{X} \text{cov}(\hat{\beta}_1, \bar{y}) = \text{cov}(\bar{y}, \bar{y}) - \bar{X} \text{cov}(\sum c_i y_i, \frac{1}{n} \sum y_j) \\ &= \frac{\sigma^2}{n} - \bar{X} \frac{1}{n} \sum c_i \text{cov}(y_i, y_i) = \frac{\sigma^2}{n} - \bar{X} \frac{\sigma^2}{n} \sum c_i = \frac{\sigma^2}{n} \end{aligned}$$

(پاسخ آمار ۹۲)

۳۹. گزینه ۴)

فرض صفر بودن میانگین باقیمانده‌ها، $E(e_i)$ و ناهمبسته بودن آن‌ها، $\text{cov}(e_i, e_j)$ ، باید در مدل رگرسیونی برقرار باشد.

(پاسخ آمار ۹۳)

۴۰. گزینه ۲)



$$\hat{\beta}_1 = \frac{\sum_{i=1}^{\Delta} x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^{\Delta} x_i^2 - n \bar{x}^2} = \frac{-2y_1 - y_2 + y_3 + 2y_4}{4 + 1 + 1 + 4} = -0.2y_1 - 0.1y_2 + 0.1y_3 + 0.2y_4$$

(پاسخ آمار ۹۳)

(پاسخ آمار ۹۵)

۴۱. گزینه ۲)

$$SSR = \hat{\beta}_1^2 S_{xx} \rightarrow E(SSR) = S_{xx} E(\hat{\beta}_1^2) = S_{xx} (\beta_1^2 + \frac{\sigma^2}{S_{xx}}) = S_{xx} \beta_1^2 + \sigma^2 = \beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 + b^2$$

(پاسخ آمار ۹۵)

۴۲. گزینه ۱)

$$\beta_1 \leq 3$$

(پاسخ آمار ۹۶)

۴۳. گزینه ۲)

$$\text{cov}(y_i, \hat{y}_i) = \text{var}(\hat{y}_i)$$

(پاسخ آمار ۹۶)

۴۴. گزینه ۳)

(پاسخ آمار ۹۶)

۴۵. گزینه ۱)

$$\text{var}(\hat{y}(\bar{x})) = \frac{\sigma^2}{n} \quad \text{می دانیم} \quad \text{var}(\hat{y}(x_0)) = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

$$\text{var}(\hat{y}(\bar{x} + a)) = \sigma^2 \left(\frac{1}{n} + \frac{a^2}{S_{xx}} \right) \Rightarrow \frac{\text{var}(\hat{y}(\bar{x} + a))}{\text{var}(\hat{y}(\bar{x}))} = \frac{\sigma^2 \left(\frac{1}{n} + \frac{a^2}{S_{xx}} \right)}{\frac{\sigma^2}{n}} = 1 + \frac{na^2}{S_{xx}}$$

(پاسخ آمار ۹۷)

۴۶. گزینه ۴)

$$\text{var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - (n-1) \sum_{i=1}^n x_i^2}$$

۴۷. گزینه ۴)

(پاسخ آمار ۹۷)

اگر $\beta_1 > 0$ باشد متوسط درآمد مردان از زنان هم‌سن آنها بیشتر است.