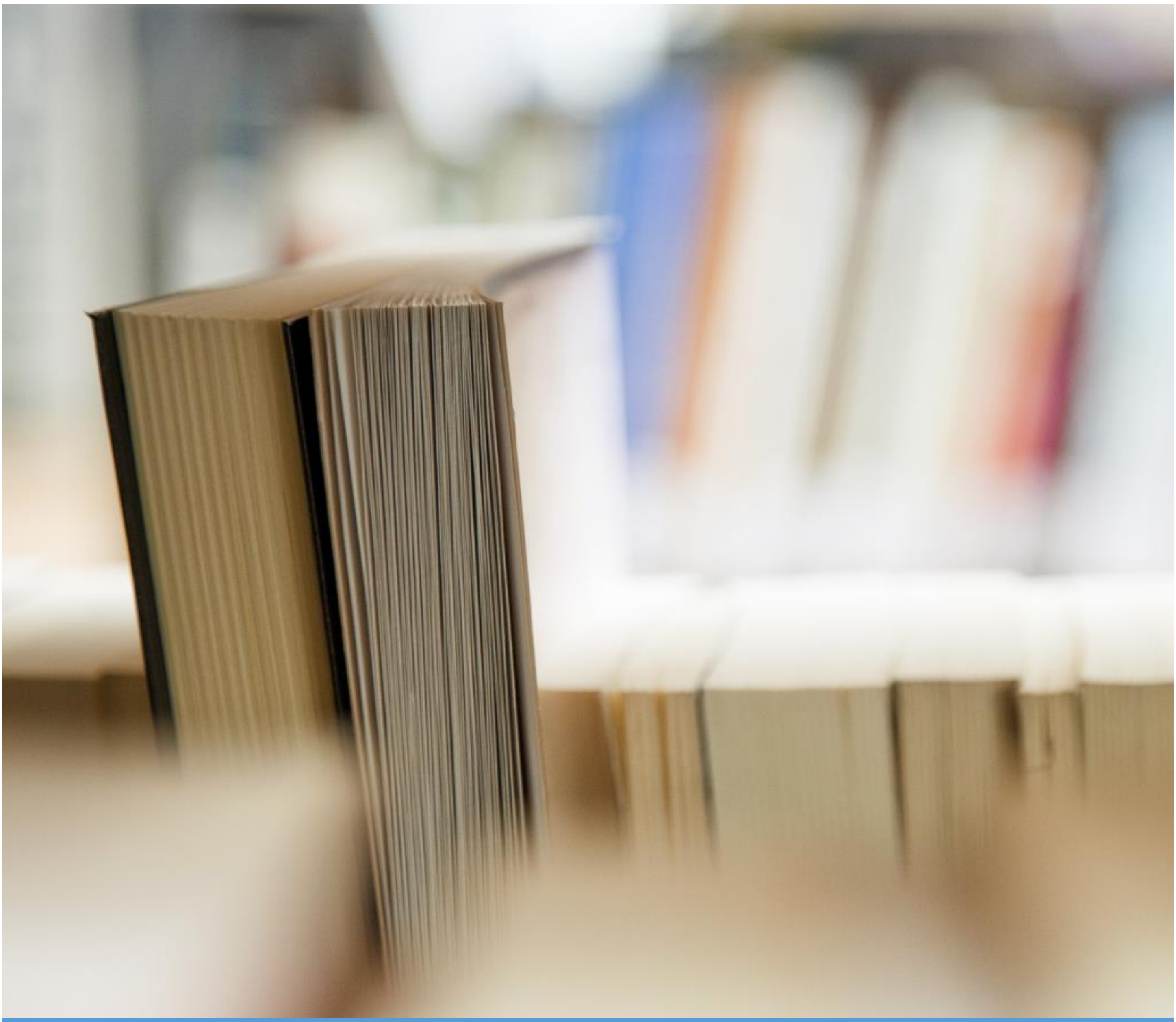


تجزیه و تحلیل مدل های کمی در تصمیم گیری های مدیریت

جلسه اول

ضرایب همبستگی

م سأله ۱) میزان سرمایه گذاری در بخش تحقیق و تو سعه و میزان سودآوری سالانه طی یک نمونه ۶ ساله در جدول زیر آمده است. در راستای پیش بینی سود ناشی از سرمایه گذاری در بخش فوق به کمک این نمونه مطلوب است:

① محاسبه ضریب همبستگی و تفسیر آن؛

② رسم نمودار پراکندگی و تفسیر نمودار پراکندگی؛

③ محاسبه معادله خط رگرسیون و رسم آن و پیش بینی سود سالانه برای سرمایه گذاری $x = 8$ و

④ آزمون معنی دار بودن ضریب همبستگی به دو روش.

سرمایه گذاری	سود سالانه	انحراف از میانگین سرمایه گذاری	انحراف از میانگین سود سالانه	$(x_i-\overline{x})(y_i-\overline{y})$	$(x_i-\overline{x})^2$	$(y_i-\overline{y})^2$
x	y	$x_i-\overline{x}$	$y_i-\overline{y}$			
5	31	5-5= 0	31 - 30 = 1	(0).(1) = 0	0	1
11	40	11-5= 6	40 - 30 = 10	(6).(10) = 60	36	100
4	30	4-5= -1	30 – 30 = 0	(-1).(0) = 0	1	0
5	34	5-5= 0	34 – 30 = 4	(0).(4) = 0	0	16
3	25	3-5 = -2	25 – 30 = -5	(-2).(-5) = 10	4	25
2	20	2-5 = -3	20 – 30 = -10	(-3).(-10) = 30	9	100
$\sum x_i = 30$	$\sum y_i = 180$	$\sum (x_i-\overline{x}) = 0$	$\sum (y_i-\overline{y}) = 0$	$\sum (x_i-\overline{x})(y_i-\overline{y}) = 100$	$\sum (x_i-\overline{x})^2 = 50$	$\sum (y_i-\overline{y})^2 = 242$

① محاسبه ضریب همبستگی و تفسیر آن

توجه ❶ جمع سرمایه گذاری را پیدا می کنیم و از روی آن میانگین سرمایه گذاری را محاسبه می نماییم.

$$\sum x_i = 30 \quad \Rightarrow \quad \overline{x} = \frac{\sum x_i}{n} = \frac{30}{6} = 5$$

توجه ❷ جمع سود سالانه را محاسبه کرده و میانگین آن را بدست می آوریم.

$$\sum y_i = 180 \quad \Rightarrow \quad \overline{y} = \frac{\sum y_i}{n} = \frac{180}{6} = 30$$

توجه ❸ انحراف از میانگین سرمایه گذاری $(x_i - \overline{x})$ را پیدا می کنیم.

توجه ❹ انحراف از میانگین سود سالانه $(y_i - \overline{y})$ را محاسبه می کنیم.

توجه ❺ انحراف از میانگین x ها را در انحراف از میانگین y ها ضرب کرده و با هم جمع می کنیم.

$$\sum (x_i-\overline{x})(y_i-\overline{y}) = 100$$

$$S = x + y \ \ \& \ \ P = xy \quad \Longrightarrow \quad \sum (x_i-\overline{x})(y_i-\overline{y}) = SP_{xy}$$

توجه ❻ انحراف از میانگین سرمایه گذاری را به توان دو می رسانیم و ستون حاصل را جمع می کنیم.

$$\sum (x_i-\overline{x})^2 = SS_x = 50$$

توجه ❼ انحراف از میانگین سود سالانه را به توان دو رسانده و ستون حاصل را جمع می کنیم.

$$\sum (y_i - \overline{y})^2 = SS_y = 242$$

توجه ❽

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 . \sum(y_i - \overline{y})^2}} = \frac{SP_{xy}}{\sqrt{SS_x . SS_y}} = \frac{100}{\sqrt{50 * 242}} = \frac{100}{110} = 0.909 \simeq 0.91$$

توجه ❾

اگر $r = 0$ باشد، بین x و y همبستگی وجود ندارد.

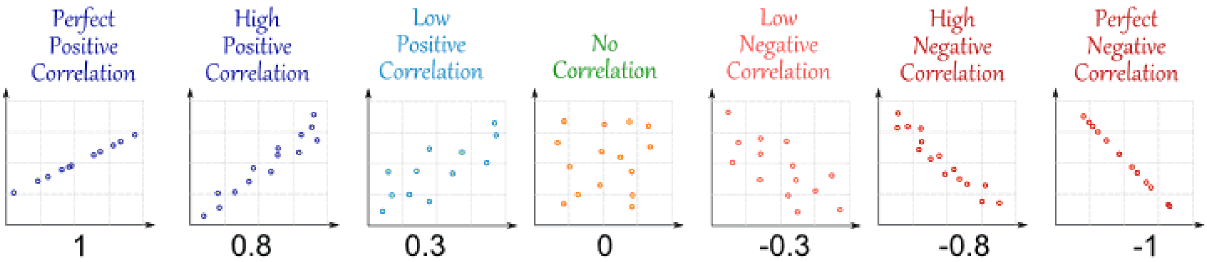
اگر $r = +1$ باشد، بین x و y همبستگی کامل و مستقیم است.

اگر $0 < r < +1$ باشد، بین x و y همبستگی ناقص و مستقیم است.

اگر $r = -1$ باشد، بین x و y همبستگی کامل و معکوس است.

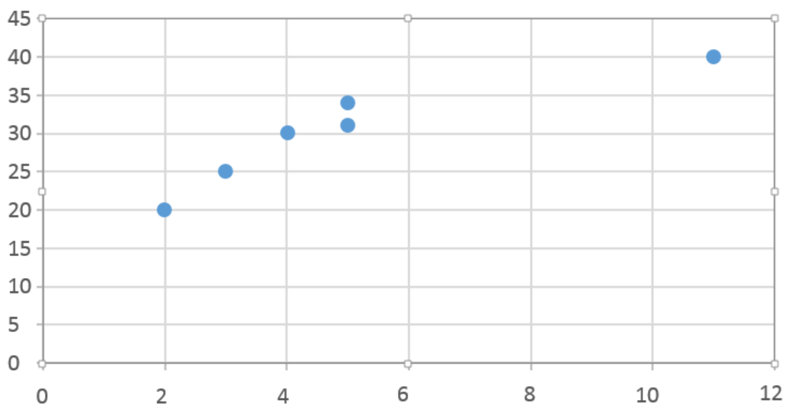
اگر $-1 < r < 0$ باشد، بین x و y همبستگی ناقص و معکوس است.

بین سرمایه گذاری و سود سالانه همبستگی ناقص و مستقیم است $\quad 0 < r = 0.91 < +1 \Rightarrow$



② رسم نمودار پراکندگی و تفسیر نمودار پراگندگی

توجه ❿ دستگاه محورهای مختصات را رسم می کنیم.



بین x و y همبستگی ناقص و مستقیم است

③ محاسبه معادله خط رگرسیون و رسم آن و پیش بینی سود سالانه برای سرمایه گذاری $x = 8$

توجه ⓫ معادله خط در حالت کلی بصورت زیر می باشد:

$$y = a + bx$$

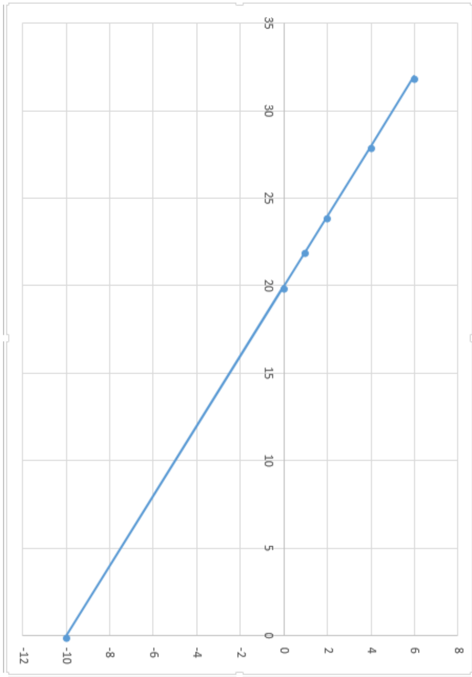
$$b = \frac{SP_{xy}}{SS_x} = \frac{100}{50} = 2$$

$$a = \overline{y} - b\overline{x} = 30 - (2 * 5) = 30 - 10 = 20$$

$$y = a + bx = 20 + 2x \ \Rightarrow y = 20 + (2 * 8) = 20 + 16 = 36$$

$$x \left| \begin{matrix} 0 \\ 20 \end{matrix} \right.$$

$$y \left| \begin{matrix} -10 \\ 0 \end{matrix} \right.$$



توجه ⓬ هر چقدر نقاط دیاگرام پراکنش به خط رگرسیون نزدیک باشد، خط رگرسیون خط مطلوبی برای تصمیم گیری است.

④ آزمون معنی دار بودن ضریب همبستگی به دو روش

توجه ⓭ می خواهیم معنی دار بودن ضریب همبستگی را آزمون کنیم. بگونه ای که ملاحظه می شود، ضریب همبستگی بدست آمده $r = 0.91$ است. این 0.91 مربوط به یک نمونه ۶ عضوی می باشد. اکنون باید بررسی نمود که آیا این ضریب همبستگی نمونه برای جامعه قابل تعمیم هست یا خیر؟ برای این منظور ۲ روش وجود دارد:

روش اول: با استفاده از جدول معنی دار بودن ضریب همبستگی

الف) درجه آزادی را پیدا می کنیم.

$$DF = n - 2 = 6 - 2 = 4$$

ب) درصد اشتباه را پیدا می کنیم.

$$\alpha = 1 - 0.99 = 0.01$$

$$\alpha = 1 - 0.95 = 0.05$$

$$r_{0.01,4} = 0.9172$$

$$r_{0.05,4} = 0.811$$

$$r_{0.05,4} = 0.811 < r = 0.91 < r_{0.01,4} = 0.9172$$

توجه ⓮

$H_0: \rho = 0$ همبستگی معنادار نیست یا همبستگی وجود ندارد.

$H_1: \rho \neq 0$ همبستگی معنادار است یا همبستگی وجود دارد.

اگر قدر مطلق ضریب همبستگی بزرگ تر از هر دو r جدول باشد، آنوقت با احتمال ۹۹٪ همبستگی معنادار است و قابل تعمیم به جامعه می باشد.

اگر قدر مطلق ضریب همبستگی بین دو عدد باشد با احتمال ۹۵٪ همبستگی معنادار است.

اگر قدر مطلق ضریب همبستگی کوچکتر از $r_{0.05}$ و $r_{0.01}$ باشد با اطمینان ۹۹٪ همبستگی معنادار نیست.

در این مسأله چون r محاسبه شده بین $r_{0.05}$ و $r_{0.01}$ قرار دارد با احتمال ۹۵٪ همبستگی معنادار بوده و وجود دارد.

جلسه دوم

ضرایب همبستگی

روش دوم- آزمون ضریب همبستگی با استفاده از روش t

توجه ❶ در این روش ابتدا فرضیه های آماری را بصورت زیر می نویسیم:

$$\begin{cases} \boldsymbol{H_0: \rho = 0 = \rho_0} & \text{.بین میزان سرمایه گذاری و سود سالانه همبستگی معنادار نیست} \\ \boldsymbol{H_1: \rho \neq 0 \neq \rho_0} & \text{.بین میزان سرمایه گذاری و سود سالانه همبستگی معنادار است} \end{cases}$$

آماره آزمون و نقطه بحرانی را از جدول t با در نظر گرفتن سطح تشخیص α و درجه آزادی $DF = n - 2$ پیدا می کنیم. اگر آماره آزمون در منطقه بحرانی قرار نگیرد. فرضیه H_0 رد نمی شود، یعنی ضریب همبستگی جامعه را می توان $\rho = 0$ در نظر گرفت و اگر آماره آزمون در ناحیه بحرانی قرار گیرد H_0 رد می شود.

$$\begin{cases} \boldsymbol{H_0: \rho = 0 = \rho_0} \\ \boldsymbol{H_1: \rho \neq 0 \neq \rho_0} \end{cases}$$

توجه ❷ آماره آزمون را می نویسیم.

$$Q_c = t = \frac{r - \rho_0}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{0.91 - 0}{\sqrt{\frac{1 - (0.91)^2}{6 - 2}}} = 4.4$$

توجه ❸

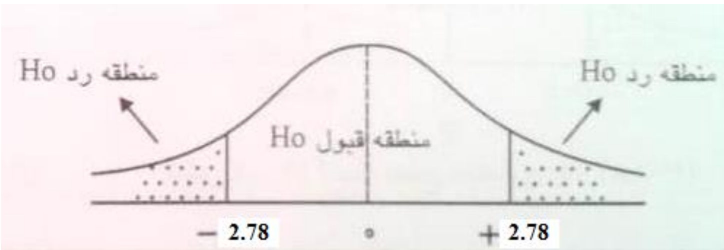
$$t_{1-\frac{\alpha}{2},n-2} \begin{cases} 0.95 & 1 - \frac{0.05}{2}, 6 - 2 = t_{0.975,4} = 2.78 \\ 0.99 & 1 - \frac{0.01}{2}, 6 - 2 = t_{0.995,4} = 4.6 \end{cases}$$

$4.4 \notin w$، فرض H_0 رد نمی شود. با اطمینان 0.99 بین میزان سرمایه گذاری و سود سالانه همبستگی معنادار نبوده و وجود ندارد.

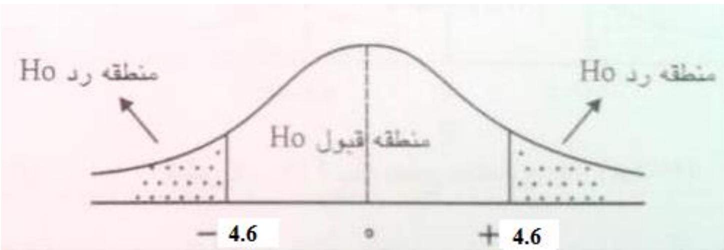
توجه ❹

برای ± 2.78 منحنی را رسم می کنیم.

فرض H_0 با اطمینان *0.95* رد می شود یعنی $\rho \neq 0$. پس بین میزان سرمایه گذاری و سود سالانه همبستگی معنادار بوده و وجود دارد.



برای ± 4.4 منحنی را رسم می کنیم. فرض H_0 با اطمینان *0.99* رد نمی شود یعنی $\rho = 0$. پس بین میزان سرمایه گذاری و سود سالانه همبستگی معنادار نبوده و وجود ندارد.



توجه ❺ به e_i ا شتباه ت صادفی مدل رگر سیونی می گویند و انحراف معیار e_i ها را با S_e نشان می دهند و به ا سم خطای معیار برآورده می باشد. واریانس e_i ها را با ${S_e}^2$ نشان می دهند. ${S_e}^2$ بصورت زیر محاسبه می شود:

$$
{S_e}^2 = \frac{SS_y - \frac{{SP_{xy}}^2}{SS_x}}{n-2} = \frac{242 - \frac{(100)^2}{50}}{6-2} = 10.5
$$

$$
S_e = \sqrt{{S_e}^2} = \sqrt{10.5} = 3.24
$$

تخمین نقطه ای و فاصله ای شیب و مقدار ثابت معادله رگرسیون جامعه با اطمینان 0.95

بطوریکه می دانیم معادله خط رگر سیون $y = a + bx$ از روی اطلاعات نمونه به د ست آمده ا ست. بنابراین برآوردی از معادله خط رگر سیون واقعی جامعه یعنی $y = \alpha + \beta x$ می با شد. به کمک برآورد کننده نقطه ای a و b و با در نظر گرفتن احتمال P می توان برای ضرایب α و β خط رگرسیون جامعه، فاصله اعتماد را به شرح ذیل به دست آورد:

$$
a = 20 \qquad b = 2
$$

برآورد فاصله ای برای α به ازای معنادار بودن (هر کدام که معنادار شد)

$$
\boldsymbol{\alpha = a \pm T * S_e * \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_x}}}
$$ برآورد فاصله ای

برآورد فاصله ای برای β به ازای معنادار بودن (هر کدام که معنادار شد)

$$
\boldsymbol{\beta = b \pm T * S_e * \sqrt{\frac{1}{SS_x}}}
$$ برآورد فاصله ای

$$
\alpha = 20 \pm 2.78 * 3.24 * \sqrt{\frac{1}{6} + \frac{5^2}{50}} = 20 \pm 7.3 = \begin{cases} +27.3 \\ +12.7 \end{cases}
$$ برآورد فاصله ای

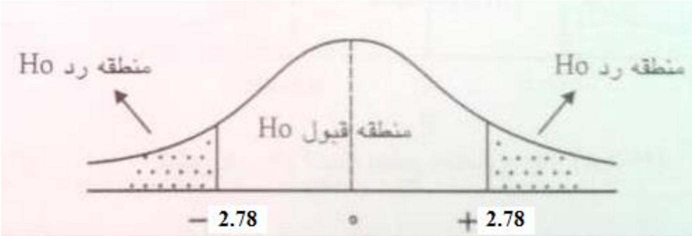
$$
\beta = 2 \pm 2.78 * 3.24 * \sqrt{\frac{1}{50}} = 2 \pm 9.0072 * \sqrt{0.02} = \begin{cases} +3.27 \\ +0.73 \end{cases}
$$ برآورد فاصله ای

آزمون α و آزمون β

توجه ❻ آزمون فر ضیه $H_0{:}\,\alpha = 25$ و آزمون فر ضیه $H_1{:}\,\beta = 0$ ا ست که در آن α و β به ترتیب ضرایب معادله خط رگرسیون جامعه می باشد. برای اجرای آزمون، در ابتدا آماره آزمون را می نویسیم.

$$
\begin{aligned} H_0&{:}\,\alpha = 25 = \alpha_0 \\ H_1&{:}\,\alpha \neq 25 \neq \alpha_0 \end{aligned} \qquad Q_c = \frac{a - \alpha_0}{S_e} . \sqrt{\frac{n.SS_x}{SS_x + n\bar{x}^2}} = \frac{20-25}{3.24} . \sqrt{\frac{6 * 50}{50 + (6 * 5^2)}} = -1.89
$$

$$
\begin{aligned} H_0&{:}\,\beta = 0 = \beta_0 \\ H_1&{:}\,\beta \neq 0 \neq \beta_0 \end{aligned} \qquad Q_c = \frac{b - \beta_0}{S_e} . \sqrt{SS_x} = \frac{2-0}{3.24} . \sqrt{50} = 4.36
$$



$-1.89 \notin \boldsymbol{w}$ **در نتیجه فرض** H_0 رد نمی شود،
یعنی $\alpha = 25$.

$4.36 \in \boldsymbol{w}$ **در نتیجه فرض** H_1 رد نمی شود،
یعنی $\beta \neq 0$.

تعیین ضریب تشخیص یا ضریب تعیین و تفسیر آن

توجه❼ ضریب تشخیص نشان می دهد که چند درصد از تغییرات سود سالانه ناشی از تغییرات سرمایه گذاری می باشد و چند درصد از تغییرات سود سالانه ارتباطی به سرمایه گذاری ندارد.

ضریب تشخیص = (ضریب همبستگی)^2

$$\boldsymbol{R^2 = r^2 = 0.91^2 = 0.83}$$

تفسیر: ۸۳٪ از تغییرات سود سالانه مربوط به تغییرات سرمایه گذاری می باشد و ۱۷٪ از تغییرات سود سالانه ارتباطی به سرمایه گذاری ندارد و مربوط به سایر عوامل غیر از سرمایه گذاری می باشد.

مسأله ۲) به منظور بررسی همبستگی بین تعداد کارشناسان اداری (x) و میزان کارآیی (y) موسسات دولتی، نمونه ای تصادفی به حجم $n = 10$ موسسه انتخاب و نتایج زیر حاصل گردید.

$$Cov_{(x,y)} = 12 \quad / \quad \sum x_i = \sum y_i = 50 \quad / \quad \delta_x = 4 \quad / \quad \delta_y = 5$$

❶ ضریب همبستگی و معادله خط رگرسیون y/x را در این نمونه حساب کنید.

❷ با استفاده از آزمون t معنی دار بودن ضریب همبستگی را در سطح $\alpha = 0.05$ آزمون کنید.

$$r = \frac{Cov_{(x,y)}}{\delta_x \delta_y} = \frac{12}{20} = 0.6$$

بین تعداد کارشناسان اداری و میزان کارآیی، همبستگی ناقص و مستقیم وجود دارد.

$$y = a + bx \quad \Rightarrow \quad b = \frac{SP_{xy}}{SS_x} = \frac{\frac{SP_{xy}}{n}}{\frac{SS_x}{n}} = \frac{Cov_{(x,y)}}{{\delta_x}^2}$$

$$b = \frac{Cov_{(x,y)}}{{\delta_x}^2} = \frac{12}{4^2} = 0.75$$

$$a = \overline{y} - b\overline{x} \qquad \Rightarrow \quad a = 5 - (0.75)(5) = 1.25$$

$$\overline{x} = \frac{\sum x_i}{n} = \frac{50}{10} = 5$$

$$\overline{y} = \frac{\sum y_i}{n} = \frac{50}{10} = 5$$

$$y = a + bx = 1.25 + 0.75x$$

$$\begin{cases} \boldsymbol{H_0: \rho = 0 = \rho_0} \\ \boldsymbol{H_1: \rho \neq 0 \neq \rho_0} \end{cases}$$

بین تعداد کارشناسان اداری و میزان کارایی همبستگی معنادار نیست .

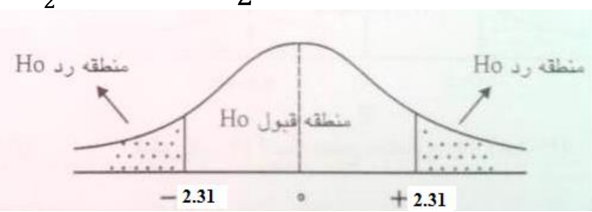
بین تعداد کارشناسان اداری و میزان کارایی همبستگی معنادار است .

$$Q_c = t = \frac{r - \rho_0}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{0.6 - 0}{\sqrt{\frac{1 - (0.6)^2}{10 - 2}}} = 2.12$$

$$DF = n - 2 = 10 - 2 = 8$$

$$\alpha = 0.05$$

$$t_{1-\frac{\alpha}{2}, n-2} \Rightarrow 1 - \frac{0.05}{2}, 10 - 2 = \ t_{0.975,8} = 2.31$$



چون $Q_c \notin w$ فرضیه $\boldsymbol{H_0}$ رد نمی شود و همبستگی معنادار نیست، بین x و y بستگی خطی وجود ندارد.

جلسه سوم

ضرایب همبستگی

م سأله ۱) پز شک محقق ثابت کرده ا ست که سیگار ک شیدن می تواند منجر به چین خوردگی پو ست در اطراف چ شم ها شود. برای این کار، جدول توافقی برای یک نمونه ت صادفی ۵۰۰ نفری را در نظر می گیریم. با ا ستفاده از انواع همبستگی های نسبت ها، تشخیص بدهید که آیا چین خوردگی اطراف چشم با سیگار کشیدن مربوط است یا نه؟

چین خوردگی / عادت به ...	آشکار	غیر آشکار	جمع
سیگاری	$a = 95$	$b = 55$	$g = 150$
غیر سیگاری	$c = 103$	$d = 247$	$h = 350$
جمع	$e = 198$	$f = 302$	$N = 500$

حل مسأله:

توجه ۱) همبستگی نسبت ها را نام ببرید.

1. ضریب توافق چوپروف
2. ضریب فی
3. ضریب توافق پیرسون
4. ضریب توافق کرامر
5. ضریب توافق یول

توجه ۲) همبستگی نسبت ها (ضریب توافق چوپروف) در بحث های قبلی، وقتی دو صفت متغیر مورد مطالعه کمی و یا کیفی ولی رتبه ای بودند را مورد مطالعه قرار دادیم.

سوال) اینک این مطلب مطرح ا ست که اگر دو صفت مورد مطالعه، ترکیبی از صفات کمی و کیفی و رتبه ای و یا بطور کلی، یکی یا هر دو صفت کیفی غیر رتبه ای (ا سمی با شند) آنگاه چگونه می توان میزان وابستگی صفات را مورد مطالعه قرار داد؟

در این حالت تنها اعدادی که در اختیار قرار می گیرند، فراوانی های توزیع ستاد می باشند (چون متغیرها اسمی هستند). پس مسأله ساختن مشخص کننده های کمی (شاخص) فقط به فراوانی ها مربوط می شود. در این حالت برای تعیین میزان شدت وابستگی بین صفات از فرمول چوپروف استفاده می کنند.

توجه ۳) فرمول چوپروف فقط از فراوانی های جدول تبعیت می کنند و متغیرها در این فرمول دخالتی ندارند.

$$T_c = \sqrt{\frac{\chi^2}{N.\sqrt{(K-1)(I-1)}}}$$

توجه ۴) مقدار T_c بین ۰ و ۱ در نوسان است. $0 \leq T_c \leq 1$

مانند فرمول قبل، اگر به مقدار عددی 1 نزدیک با شد، پیوند شدید صفات ن سبت به هم را ن شان می دهد. در هر صورت تا حدودی به حجم نمونه بستگی دارد. اگر حجم نمونه بیش از ۱۰۰ با شد، در صورتیکه T_c محا سبه شده ۰،۲ و یا بیشتر با شد، تقریباً نشانگر میزان وابستگی دو صفت نسبت به هم می باشد.

شدت وابستگی زیاد است	$T_c \geq 0.2$

در فرمول بالا χ^2 یک عدد است.

توجه ۶) K و I را به ترتیب تعداد سطرها و ستون های جدول مورد مطالعه می نامند.

توجه ۷) یک نکته اساسی در کاربرد این فرمول زمانی است که به کمک χ^2 متوجه شده باشیم که بین دو صفت مورد مطالعه، اثر متقابل (پیوند) وجود دارد. یعنی فرض H_1 مورد قبول باشد.

توجه ۸) در حالت خاص که تعداد سطر و ستون جدول داده ها مساوی باشند، یعنی $K = I$ گردد، فرمول چوپروف بصورت زیر محاسبه می شود.

$$T_c = \sqrt{\frac{\chi^2}{N.\sqrt{(K-1)(I-1)}}} = \sqrt{\frac{\chi^2}{N.(K-1)}}$$

توجه ۹) اگر $K = I = 2$ باشد یعنی جدول توافقی بصورت چهارخانه گردد، آن وقت:

$$T_c = \sqrt{\frac{\chi^2}{N.(2-1)}} = \sqrt{\frac{\chi^2}{N}}$$

یعنی $T_c = \sqrt{\frac{\chi^2}{N}}$ که معمولاً آن را با فی (φ) نشان می دهند.

$$\varphi = \sqrt{\frac{\chi^2}{N}}$$

توجه ۱۰) در این صورت مقدار χ^2 را می توان از فرمول زیر به دست آورد:

$$\chi^2 = \frac{N.(ad-bc)^2}{e.f.g.h}$$

$$\varphi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{\frac{N(ad-bc)}{e.f.g.h}}{N}} = \frac{|ad-bc|}{\sqrt{e.f.g.h}}$$

$$\varphi = \frac{|ad-bc|}{\sqrt{e.f.g.h}} = \frac{|(95*247)-(55*103)|}{\sqrt{198*302*150*350}} = 0.318$$

توجه ۱۱) فرمول ضریب یول

$$Q = \frac{|ad-bc|}{ad+bc} = \frac{|(95*247)-(55*103)|}{(95*247)+(55*103)} = 0.61$$

توجه ۱۲) مقادیر φ و Q گویای این مطلب هستند که چین خوردگی اطراف چشم به سیگار کشیدن مربوط است و این مربوط بودن چون هر دو بزرگ تر از ۰٫۲ هستند، این مربوط بودن زیاد است.

توجه ۱۳) لازم به یادآوری است که وقتی صفات مورد مطالعه کیفی و به صورت اسمی باشند مانند چین خوردگی، آنگاه در صورت فرمول یول علامت قدر مطلق می گذارند. زیرا انتخاب جا برای حالات صفات اختیاری است.

توجه ۱۴) فرمول ضریب توافق پیرسون به شرح ذیل می باشد:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

توجه ۱۵) فرمول ضریب توافق پیرسون، اندازه، درجه، بستگی یا ارتباط طبقه بندی ها در یک جدول توافقی را نشان می دهد.

توجه ۱۶) ضریب توافق پیرسون را اگر محاسبه کنیم، مقدار آن بین ۰ و ۱ نوسان می کند.

توجه ۱۷) هرچه C بزرگ تر باشد، درجه بستگی صفات نسبت به هم بیشتر است و مقدار Max آن برابر است با:

$$C_{Max} = \sqrt{\frac{k-1}{k}}$$

در نتیجه $\sqrt{\frac{k-1}{k}} \geq C \geq 0$ قرار می گیرد.

توجه ۱۸) در فرمول بالا k عبارت است از تعداد سطرها و ستون هایی که با هم برابرند. با توجه به مقدار Max معلوم می شود که مقدار C هرگز بزرگتر از 1 نمی بشد.

توجه ۱۹) مقدار C_{Max} وقتی حاصل می شود که دو گروه کاملاً وابسته یا به هم مربوط باشند.

توجه ۲۰) حال C را محاسبه می کنیم. برای محاسبه C، χ^2 را بدست می آورید.

$$\chi^2 = \frac{N.(ad-bc)^2}{e.f.g.h} = \frac{500.(95*247-55*103)^2}{198*302*350*150} = 50.46$$

$$C = \sqrt{\frac{\chi^2}{\chi^2+N}} = \sqrt{\frac{50.46}{50.46+500}} = 0.303$$

توجه ۲۱) ضرایب توافق پیرسون را معمولاً برای جدول مربع به کار می برند.

توجه ۲۲) اگر توافقی بصورت مستطیل باشد، آنگاه از ضریب توافق کرانر که آن را با V نمایش می دهیم، استفاده می کنیم.

توجه ۲۳) فرمول ضریب توافق کرانر بصورت زیر می باشد:

$$V = \sqrt{\frac{\chi^2}{N.Min(k-1,I-1)}}$$

توجه ۲۴) اگر جدول ۴*۶ باشد، آنگاه:

$$Min(k-1,I-1) = Min(4-1,6-1) = Min(3,5) = 3$$

توجه ۲۵) در حالت خاص که جدوال توافق ۲*۲ می باشد، آنگاه فرمول V با فرمول φ یکسان خواهد بود.

توجه ۲۶) گفتیم χ^2 برابر ۵۰,۴۶ می باشد و $Min(k-1,I-1)$ برابر با $Min(2-1,2-1) = 1$

توجه ۲۷) V را محاسبه می کنیم.

$$V = \sqrt{\frac{\chi^2}{N.Min(k-1,I-1)}} = \sqrt{\frac{50.46}{500*1}} = 0.317$$

توجه ۲۸) جواب V همان جواب φ است.

مسأله ۲) به منظور مطالعه میزان وابستگی مشارکت دانشجویان در مسائل اجتماعی و انگیزه پیشرفت آنان از بین کلیه دانشجویان دانشگاه ۲۴۰ نفر را به عنوان نمونه انتخاب کرده، پس از تکمیل پر سشنامه، جدول زیر را تنظیم کردیم. (توجه دا شته با شید که هر دو صفت از بالا به پایین مرتب شده ا ست) با ا ستفاده از همبستگی های رتبه ای، آیا می توان گفت مشارکت دانشجویان در مسائل اجتماعی به انگیزه پیشرفت آن ها مربوط است.

انگیزه پیشرفت	مشارکت در مسائل اجتماعی				جمع
	بالا ۱	نسبتاً بالا ۲	متوسط ۳	پایین ۴	
بالا ۱	۲۶	۲۲	۱۲	۷	۶۷
نسبتاً بالا ۲	۲۰	۱۸	۱۱	۹	۵۸
متوسط ۳	۱۰	۱۴	۲۳	۱۷	۶۴
پایین ۴	۵	۶	۱۶	۲۴	۵۱
جمع	۶۱	۶۰	۶۲	۵۷	$N = 240$

حل مسأله)

توجه ۲۹) همبستگی های رتبه ای را نام ببرید.

1. ضریب همبستگی کندال τ_a (تا اُ آ)
2. ضریب گاما γ
3. ضرایب سامرز $dx|y$ و $dy|x$
4. میانگین هندسی ضرایب سامرز τ_b
5. τ_c

توجه ۳۰) فرمول ضریب همبستگی کندال τ_a به شرح ذیل است:

$$\tau_a = \frac{NS - Nd}{T}$$

توجه ۳۱) تعداد کل زوج ها را T می گویند.

توجه ۳۲) T را محاسبه می کنیم.

$$T = C_N^2 = \frac{N!}{2!\,(N-2)!} = \frac{N(N-1)(N-2)!}{(2*1)(N-2)!} = \frac{N(N-1)}{2}$$

$$T = \frac{N(N-1)}{2} = \frac{240(240-1)}{2} = 28680$$

توجه ۳۳) فرمول τ_a بر اساس زوج ها پایه گذاری شده است.

توجه ۳۴) NS عبارت است از تعداد زوج هایی که با افزایش یا کاهش رتبه xها و yها افزایش یا کاهش یابد.

توجه ۳۵) یک جدول ۲*۳ داریم: اندیس به معنی رتبه در نظر گرفته شده است. NS به شرح ذیل محاسبه می شود:

A_i \ B_j	A_1	A_2
B_1	n_{11}	n_{12}
B_2	n_{21}	n_{22}
B_3	n_{31}	n_{32}

$$NS = n_{11}(n_{22} + n_{32}) + n_{21}.n_{32}$$

توجه ۳۶) در مسأله بالا NS را محاسبه می کنیم.

$$\begin{aligned} NS = 26&(18 + 11 + 9 + 14 + 23 + 17 + 6 + 16 + 24) \\ &+ 20(14 + 23 + 17 + 6 + 16 + 24) + 10(6 + 16 + 24) \\ &+ 22(11 + 9 + 23 + 17 + 16 + 24) + 18(23 + 17 + 16 + 24) \\ &+ 14(16 + 24) + 12(9 + 17 + 24) + 11(17 + 24) + 23(24) = 11851 \end{aligned}$$

توجه ۳۷) NS عبارت ا ست از تعداد زوج هایی که جهت تغییرات رتبه های X و Y آن ها هماهنگ و یک سان با شد. یعنی اگر رتبه X افزایش یا کاهش نشان دهد رتبه Y نیز در همان جهت باشد. به سخن دیگر با افزایش رتبه X، رتبه Y نیز کاهش نشان می دهد.

توجه ۳۸) Nd عبارت است از تعداد زوج هایی که با افزایش رتبه Xها، رتبه Yها کاهش می یابد و یا برعکس.

توجه ۳۹) Nd را در جدول ۲*۳ محاسبه کنید.

$$Nd = n_{12}(n_{21} + n_{31}) + n_{22}.n_{31}$$

توجه ۴۰) Nd عبارت اســـت از تعداد زوج هایی که جهت تغییرات رتبه های X و Y آن ها ناهماهنگ و به عبارتی جهت تغییرات آن ها مخالف هم باشند. یعنی اگر رتبه Xها صعودی می باشد، رتبه Yها نزولی شوند و یا بر عکس.

توجه ۴۱) در این مسأله Nd را محاسبه می کنیم.

$$\begin{aligned} Nd = 7&(11 + 18 + 20 + 23 + 14 + 10 + 16 + 6 + 5) + 9(23 + 14 + 10 + 16 + 6 + 5) \\ &+ 17(16 + 6 + 5) + 12(18 + 20 + 14 + 10 + 6 + 5) \\ &+ 11(14 + 10 + 6 + 5) + 23(6 + 5) + 22(20 + 10 + 5) + 18(10 + 5) \\ &+ 18(10 + 5) + 14(5) = 4610 \end{aligned}$$

توجه ۴۲) T_x عبارت است از تعداد زوج هایی که رتبه Xها برابر ولی رتبه Yها نابرابر باشد.

توجه ۴۳) در جدول ۲*۳، T_x را محاسبه کنید.

$$T_x = n_{11} * n_{12} + n_{21} * n_{22} + n_{31} * n_{32}$$

توجه ۴۴) در مسأله بالا T_x را محاسبه می کنیم.

$$\begin{aligned} T_x = 26&(22 + 12 + 7) + 22(12 + 7) + 12(7) + 20(18 + 11 + 9) + 18(11 + 9) + 11(9) \\ &+ 10(14 + 23 + 17) + 14(23 + 17) + 23(17) + 5(6 + 16 + 24) \\ &+ 6(16 + 24) + 16(24) = 5132 \end{aligned}$$

توجه ۴۵) T_y عبارت است از تعداد زوج هایی که رتبه Yها برابر ولی رتبه Xها نابرابر باشد.

توجه ۴۶) در جدول ۲*۳، T_y را محاسبه می کنیم.

$$T_y = n_{11}(n_{21} + n_{31}) + n_{21} * n_{31} + n_{12}(n_{22} + n_{32}) + n_{22} * n_{32}$$

توجه ۴۷) T_y را در مسأله بالا به صورت زیر محاسبه می کنیم:

$$T_y = 26(20+10+5)+20(10+5)+10(5)+22(18+14+6)+18(14+6)+14(6) + 12(11+23+16)+11(23+16)+23(16)+7(9+17+24) + 9(17+24)+17(24) = 5064$$

$$\tau_a = \frac{NS-Nd}{T} = \frac{11851-4610}{28680} = 0.25$$

توجه ۴۸) اگر جمع نمونه ۱۰۰ یا بشتر باشد ($N \geq 100$) و در صورتیکه مقدار محاسبه شده هرکدام از ضرایب یاد شده برابر با ۰,۲ یا بیشتر باشند آنگاه میزان وابستگی دو متغیر تا حد زیاد است و چون τ_a برابر ۰,۲۵ و بزرگ تر از ۰,۲ است پس با توجه به τ_a بدست آمده می توان گفت مشارکت دانشجویان در مسائل اجتماعی به انگیزه پیشرفت آنان مربوط می شود.

$$\gamma = \frac{NS-Nd}{NS+Nd} = \frac{11851-4610}{11851+4610} = 0.44$$

چون $\gamma = 0.44 > 0.2$ تفسیر τ_a در مورد γ نیز صادق است.

توجه ۴۹) ضرایب سامرز را محاسبه می کنیم:

$$dy|x = \frac{NS-Nd}{NS+Nd+T_y} = \frac{11851-4610}{11851+4610+5064} = 0.3364$$

$$dx|y = \frac{NS-Nd}{NS+Nd+T_x} = \frac{11851-4610}{11851+4610+5132} = 0.3353$$

توجه ۵۰) میانگین هندسی ضرایب سامرز را محاسبه می کنیم که با τ_b نشان می دهند.

$$\tau_b = \sqrt{dy|x * dx|y} = \sqrt{0.3364 * 0.3353} = 0.3358$$

توجه ۵۱) چون $\tau_b = 0.3358 > 0.2$ است، پس مشارکت دانشجویان در مسائل اجتماعی به انگیزه پیشرفت آنان مربوط است، این مربوط بودن زیاد می باشد.

توجه ۵۲) τ_b مناسب ترین شاخصی است که میزان وابستگی صفات را نشان می دهد ولی در بعضی موارد γ مناسب تر است.

توجه ۵۳) بطوریکه ملاحظه می شود:

$$\tau_a \leq \tau_b \leq |\gamma| \qquad \Rightarrow \qquad 0.25 < 0.3358 < 0.44$$

توجه ۵۴) برای محاسبه τ_c می گوییم:

$$\tau_c = \frac{NS-Nd}{\frac{N^2(m-1)}{2m}}$$

توجه ۵۵) در فرمول τ_c، که m بکار رفته است، m یعنی مینیمم تعداد سطر و ستون. مثلاً اگر جدول دارای ۳ سطر و ۵ ستون باشد، آنگاه $m = 3$

توجه ۵۶) این فرمول برای جداولی که بصورت مربع یا مستطیل باشند بکار برده می شوند.

توجه ۵۷) ایراد این فرمول این است که چون مخرج آن فقط به تعداد سطرها و ستون ها مربوط است و به توزیع حاشیه ای وابسته نیست، لذا این امر تا حدودی تفسیر τ_c را دشوار می کند. به همین دلیل از مطلوبیت کمتری نیست به τ_b برخوردار است.

توجه ۵۸) این مسأله دارای ۴ سطر و ۴ ستون می باشد، پس $m = 4$

توجه ۵۹) τ_c را محاسبه می کنیم:

$$\tau_c = \frac{NS - Nd}{\frac{N^2(m-1)}{2m}} = \frac{11851 - 4610}{\frac{240^2(4-1)}{2 * 4}} = 0.3352$$

توجه ۶۰) چون $\tau_c = 0.3352 > 0.2$ با توجه یه ضرایب بدست آمده معلوم می شود که مشارکت دانشجویان در مسائل اجتماعی به انگیزه پیشرفت آنان مربوط است یعنی هر چقدر انگیزه پیشرفت آن ها زیادتر باشد در مسائل اجتماعی بیشتر شرکت می کنند.

توجه ۶۱) این سوال پیش می آید که کدام یک از ضرایب همبستگی رتبه ای بطور کامل محبوب است؟

در حال حاضر این مسأله که کدام یک از ضرایب همبستگی رتبه ای مطلوب تر هستند، همچنان یک سوال باز است یعنی در حال حاضر هیچکدام به قدر کافی از مطلوبیت تام برخوردار نیستند. بدین لحاظ یک محقق ناگزیر است که ضریب همبستگی رتبه ای را از فرمول های متفاوت محاسبه نماید و ببینند که آیا با توجه به یافته های مورد نظر وی نتایج مشابهی به دست می دهد یا نه؟

جلسه چهارم

ضرایب همبستگی

مسأله ۱) جدول زیر نتیجه آمارگیری نیروی انسانی در یکی از شهرهای بزرگ را نشان می دهد. مطلوب است:

الف) ضریب کروسکال

ب) ضریب گاتمن

ج) تفسیر آن ها

جنس / علل بی سوادی	فقر مالی	مخالفت والدین	نبودن مدرسه	عدم علاقه	$n_{i\circ}$
مرد	۵۷	۸	۱۱	۱۲	۸۸
زن	۲۷	۴۸	۱۰	۷	۹۲
جمع $n_{\circ j}$	۸۴	۵۶	۲۱	۱۹	۱۸۰

توجه ۶۲) همبستگی غیر رتبه ای را نام ببرید.

1. ضریب کروسکال
2. ضریب گاتمن

توجه ۶۳) ضرایب وابستگی غیر رتبه ای (اسمی) می باشد.

توجه ۶۴) ضریب کروسکال: این ضریب اعم از اینکه متغیر کیفی یا کمی باشد، به کار برده می شود زیرا فقط از فراوانی های جدول تبعیت می کند و به مقادیر متغیر ارتباطی ندارد.

توجه ۶۵) فرمول ضریب کروسکال به شرح زیر است.

$$\tau_y = \frac{E_1 - E_2}{E_1}$$

توجه ۶۶) E_1 از فرمول زیر به دست می آید.

$$E_1 = \sum_{j=1}^{I} \left(\frac{N - n_{\circ j}}{N}\right) n_{\circ j}$$

توجه ۶۷) E_2 از فرمول زیر به دست می آید.

$$E_2 = \sum_{i=1}^{K} \sum_{j=1}^{I} \left(\frac{n_{i\circ} - n_{ij}}{n_{i\circ}}\right) n_{ij}$$

توجه ۶۸) لازم به یادآوری است که $n_{i\circ}$ فراوانی توزیع حاشیه ای x ها است.

توجه ۶۹) $n_{\circ j}$ فراوانی توزیع y ها است.

توجه ۷۰) ضریب گاتمن: این ضریب نیز از فراوانی های جدول توافقی تبعیت می کند. اغلب در مواردی به کار برده می شود که نتوان صفات مورد مطالعه را بصورت رتبه ای طبقه بندی کرد.

توجه ۷۱) فرمول ضریب گاتمن به شرح ذیل است:

$$\lambda = \frac{\sum my - My}{N - My}$$

توجه ۷۲) My عبارت است از بزرگ ترین فراوانی توزیعی یعنی $Max\ n_{\circ j}$

توجه ۷۳) my عبارت است از بزرگ ترین فروانی توزیع های شرطی y ها و N (حجم نمونه).

توجه ۷۴) طبیعی است My هرگز بزرگ تر از $\sum my$ نمی شود یعنی $My \leq my$

توجه ۷۵) $0 \leq \lambda \leq 1$

توجه ۷۶) هر چقدر به عدد 1 نزدیک تر باشد میزان وابستگی صفات مورد مطالعه بیشتر است.

توجه ۷۷) فرمول هایی را که فقط از فراوانی ها تبعیت می کنند می توان آن ها را برای صفات کیفی اعم از رتبه ای و غیر رتبه ای و کمی بکار برد.

توجه ۷۸) فرمول هایی که علاوه بر فراوانی ها از مقادیر صفت نیز تبعیت می کنند فقط برای صفات کمی بکار می برند.

توجه ۷۹) حال ضریب کروسکال را محاسبه می کنیم.

$$E_1 = \sum_{j=1}^{I} \left(\frac{N - \mathrm{n}_{\circ j}}{N}\right)\mathrm{n}_{\circ j} = \left(\frac{180-84}{180} * 84\right) + \left(\frac{180-56}{180} * 56\right) + \left(\frac{180-21}{180} * 21\right) + \left(\frac{180-19}{180} * 19\right) = 118.5$$

توجه ۸۰) برای توجیه فرمول بالا می توان گفت:

احتمال خطای مربوط به صنعتی که فراوانی توزیع حاشیه آن برابر $\mathrm{n}_{\circ j}$ است $\frac{N - \mathrm{n}_{\circ j}}{N}$

توجه ۸۱) در مسأله بالا احتمال خطا در مورد فقر مالی برابر است با:

$$\frac{N - \mathrm{n}_{\circ j}}{N} = \frac{180 - 84}{180} = 0.53$$

توجه ۸۲) در مسأله بالا احتمال خطا در مورد مخالفت والدین برابر است با:

$$\frac{N - \mathrm{n}_{\circ j}}{N} = \frac{180 - 56}{180} = 0.69$$

توجه ۸۳) E_2 را بصورت زیر محاسبه می کنیم:

$$E_2 = \sum_{i=1}^{K}\sum_{j=1}^{I} \left(\frac{n_{i\circ} - n_{ij}}{n_{i\circ}}\right) n_{ij} = \left(\frac{88-57}{88} * 57\right) + \left(\frac{88-8}{88} * 8\right) + \left(\frac{88-11}{88} * 11\right) + \left(\frac{88-12}{88} * 12\right) + \left(\frac{92-27}{92} * 27\right) + \left(\frac{92-48}{92} * 48\right) + \left(\frac{92-10}{92} * 10\right) + \left(\frac{92-7}{92} * 7\right) = 104.7$$

توجه ۸۴) ضریب کروسکال را محاسبه می کنیم:

$$\tau_y = \frac{E_1 - E_2}{E_1} = \frac{118.5 - 104.7}{118.5} = 0.12$$

چون $0.2 > 0.12$ است، میزان شدت وابستگی بین علل وابستگی و جنس را نشان می دهد و این میزان شدت وابستگی کم است.

توجه ۸۵) ضریب گاتمن را محاسبه می کنیم:

$$\lambda = \frac{\sum my - My}{N - My} = \frac{(57 + 48) - 84}{180 - 84} = 0.22$$

۰,۲۲ میزان شدت وابستگی علل بی سوادی و جنس را نشان می دهد و چون $0.2 < 0.22$ پس این وابستگی تا حد زیادی است.

نتیجه گیری: تا حدودی می توان گفت که علل بی سوادی به جنس مربوط است.

توجه ۸۶) ممکن ا ست بخواهیم بدانیم کدامیک از فرمول های یاد شده در همبستگی غیر رتبه ای مطلوبیت بیشتری دارد. با توجه به مطالب یاد شـــده در بالا می توان گفت این معیارها به طور پایه ای انواعی از یکدیگرند. همچنین اینکه برای اســـتفاده کدام بر دیگری برتری دارد، موضوعی است که اغلب به عادت افراد در رشته های مختلف مربوط می شود.

آنچه مســـلم اســـت اینکه مقادیر بزرگ این معیارها بطور حتم نشـــانگر یک پیوند قوی هســـتند ولی تعبیری کلی برای مقادیر کوچک و متوســـط وجود ندارد زیرا یک حجم بزرگ نمونه بزرگ n، معیارها را به کوچک شـــدن گرایش می دهد. همانگونه که پیشتر نیز ا شاره شد، وقتی حجم نمونه تا حدودی بزرگ با شد یعنی $n \geq 100$ اگر این معیارها از ۰,۲ بزرگ تر با شند تا حدودی نشانگر وابستگی صفات مورد مطالعه نسبت به هم می باشند.

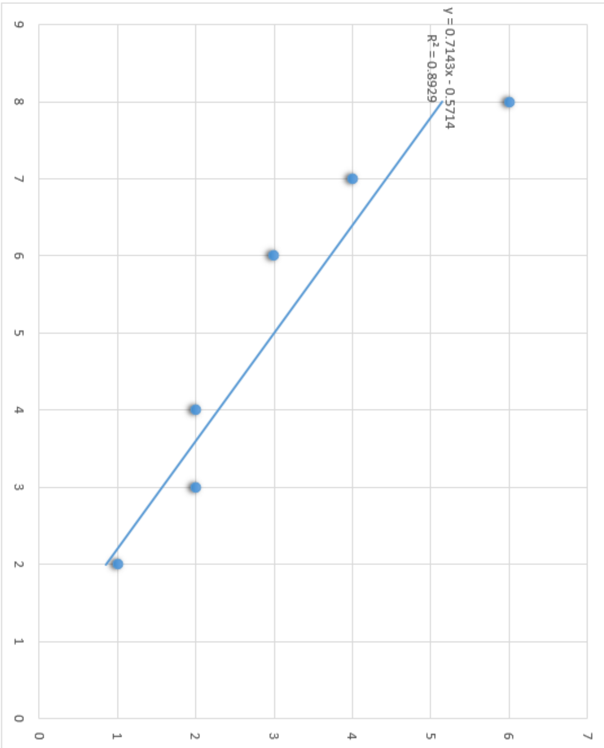
جلسه چهارم

رگرسیون و همبستگی

مسأله ۱) به منظور بررسی همبستگی بین سابقه کار (x) و درآمد روزانه (y) کارکنان یک واحد تولیدی $n = 6$ نفر را به طور تصادفی از بین کارکنان موسسه، انتخاب و پس از تحقیق نتایج زیر حاصل شده است.

سابقه کار	درآمد روزانه	انحراف از میانگین سابقه کار	انحراف از میانگین درآمد روزانه	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
x	y	$x_i - \bar{x}$	$y_i - \bar{y}$			
1	2	-2	-3	6	4	9
2	3	-1	-2	2	1	4
2	4	-1	-1	1	1	1
3	6	0	1	0	0	1
4	7	1	2	2	1	4
6	8	3	3	9	9	9
$\sum x_i = 18$	$\sum y_i = 30$	$\sum (x_i - \bar{x}) = 0$	$\sum (y_i - \bar{y}) = 0$	$\sum (x_i - \bar{x})(y_i - \bar{y}) = 20$	$\sum (x_i - \bar{x})^2 = 16$	$\sum (y_i - \bar{y})^2 = 28$

❶ نمودار پراکندگی را رسم کنید



بین سابقه کار و درآمد روزانه همبستگی ناقص و مستقیم وجود دارد.

نمودار پراکندگی نشان می دهد که احتمالاً بین این دو صفت رابطه خطی وجود دارد.

نمودار پراکندگی یکی از روش های ساده ای است که می توان به طور تقریب و شهودی به وجود رابطه و همبستگی بین x و y و بخصوص نوع این همبستگی پی برد.

❷ در مسأله فوق ضریب همبستگی را پیدا کنید.

$$\sum x_i = 18 \quad \Rightarrow \quad \bar{x} = \frac{\sum x_i}{n} = \frac{18}{6} = 3$$

$$\sum y_i = 30 \quad \Rightarrow \quad \bar{y} = \frac{\sum y_i}{n} = \frac{30}{6} = 5$$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 . \sum (y_i - \bar{y})^2}} = \frac{SP_{xy}}{\sqrt{SS_x . SS_y}} = \frac{20}{\sqrt{16 * 28}} = 0.945$$

چون $1 > 0.945 > 0$ می باشد پس بین سابقه کار و درآمد روزانه همبستگی مستقیم و ناقص برقرار است.

ضریب همبستگی بهترین معیار تشخیص وجود یا عدم وجود همبستگی می باشد.

توجه) اگر متغیرهای x و y یک جامعه دو بعدی را به متغیرهای استاندارد به شرح زیر تبدیل کنیم خواهیم داشت:

$$Z_x = \frac{x - \overline{x}}{\delta_x}$$

$$Z_y = \frac{y - \overline{y}}{\delta_y}$$

به امید ریاضی حاصل ضرب متغیرهای استاندارد Z_x و Z_y که آن را ρ_{xy} نشان می دهیم، ضریب همبستگی دو متغیر x و y می گویند.

$$\rho_{xy} = E(Z_x . Z_y) = E\left(\frac{x - \overline{x}}{\delta_x}\right)\left(\frac{y - \overline{y}}{\delta_y}\right) = \frac{E(x - \overline{x})(y - \overline{y})}{\delta_x \delta_y} = \frac{Cov(x . y)}{\delta_x \delta_y}$$

ضریب همبستگی نمونه را با (r) نشان می دهند و $-1 \leq r \leq +1$

یکی از کاربردهای مهم خط رگر سیون پیش بینی y به ازای هر x معلوم می با شد. برای این منظور مقدار x را در معادله خط رگرسیون قرار می دهیم تا y_i بدست آید.

3 در مسأله فوق معادله خطر رگرسیون را بدست آورید و خط رگرسیون را رسم کنید و به ازای $x = 5$ مقدار y را تخمین بزنید.

$$y = a + bx$$

$$b = \frac{SP_{xy}}{SS_x} = \frac{20}{16} = 1.25$$

$$a = \overline{y} - b\overline{x} = 5 - (1.25 * 3) = 5 - 3.75 = 1.25$$

$$y = a + bx = 1.25 + 1.25x \ \Rightarrow y = 1.25 + (1.25 * 5) = 7.5$$

$$x \Big|_{1.25}^{0} \qquad y \Big|_{0}^{-1}$$

4 در مسأله فوق خطای معیار برآورد را محاسبه کنید.

$$S_e{}^2 = \frac{SS_y - \frac{SP_{xy}{}^2}{SS_x}}{n - 2} = \frac{\frac{28 - (20)^2}{16}}{6 - 2} = 0.75$$

$$S_e = \sqrt{S_e{}^2} = \sqrt{0.75} = 0.866$$

5 در مسأله فوق انحراف معیارهای a و b را محاسبه کنید.

در این مسأله اول S_e را محاسبه می کنیم.

$$S_e = \sqrt{S_e{}^2} = \sqrt{0.75} = 0.866$$

$$S_a = S_e * \sqrt{\frac{1}{n} + \frac{\overline{x}^2}{SS_x}} = 0.866 * \sqrt{\frac{1}{6} + \frac{9}{16}} = 0.739$$

$$S_b = S_e * \sqrt{\frac{1}{SS_x}} = 0.866 * \sqrt{\frac{1}{16}} = 0.2165$$

محاسبه واریانس خطا

امید ریاضی U^2 (خطا در جامعه)

$$E(U^2) = \delta_U^2$$

برای تحلیل و اســـتنباط آماری خط رگرســـیون، احتیاج به دانســـتن واریانس مجهول δ_U^2 و انحراف معیار آن δ_U داریم. δ_U انحراف معیاری است که پراکندگی نقاط بالا و پایین خط رگرسیون را نشان می دهد.

چون δ_U^2 را نمی توان از روی اطلاعات تمام جامعه بدســـت آورد، بنابراین باید از برآورد آن یعنی ${S_e}^2$ که مربوط به نمونه تصادفی باشد به شرح ذیل استفاده نمود:

$${S_e}^2 = \frac{\sum e_i^2}{n-2} \quad \Rightarrow \begin{cases} e_i = y_i - \hat{y} \\ E_{(e_i)} = 0 \end{cases} \quad \Rightarrow {S_e}^2 = \frac{\Sigma(y_i - \overline{y})^2}{n-2} \quad \Longrightarrow S_e = \sqrt{\frac{\Sigma(y_i - \overline{y})^2}{n-2}} = \sqrt{\frac{SS_y - \frac{{SP_{xy}}^2}{SS_x}}{n-2}}$$

برای مطالعه بیشتر

برای پیدا کردن نمونه آماری از فرمول زیر استفاده می گردد:

$$\boldsymbol{\varepsilon = Z_{1-\frac{\alpha}{2},\infty} . \sqrt{\frac{p(1-p)}{n}} . \sqrt{\frac{N-n}{N-1}}} \qquad \boldsymbol{\varepsilon = 0.05}$$

در جوامع با تعداد نامعلوم

$$\boldsymbol{\sqrt{\frac{N-n}{N-1}} = 1}$$

جلسه پنجم

رگرسیون

فاصله اعتماد برای ضرایب α و β خط رگرسیون

اکنون که میانگین و واریانس ضرایب a و b خط رگر سیون نمونه را محا سبه نمودیم، می توانیم برای ضریب خط رگر سیون جامعه یعنی α و β فاصله اطمینان در سطح $\alpha = 1 - p$ محاسبه می کنیم.

$$\alpha = \mathrm{a} \pm T_{n-2\,,1-\frac{\alpha}{2}}.S_a$$

$$\beta = b \pm T_{n-2\,,1-\frac{\alpha}{2}}.S_b$$

در مسأله فوق برای ضرایب α و β خط رگرسیون جامعه ۹۵٪ را محاسبه نمایید.

$$T_{n-2\,,1-\frac{\alpha}{2}} = T_{6-2\,,1-\frac{0.05}{2}} = T_{4\,,0.975} = 2.78$$

$$\alpha = \mathrm{a} \pm T_{n-2\,,1-\frac{\alpha}{2}}.S_a = 1.25 \pm (2.78 * 0.739) = \begin{cases} +3.30442 \\ -0.80442 \end{cases}$$

$$\beta = b \pm T_{n-2\,,1-\frac{\alpha}{2}}.S_b = 1.25 \pm (2.78 * 0.2165) = \begin{cases} +1.85187 \\ +0.64813 \end{cases}$$

آزمون ضرایب α و β خط رگرسیون

در مواردی لازم است که آزمون نماییم آیا عدد $\alpha_\circ$ می تواند عرض از مبدأ خط رگرسیون جامعه باشد یا خیر؟ همچنین آیا شیب خط رگر سیون می تواند مقدار $\beta_\circ$ با شد یا خیر؟ در این صورت با انتخاب یک نمونه تصادفی و محا سبه خط رگر سیون نمونه می توان آزمون های ذیل را انجام داد.

الف) آزمون α

$$\begin{cases} H_\circ: \alpha = \alpha_\circ \\ H_1: \alpha \neq \alpha_\circ \end{cases} \qquad t = \frac{\alpha - \alpha_\circ}{S_a}$$

در این مدل آزمون t جدول را محاسبه می کنیم. اگر آماره آزمون در ناحیه بحرانی قرار نگیرد، فرضیه $H_\circ$ رد نمی شود و $\alpha = \alpha_\circ$

ب) آزمون β

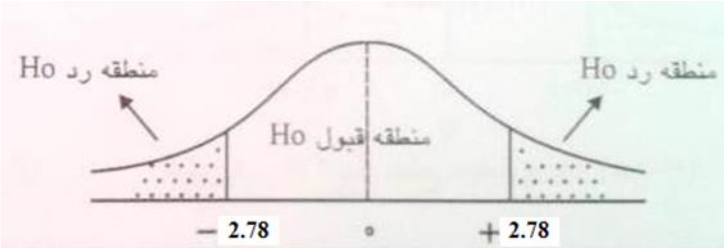
$$\begin{cases} H_\circ: \beta = \beta_\circ \\ H_1: \beta \neq \beta_\circ \end{cases} \qquad t = \frac{b - \beta_\circ}{S_b}$$

در این آزمون t جدول را محاسبه می کنیم و با t محاسبه شده مقایسه می کنیم. اگر آماره آزمون در ناحیه بحرانی قرار نگیرد، فرضیه $H_\circ$ رد نمی شود و عدد $\beta_\circ$ را می توان به عنوان شیب خط رگرسیون جامعه در نظر گرفت.

در مســأله فوق تحقیق نمایید عرض از مبدأ و شــیب خط رگرســیون جامعه می تواند به ترتیب $\boldsymbol{\alpha = 2}$ و $\boldsymbol{\beta = 0}$ باشد.

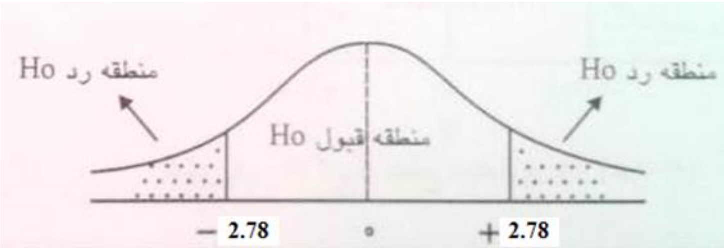
$$t = \frac{\alpha - \alpha_\circ}{S_a} = \frac{1.25 - 2}{0.739} = -1.015 \qquad t_{\text{جدول}} = 2.78 \qquad -1.015 \notin w$$

آماره آزمون در ناحیه بحرانی قرار نگرفته اســت پس فرضــیه $H_\circ$ رد نمی شــود و $\alpha = \alpha_\circ = 2$ می باشد.



$$t = \frac{\beta - \beta_\circ}{S_b} = \frac{1.25 - 0}{0.2165} = 5.77 \qquad t_{\text{جدول}} = 2.78 \qquad 5.77 \in w$$

آماره آزمون در ناحیه بحرانی قرار گرفته اســت پس فرضــیــه H_1 رد نمی شــود و $\beta \neq 0$ می باشد.



فاصله اطمینان $\boldsymbol{\mu_y}$ به ازای مقدار خاص $\boldsymbol{x}$

در مدل رگرسیون هدف اصلی از محاسبه خط رگرسیون برآورد y به ازای مقدار خاص مفروض x می باشد. زیرا مقدار $\hat{y}$ که ازخط رگرسیون نمونه بصورت زیر به دست می آید، برآوردی از میانگین حقیقی جامعه یعنی μ_y است.

$$\hat{y} = a + bx$$

$$\mu_y = \hat{y} \pm T_{n-2\,,1-\frac{\alpha}{2}}.S_e.\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{SS_x}}$$

در مسأله فوق یک فاصله اطمینان ۹۵٪ برای میانگین میزان کارایی کارکنان به ازای سابقه $\boldsymbol{x = 5}$ محاسبه کنید.

$$\hat{y} = a + bx = 1.25 + (1.25 * 5) = 7.5$$

$$T_{n-2\,,1-\frac{\alpha}{2}} = T_{4\,,0.975} = 2.78$$

$$\mu_y = \hat{y} \pm T_{n-2\,,1-\frac{\alpha}{2}}.S_e.\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{SS_x}} = 7.5 \pm 2.78 * 0.866.\sqrt{\frac{1}{6} + \frac{(5-3)^2}{16}} = \begin{cases} 5.95 \\ 9.05 \end{cases}$$

مفهوم آن این است که متوسط میزان کارآیی کارمندی که ۵ سال سابقه کار دارد بین ۵,۹۵ و ۹,۰۵ می باشد.

فاصله اطمینان y خاص به ازای مقدار خاص x

در این حالت به جای محاسبه فاصله اطمینان برای میانگین y به ازای یک مقدار خاص x می خواهیم فاصله اطمینان را برای یک مقدار خاص y به ازای یک مقدار خاص x محاسبه نماییم. این فاصله اطمینان به شرح ذیل است.

$$y = \hat{y} \pm T_{n-2\,,1-\frac{\alpha}{2}}.S_e.\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{SS_x}}$$

بطوریکه ملاحظه می شود، انحراف معیار این y خاص، بزرگ تر از حالتی است که انحراف معیار برای μ_y بکار می رود.

در مسأله فوق، فاصله اطمینان ۹۵٪ میزان کارآیی یک کارمند را به ازای سابقه $x = 5$ بدست آورید.

$$y = \hat{y} \pm T_{n-2\,,1-\frac{\alpha}{2}}.S_e.\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{SS_x}} = 7.5 \pm (2.78 * 0.866 * \sqrt{1+\frac{1}{6}+\frac{(5-3)^2}{16}} = \begin{cases} 4.64 \\ 10.36 \end{cases}$$

بطوریکه ملاحظه می شود، فاصله اطمینان برای y بزرگ تر از فاصله اطمینان برای μ_y می باشد. مفهوم آن این است که میزان کارآیی کارمندی که ۵ سال سابقه کار دارد بین ۴٫۶۴ و ۱۰٫۳۶ می باشد.

استفاده از تجزیه واریانس در رگرسیون

قبلاً نحوه آزمون فرضیه $H_\circ: \beta = 0$ را به کمک آماره t شرح دادیم. با استفاده از آماره F این آزمون را نیز می توان انجام داد. در رگرسیون خطی فرض بر این است که y تابعی از متغیر مستقل x می باشد. بنابراین در صورت وجود این رابطه خطی، می توان انتظار داشت که هر تغییری در y توسط متغیر x توجیه گردد. در تجزیه واریانس باید پراکندگی کل را به دو یا چند پراکندگی تجزیه نمود و در مدل رگرسیون نیز پراکندگی کل، متغیر وابسته y را به پراکندگی که توسط خط رگرسیون توجیه می گردد و به آن پراکندگی می گویند، انجام داد.

برای اینکه بهتر بتوانیم این آزمون را انجام دهیم، اول فرضیات آماری را می نویسیم.

$$\begin{cases} H_\circ: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

$$SST = SSR + SSe$$

$$SST = \sum (y_i - \overline{y})^2 \qquad or \qquad SST = \sum y_i^2 - \frac{1}{n}(\sum y_i)^2$$

$$SSR = \sum (\hat{y}_i - \overline{y})^2 = \frac{(SPxy)^2}{SSx}$$

$$SSe = SST - SSR$$

برای اجرای آزمون، فرضیه های زیر را می نویسیم:

$$\begin{cases} H_\circ: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

منبع تغییرات	DF	SS	MS	F
R	C-1=2-1=1	SSR	MSR=SSR	MSR/MSe
e	DFe=n-2=6-2=4	SSe	MSe=SSe/DFe	
کل	DF=n-1=6-1=5	SST		

F جدول را پیدا می کنیم و سپس F محاسبه شده را با آن مقایسه می کنیم و نتیجه گیری می نماییم.

در مسأله فوق آزمون زیر را کمک تجزیه واریانس انجام دهید.

$$\begin{cases} H_\circ: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

$$\hat{y} = a + bx = 1.25 + (1.25x)$$

$$SST = \sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2 = 178 - \frac{1}{6}(30)^2 = 28$$

$$SSR = \sum(\hat{y_i} - \overline{y})^2 = \frac{(SPxy)^2}{SSx} = \frac{(20)^2}{16} = 25$$

$$SSe = SST - SSR = 28 - 25 = 3$$

جدول تجزیه واریانس را تشکیل می دهیم.

منبع تغییرات	DF	SS	MS	F
R	C-1=2-1=1	25	25	33.33
e	DFe=n-2=6-2=4	3	0.75	
کل	DF=n-1=6-1=5	28		

مقدار F از جدول برای درجه آزادی $DF_R = n_1 = 1$ و $DF_e = n_2 = 4$ درسطح ۰,۰۵ و ۰,۰۱ برابر است با:

$$DF_R = n_1 = 1 \qquad \alpha = 0.05 \qquad 7.71$$

$$DF_e = n_2 = 4 \qquad \alpha = 0.01 \qquad 21.2$$

در سطح $\alpha = 0.01$ فرض $H_\circ$ رد می شود زیرا F محاسبه شده بزرگ تر از F ۰,۰۱ و ۰,۰۵ جدول است و همبستگی خطی برقرار است.

ضریب تشخیص یا ضریب تعیین

ضریب تشخیص نشان می دهد که چند درصد از تغییرات y ناشی از تغییرات x می باشد و چند درصد از تغییرات y مربوط به تغییرات x نیست و از فرمول زیر به دست می آید:

$$R^2 = 1 - \frac{SSe}{SSy} = \frac{SSR}{SSy}$$

$$R^2 = \frac{(SPxy)^2}{SSx.SSy} = r^2$$ ضریب همبستگی

در مسأله فوق، ضریب تشخیص را بدست آورده و آن را تفسیر نمایید.

$r = 0.945$

$R^2 = r^2 = (0.945)^2 = 0.893$

تفسیر آن این است که ۸۹٪ از تغییرات y ناشی از تغییرات x است و ۱۱٪ مربوط به سایر عوامل می باشد.

رابطه بین شیب خطوط رگرسیون y/x و x/y

معادله خط رگرسیون را می توان به دو صورت زیر بدست آورد:

1. y/x که در آن y تابع و x متغیر مستقل است.
2. x/y که در آن x تابع و y متغیر مستقل است.

$\overline{y} = \alpha + \beta x$

$\overline{x} = \alpha + \beta` y$

بین شیب این خطوط رگرسیون و ضریب همبستگی تساوی زیر برقرار است:

$r^2 = \beta\beta`$

هنگام جذر گرفتن باید توجه نمود که اگر شیب خطوط رگرسیون مثبت باشد برای r علامت (+) و اگر شیب خطوط رگرسیون منفی باشد برای r علامت (-) در نظر می گیریم.

$$\beta = \frac{Cov(x,y)}{\delta_x^2} \qquad \beta` = \frac{Cov(x,y)}{\delta_y^2}$$

$$\beta\beta` = \frac{Cov(x,y).Cov(x,y)}{\delta_x^2\delta_y^2} = \left(\frac{Cov(x,y)}{\delta_x\delta_y}\right)^2 = r^2$$

در جامعه ای که توزیع مشترک چگالی احتمال های دو متغیر تصادفی x و y بر طبق قانون نرمال توزیع می شود. دو معادله همبستگی در نمونه ای به حجم $n = 10$ و $\overline{y}_x = 13.6 - 1.9x$ و $\overline{x}_y = 7 - 0.5y$ می باشد. ضریب همبستگی خطی را تخمین بزنید.

$\overline{y}_x = 13.6 - 1.9x$

$\overline{x}_y = 7 - 0.5y$

$r^2 = \beta\beta` = (-1.9)(-0.5) = 0.95$

$r = -\sqrt{0.95} = -0.975$

شیب خط رگرسیون منفی است. چون هر دو شیب منفی است پس $r = -0.975$ خواهد بود.

جلسه ششم

تجزیه واریانس رگرسیون چندگانه

تجزیه واریانس رگرسیون چندگانه

به منظور بررسی معنی دار بودن رگرسیون چندگانه می توان از تجزیه واریانس استفاده نمود و فرضیه زیر را آزمون نمود.

$H_\circ: \beta_1 = \beta_2 = \cdots = \beta_K$

یعنی تمام متغیرهای مستقل در تعیین تغییرات مقادیر y تأثیری ندارند.

$H_1:$ حداقل 2 تا از بتاها با هم متفاوت هستند

برای اینکار، جدول تجزیه واریانس ANOVA به ترتیب زیر تشکیل می شود.

منبع تغییرات	DF	SS	MS	F
رگرسیون R	K	SSR	MSR=SSR/DFR	F=MSR/MSe
اشتباه e	n-k-1	SSe	MSe=SSe/DFe	
کل T	n-1	SST		

مقدار بحرانی F هم در سطح α و با در نظر گرفتن درجات آزادی صورت و مخرج از جدول F استخراج و نتیجه گیری صورت می گیرد.

ه سأله) به منظور برر سی رابطه بین درآمد (y) و سابقه کار (x_1) و تح صیلات دان شگاهی (x_2) کارکنان از نمونه تصادفی ۵ تایی نتایج زیر حاصل شده است:

x_1	x_2	y
1	2	4
2	2	5
2	3	6
3	4	6
4	4	8

الف) معادله رگرسیون خطی بین درآمد و سابقه کار و تحصیلات دانشگاهی را بدست آورید و درآمد یک کارمند با سابقه$x_1 = 5$ وتحصیلات دانشگاهی $x_2 = 4$ را بدست آورید.

ب) ضریب تشخیص و ضریب همبستگی چندگانه را حساب کنید.

ج) با استفاده از جدول تجزیه واریانس معادله رگرسیون را ارزشیابی نمایید.

حل مسأله)

برای حل مسأله، جدول محاسباتی را تشکیل می دهیم:

x_1	x_2	y	x_1^2	x_2^2	y^2	x_1x_2	x_1y	x_2y
1	2	4	1	4	16	2	4	8
2	2	5	4	4	25	4	10	10
2	3	6	4	9	36	6	12	18
3	4	6	9	16	36	12	18	24
4	4	8	16	16	64	16	32	32
$\sum x_1 = 12$	$\sum x_2 = 15$	$\sum y = 29$	$\sum x_1^2 = 34$	$\sum x_2^2 = 49$	$\sum y^2 = 177$	$\sum x_1x_2 = 40$	$\sum x_1y = 76$	$\sum x_2y = 92$

برای محاسبه صفحه رگرسیون از فرمول های زیر استفاده می کنیم:

$$na + b_1\sum x_1 + b_2\sum x_2 = \sum y$$

❶ $5a + 12b_1 + 15b_2 = 29$

$$a\sum x_1 + b_1\sum {x_1}^2 + b_2\sum x_1x_2 = \sum x_1y$$

❷ $12a + 34b_1 + 40b_2 = 76$

$$a\sum x_2 + b_1\sum x_1x_2 + b_2\sum {x_2}^2 = \sum x_2y$$

❸ $15a + 40b_1 + 49b_2 = 92$

$$\begin{bmatrix} 5 & 12 & 15 & 29 \\ 12 & 34 & 40 & 76 \\ 15 & 40 & 49 & 92 \end{bmatrix} \rightarrow \begin{bmatrix} 5 & 12 & 15 & 29 \\ (-2.4*5)+12=0 & (-2.4*12)+34=5.2 & (-2.4*15)+40=4 & (-2.4*29)+76=6.4 \\ 15 & 40 & 49 & 92 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 5 & 12 & 15 & 29 \\ 0 & 5.2 & 4 & 6.4 \\ 15 & 40 & 49 & 92 \end{bmatrix} \rightarrow \begin{bmatrix} 5 & 12 & 15 & 29 \\ 0 & 5.2 & 4 & 6.4 \\ (5)(-3)+15=0 & (12)(-3)+40=4 & (15)(-3)+49=4 & (29)(-3)+92=5 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 5 & 12 & 15 & 29 \\ 0 & 5.2 & 4 & 6.4 \\ 0 & 4 & 4 & 5 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 5 & 12 & 15 & 29 \\ 0 & 5.2 & 4 & 6.4 \\ (0)\left(-\frac{10}{13}\right)+0=0 & \left(\frac{26}{5}\right)\left(-\frac{10}{13}\right)+4=0 & (4)\left(-\frac{10}{13}\right)+4=\frac{12}{13} & \left(\frac{32}{5}\right)\left(-\frac{10}{13}\right)+5=\frac{1}{13} \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 5 & 12 & 15 & 29 \\ 0 & \frac{26}{5} & 4 & \frac{32}{5} \\ 0 & 0 & \frac{12}{13} & \frac{1}{13} \end{bmatrix}$$

$$5a + 12b_1 + 15b_2 = 29 \rightarrow 5a + 12\left(\frac{7}{6}\right) + 15\left(\frac{1}{12}\right) = 29 \rightarrow 5a + \frac{84}{6} + \frac{15}{12} = 29 \rightarrow 5a = 13.75 \rightarrow a = 2.75 = \frac{11}{4}$$

$$0a + \frac{26}{5}b_1 + 4b_2 = \frac{32}{5} \rightarrow \frac{26}{5}b_1 + \left(4*\frac{1}{12}\right) = \frac{32}{5} \rightarrow \frac{26}{5}b_1 = \frac{32}{5} - \frac{1}{3} = \frac{91}{15} \rightarrow b_1 = \frac{\frac{91}{15}}{\frac{26}{5}} = \frac{91}{78} = \frac{7}{6}$$

$$0a + 0b_1 + \frac{12}{13}b_2 = \frac{1}{13} \rightarrow b_2 = \frac{\frac{1}{13}}{\frac{12}{13}} = \frac{1}{12}$$

از حل این دستگاه داریم:

$$a = \frac{11}{4}\,, \qquad b_2 = \frac{1}{12}\,, \qquad b_1 = \frac{7}{6}$$

الف) معادله صفحه رگرسیون را می نویسیم.

$$y = a + b_1x_1 + b_2x_2 = \frac{11}{4} + \frac{7}{6}x_1 + \frac{1}{12}x_2$$

$$\begin{cases} x_1 = 5 \\ x_2 = 4 \end{cases} \qquad y = \frac{11}{4} + \frac{7}{6}x_1 + \frac{1}{12}x_2 = \frac{11}{4} + \left(\frac{7}{6} * 5\right) + \left(\frac{1}{12} * 4\right) = 8.92$$

$$R^2 = \frac{\sum(\hat{y} - \overline{y})^2}{\sum(y - \overline{y})^2}$$

$$\sum(\hat{y} - \overline{y})^2 = b_1\left[\sum x_1y - \frac{1}{n}(\sum x_1)(\sum y)\right] + b_2\left[\sum x_2y - \frac{1}{n}(\sum x_2)(\sum y)\right]$$

$$\sum(\hat{y} - \overline{y})^2 = \frac{7}{6}\left[76 - \frac{1}{5}(12)(29)\right] + \frac{1}{12}\left[92 - \frac{1}{5}(15)(29)\right] = 7.88$$

$$\sum(y - \overline{y})^2 = \sum {y_i}^2 - \frac{1}{n}(\sum y_i)^2 = 177 - \frac{1}{5}(29)^2 = 8.8$$

$$R^2 = \frac{\sum(\hat{y} - \overline{y})^2}{\sum(y - \overline{y})^2} = \frac{7.88}{8.8} = 0.9 \qquad R = \sqrt{\frac{\sum(\hat{y} - \overline{y})^2}{\sum(y - \overline{y})^2}} = \sqrt{0.9} = 0.95$$

ج)

$H_0: \beta_1 = \beta_2 = \cdots = \beta_K$

H_1: حداقل 2 تا از بتاها با هم متفاوت هستند

$$SST = \sum(y - \overline{y})^2 = 8.8 = SSy$$

$$SSR = \sum(\hat{y} - \overline{y})^2 = 7.88$$

$$SSe = SST - SSR = 8.8 - 7.88 = 0.92$$

منبع تغییرات	DF	SS	MS	F
R	K=3-1=2	SSR=7.88	MSR=SSR/DFR	F=MSR/MSe 3.94/0.46=8.57
e	DFe=n-k-1=2	SSe=0.92	MSe=SSe/DFe	
کل	DF=n-1=5-1=4	SST= 8.8		

چون F محاسبه شده ۸,۵۷ کوچکتر از $F0.05$ و $F0.01$ است پس فرضیه H_0 رد نمی شود.

$F_{0.05,\,2,\,2}=19$, $F_{0.01,\,2,\,2}=99$

جلسه‌ی هفتم

تجزیه و تحلیل واریانس فریدمن

مسأله) از ۹ نفر کارشناس مدیریت خواسته شده است در مورد ❸ طرح طبقه بندی مشاغل اظهار نظر و آن ها را به ترتیب مزیتشان، رتبه بندی نمایند. در این آزمون فرضیه صفر دلالت بر عدم مزیت بین ❸ طرح را دارد و فرضیه مخالف نشانگر این است که مزیت ❸ طرح با هم تفاوت معنادار دارد.

به کمک آزمون فریدمن، تشخیص دهید که آیا ستون های رتبه ها از یک جمله انتخاب شده اند یا خیر؟ اظهار نظر و رتبه بندی کارشناسان مذکور در جدول زیر منعکس است.

کارشناسان	I	II	III
1	2	3	1
2	2	3	1
3	2	3	1
4	1	3	2
5	3	2	1
6	1	2	3
7	2	3	1
8	1	3	2
9	1	3	2
	$R_1 = 15$	$R_2 = 25$	$R_3 = 14$

برای حل:

توجه ❶ همانطور که گاهی به روش ناپارامتری مشابه با تجزیه و تحلیل واریانس یک طرفه نیاز داریم، در مواردی لازم است داده های آماری را در رده بندی دو طرفه که مشابه تجزیه و تحلیل واریانس دو طرفه است، تجزیه و تحلیل نمایید.

توجه ❷ معمولاً هنگامی از این روش ناپارامتری استفاده می شود که به علت ضعف مقیاس اندازه گیری، شرایط لازم برای تجزیه و تحلیل پارامتری وجود نداشته باشد و یا بخواهیم با سرعت و عجله نتایج را بدست آوریم.

توجه ❸ آزمونی که تحت این شرایط بکار می رود، تجزیه و تحلیل واریانس به کمک رتبه هاست. این آزمون، هنگامی به کار می رود که داده های آماری حداقل دارای مقیاس ترتیبی باشند و بتوان با مفهوم ترتیبی، آن ها را در رده بندی دو طرفه مرتب نمود.

توجه ❹ در مسأله بالا نحوه آزمون فریدمن شرح داده می شود.

توجه ❺ اگر فرضیه صفر درست باشد، انتظار داریم که رتبه های ① و ② و ③ در هر ستون، تقریباً با فراوانی یکسان رخ دهد.

توجه ❻ اگر فرضیه صفر نادرست باشد یعنی طرح های مشاغل یکسان نباشند، انتظار داریم که رتبه های حداقل یک ستون با بقیه ستون ها تفاوت داشته باشند.

توجه ❼ در آزمون فریدمن بررسی می شود که اگر فرضیه صفر درست باشد، آیا مجموع های مشاهده شده رتبه های ستون ها آنقدر تفاوت متمایز دارند که ناشی از تصادف نباشد.

توجه ❽ ابتدا رتبه های هر ستون را جمع کرده و مجموع هر ستون را با Rj نشان می دهیم.

آماره آزمون فریدمن به شرح ذیل است:

$$\chi_2{}^2 = \frac{12}{nk(k+1)} \sum_{j=1}^{n} R_j{}^2 - 3n(k+1)$$

$$\chi_2{}^2 = \frac{12}{9 * 3(4)} [(14)^2 + (25)^2 + (15)^2] - 3(9)(4) = 8.22$$

جدول $\chi^2_{\alpha,k-1} = \chi^2_{0.05,3-1} = 5.99147$

چون $8.222 > 5.99147$ است فرض H_0 رد می شود و 3 طرح طبقه بندی مشاغل از نظر مزیت یکسان نیستند.