حل تمرین های تحلیل آماری

سوال ۱) مدیر کنترل کیفیت کارخانه ای هر روز $n = 3$ دستگاه اتومبیل را که از خط مونتاژ خارج می شود به عنوان نمونه انتخاب و تعداد تنظیم های مورد نیاز روزانه را در ۱۰۰ روز کاری به شرح زیر جمع آوری کرده است.

۳	۲	۱	۰	تعداد تنظیم ها
۴	۱۰	۲۸	۵۸	تعداد روزها

در سطح معنی دار بودن $\alpha = 0.05$ آزمون نمائید که توزیع دو جمله ای با پارامترهای $n = 3$ و $p = 0.2$ یک مدل منا سب برای توزیع فوق می با شد یا خیر؟ این آزمون را با روش های زیر انجام دهید و نتایج را با هم مقایسه کنید:

الف) آزمون کای دو x^2

ب) آزمون کولموگروف – اسمیرنوف KS

حل مسأله :

الف) آزمون کای دو x^2

تعداد تنظیم ها	تعداد روزها F_o	$P(x)$	$F_e = np$	$(F_o - F_e)^2$	$\frac{(F_o - F_e)^2}{F_e}$
0	58	0.512	51.2	$(58 - 51.2)^2 = 46.24$	$\frac{46.24}{51.2} = 0.9$
1	28	0.384	38.4	$(28 - 38.4)^2 = 108.16$	$\frac{108.16}{38.4} = 2.81$
2	10	0.096	9.6	$(10 - 9.6)^2 = 0.16$	$\frac{0.16}{9.6} = 0.016$
3	4	0.008	0.8	$(4 - 0.8)^2 = 10.24$	$\frac{10.24}{0.8} = 12.8$
	$\sum F_o = 100$		$\sum F_e = 100$		$\sum x^2 = 16.526$

$$P(x) = C_n^x p^x q^{n-x}$$

$$P(0) = C_3^0 (0.2)^0 (0.8)^{3-0} = \frac{3!}{3!\,0!}(1)(0.512) = 0.512$$

$$P(1) = C_3^1 (0.2)^1 (0.8)^{3-1} = \frac{3!}{2!\,1!}(0.2)(0.64) = 0.384$$

$$P(2) = C_3^2 (0.2)^2 (0.8)^{3-2} = \frac{3!}{1!\,2!}(0.04)(0.8) = 0.096$$

$$P(3) = C_3^3 (0.2)^3 (0.8)^{3-3} = \frac{3!}{0!\,3!}(0.008)(1) = 0.008$$

$v = R - 1 = 4 - 1 = 3$

$x^2\ (3, 0.05) = 7.81$

$$\Longrightarrow \text{محاسبه شده } x^2 < x^2 \text{ جدول}$$

چون آماره آزمون در ناحیه بحرانی قرار می گیرد، بنابراین با احتمال ٪۹۵ فرضیه H_0 رد می شود؛ یعنی توزیع دو جمله ای با پارامترهای $n = 3$ و $p = 0.2$ یک مدل مناسب برای توزیع فوق نمی باشد.

ب) آزمون کولموگروف – اسمیرنوف KS

$$\begin{cases} H_0: F(x) = F_e(x) & \text{برای تمام مقادیر} \\ H_1: F(x) \neq F_e(x) & \text{برای حداقل یک مقدار} \end{cases}$$

X	فراوانی مطلق	فراوانی تجمعی مطلق	فروانی تجمعی نسبی	توزیع دو جمله ای	$F_e(x)$	$F(x)$	$\lvert F(x) - F_e(x)\rvert$
0	58	58	0.58	0.512	0.512	0.58	MAX=0.068
1	28	86	0.86	0.384	0.896	0.86	0.036
2	10	96	0.96	0.096	0.992	0.96	0.032
3	4	100	1	0.008	1	1	0.000

$MAX\ \lvert F(x) - F_e(x)\rvert = 0.068$

$D_{\alpha, n} = D_{0.05, 100} = 0.136$

چون آماره آزمون در ناحیه بحرانی قرار نمی گیرد، بنابراین با احتمال ٪۹۵ فرضیه H_0 رد نمی شود؛ یعنی توزیع دو جمله ای با پارامترهای $n = 3$ و $p = 0.2$ یک مدل مناسب برای توزیع فوق می باشد.

سوال ۲) سازمان مدیریت و برنامه ریزی کشور می خواهد نسبت مدیران سه وزارتخانه را که رشته تحصیلی آنان با نوع وظایف سازمان مورد تصدیشان مطابقت دارد مقایسه نماید. برای این منظور از وزارتخانه های مذکور سه نمونه تصادفی به حجم های ۵۴، ۶۰ و ۶۶ مدیر را انتخاب و پس از بررسی معلوم گردید که به ترتیب ۴، ۱۲ و ۱۴ نفر تحصیلاتشان با نوع کار ارجائی آنان مطابقت دارد. در سطح $\alpha = 0.05$ یکسان بودن نسبت مدیران متخصص سه وزارتخانه را آزمون نمائید.

حل مسأله :

$$\begin{cases} H_0: \boldsymbol{n_1} = \boldsymbol{n_2} = \boldsymbol{n_3} \\ H_1: \text{حداقل دو تا از نسبت ها با هم اختلاف دارند} \end{cases}$$

وزارتخانه				
وضعیت مطابقت تحصیل با نوع کار ارجائی	①	②	③	جمع
مطابقت دارد	4	12	14	30
مطابقت ندارد	50	48	52	150
جمع	54	60	66	180

وزارتخانه				
وضعیت مطابقت تحصیل با نوع کار ارجائی	①	②	③	جمع
مطابقت دارد	$\frac{30}{180} * 54 = 9$	$\frac{30}{180} * 60 = 10$	$\frac{30}{180} * 66 = 11$	30
مطابقت ندارد	$\frac{150}{180} * 54 = 45$	$\frac{150}{180} * 60 = 50$	$\frac{150}{180} * 66 = 55$	150
جمع	54	60	66	180

F_o	F_e	$(F_o - F_e)$	$(F_o - F_e)^2$	$\frac{(F_o - F_e)^2}{F_e}$
4	9	5	25	$\frac{25}{9} = 2.78$
12	10	2	4	$\frac{4}{10} = 0.4$
14	11	3	9	$\frac{9}{11} = 0.82$
50	45	5	25	$\frac{25}{45} = 0.56$
48	50	2	4	$\frac{4}{50} = 0.08$
52	55	3	9	$\frac{9}{55} = 0.16$
$\sum F_o = 180$	$\sum F_e = 180$			$\sum x^2 = 4.8$

$v = (R-1)(C-1) = 1 * 2 = 2$

$x^2\ (2, 0.05) = 5.99$

$$x^2_{\text{جدول}} > x^2_{\text{محاسبه شده}} \Longrightarrow$$

چون آماره آزمون در ناحیه بحرانی قرار نمی گیرد، بنابراین با احتمال ۹۵٪ فر ضیه H_0 رد نمی شود؛ و فر ضیه یک سان بودن نسبت مدیران متخصص سه وزارتخانه تأیید می گردد.

سوال ۳) شرکت تولید کننده پودر لباس شویی ادعا دارد که وزن خالص پودر یک نوع بسته بندی 480g می باشد. در ۲۵ جعبه که بصورت تصادفی نمونه گیری شد، میانگین وزن 468g با انحراف معیار 12g بدست آمد. با توجه به اطلاعات فوق در سطح تشخیص $\alpha = 0.05$ آیا می توان نتیجه گرفت که وزن خالص پودر در این نوع بسته بندی 480g ادعایی نیست؟ (توزیع جامعه نرمال نیست)

$H_0: \mu = 480 = \mu_0$

$H_1: \mu \neq 480 \neq \mu_0$

$n = 25 \qquad \bar{x} = 468 \qquad S = 12$

چون جامعه نرمال نیست و حجم نمونه کوچک است از " چه بی شف " برای اجرای آزمون استفاده می کنیم.

$$K = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{468 - 480}{\frac{12}{\sqrt{25}}} = -5$$

$$K = \sqrt{\frac{1}{\alpha}} = \sqrt{\frac{1}{0.05}} = \pm 4.47$$

در " چه بی شف " به جدول نیاز نداریم.

فرض H_0 تأیید نمی شود. یعنی وزن خالص پودر در یک نوع بسته بندی 480g نیست.

سوال ۴) به منظور بررسی همبستگی بین تعداد کارشناسان اداری (x) و میزان کارائی (y) موسسات دولتی، نمونه ای تصادفی به حجم n=10 موسسه انتخاب و نتایج زیر حاصل گردید:

$$\boldsymbol{Cov\,(x,y) = 12 \qquad \sum x_i = \sum y_i = 50 \quad S_x = 4 \quad S_y = 5}$$

ضریب همبستگی و معادله خط رگرسیون y روی x را در این نمونه حساب کنید و با استفاده از آزمون t معنی دار بودن ضریب همبستگی را در سطح $\boldsymbol{\alpha = 0.05}$ آزمون نمایید.

الف) حل ضریب همبستگی

$$\mathrm{r_{X,Y}} = \frac{\mathrm{Cov\,(x,y)}}{\mathrm{S_x S_y}} = \frac{12}{4*5} = \frac{12}{20} = 0.6$$

ب) معادله خط رگرسیون y روی x

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

$$\bar{\mathrm{x}} = \bar{\mathrm{y}} = 5$$

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

$$\beta = \frac{\mathrm{Cov\,(x,y)}}{Sx^2} = \frac{12}{16} = 0.75$$

$$\bar{y} = \alpha + \beta\bar{x} \quad \Rightarrow 5 = \alpha + (0.75)(5) \;\Rightarrow\; \alpha = 5 - 3.75 = 1.25 \quad \Rightarrow \quad \alpha = 1.25$$

$$y = 1.25 + 0.75x$$

ج) بررسی معنی دار بودن ضریب همبستگی با استفاده از آزمون t

$$H_0: \rho = \rho_0$$

$$H_1: \rho \neq \rho_0$$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.6}{\sqrt{\frac{1-(0.6)^2}{10-2}}} = \frac{0.6}{\sqrt{\frac{0.64}{8}}} = \frac{0.6}{\sqrt{0.08}} = \frac{0.6}{0.28} = 2.14$$

$$\begin{cases} \alpha = 0.05 \quad \Rightarrow \; 2.262 \\ \alpha = 0.01 \;\Rightarrow\; 3.250 \\ df = n - 1 = 9 \end{cases}$$

آماره آزمون در ناحیه بحرانی قرار نمی گیرد و فرض H_0 تأیید می گردد.

سوال ۵) فرضیه ای بصورت زیر تدوین شده است:

رضایت شغلی در سازمان (الف) بیشتر از سازمان (ب) است. به منظور بررسی فرضیه فوق، یک نمونه تصادفی ۱۰۰ نفره از سازمان (الف) انتخاب شده است که ۳۰ در صد آن ها از کار خود راضی هستند. در حالیکه فقط ۴۰ نفر از یک نمونه تصادفی ۲۰۰ نفره از سازمان (ب) از کار خود راضی هستند. در سطح خطای 0.05 فرضیه را آزمون کنید.

مقدار بحرانی z=1.645

$H_0: P_1 \leq P_2$

$H_1: P_1 > P_2$

$n_1 = 100 \qquad x_1 = 30 \qquad P_1 = \frac{x_1}{n_1} = \frac{30}{100} = 0.3$

$n_2 = 200 \qquad x_2 = 40 \qquad P_2 = \frac{x_2}{n_2} = \frac{40}{200} = 0.2$

$\overline{P} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{30 + 40}{100 + 200} = \frac{70}{300} = 0.233$

$$Q_c = \frac{P_1 - P_2}{\sqrt{\overline{P}(1-\overline{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.3 - 0.2}{\sqrt{(0.233)(0.767)\left(\frac{1}{100} + \frac{1}{200}\right)}} = \frac{0.1}{\sqrt{(0.1787)(0.015)}} = \frac{0.1}{\sqrt{0.00268}} = \frac{0.1}{0.052} = 1.92$$

$W\left(Z < Z_{1-\propto,\infty}\right)$

$Z < Z_{1-\propto,\infty} = Z_{1-0.05,\infty} = Z_{0.95,\infty} = 1.645$

فرض H_0 رد می شود یعنی می گوییم رضایت شغلی در سازمان (الف) بیشتر از سازمان (ب) است.

سوال ۶) اندک زمانی بعد از آنکه رئیس جمهور x مشغول کار شد، ۱۶۰ نفر از یک نمونه ۲۰۰ نفری از ساکنان کشور او را تأیید کردند. چهار سال بعد با رشد نارضایتی از سیاست های او فقط ۱۱۰ نفر از یک نمونه ۲۰۰ نفری او را تأیید کردند. یک فاصله اطمینان برای تغییر عقیده افراد کشور، بین دو دوره را با اطمینان ۹۵٪ تعیین کنید.

$$n_1 = 200 \qquad x_1 = 160 \qquad P_1 = \frac{x_1}{n_1} = \frac{160}{200} = 0.8$$

$$n_2 = 200 \qquad x_2 = 110 \qquad P_2 = \frac{x_2}{n_2} = \frac{110}{200} = 0.55$$

$$P^* = 0.95$$

$$F_{(z)} = 1 - \frac{\alpha}{2} = 1 - \frac{0.05}{2} = 1 - 0.025 = 0.975 \quad \Rightarrow \quad z = 1.96$$

$$\delta_{P_1 - P_2} = \sqrt{\frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2}} = \sqrt{\frac{0.8(1 - 0.8)}{200} + \frac{0.55(1 - 0.55)}{200}} = \sqrt{0.002} = 0.045$$

$$e = Z_{\frac{\alpha}{2}} * \delta_{P_1 - P_2} = 1.96 * .045 = 0.088$$

$$\left(\overline{P_1} - \overline{P_2} - e\,, \quad \overline{P_1} - \overline{P_2} + e \right) = (0.25 - 0.088\,, 0.25 + 0.088) = (0.162\,, 0.338)$$

سوال ۷) در یک مو سسه تولیدی، سه خط تولید A، B و C مصنوعات پلا ستیکی تولید می کنند. ۲۰٪ تولیدات خط A، ۱۰٪ تولیدات خط B و ۵٪ تولیدات خط C معیوب هستند. تا سی را پرتاب می کنیم. اگر خال یک رو شد، یکی از تولیدات A، اگر ۲ یا ۳ رو شد، یکی از تولیدات B و اگر ۴، ۵ و ۶ رو شد، یک از تولیدات C را بر می داریم:

الف) احتمال اینکه مصنوع برداشته شده، معیوب باشد چقدر است؟

ب) اگر مصنوع برداشته شده خراب باشد، احتمال اینکه این موضوع از خط تولید A برداشته شده باشد، چیست؟

الف)

احتمال خرابی محصول A : 0.2

احتمال خرابی محصول B : 0.1

احتمال خرابی محصول C : 0.05

و اگر انتخاب محصول را مشروط به پرتاب تاس کنیم:

۱) احتمال اینکه محصول A انتخاب شود به شرط اینکه تاس ۱ بیاید برابر است با:

$$\frac{1}{6} \times 0.2 = \frac{2}{60}$$

۲) احتمال اینکه محصول B انتخاب شود به شرط اینکه تاس ۲ و۳ بیاید برابر است با:

$$\frac{2}{6} \times 0.1 = \frac{2}{60}$$

۳) احتمال اینکه محصول C انتخاب شود به شرط اینکه تاس ۴ و۵ و۶ بیاید برابر است با:

$$\frac{3}{6} \times 0.05 = \frac{15}{600}$$

در نتیجه احتمال اینکه مصنوع برداشته شده معیوب باشد برابر است با:

$$(\frac{1}{6} \times 0.2) + (\frac{2}{6} \times 0.1) + (\frac{3}{6} \times 0.05) = \frac{55}{600}$$

ب)

طبق قضیه بیز داریم:

$$\frac{\frac{1}{6} \times 0.2}{\frac{1}{6} \times 0.2 + \frac{2}{6} \times 0.1 + \frac{3}{6} \times 0.05} = \frac{4}{11}$$

یعنی احتمال اینکه مصنوع برداشته شده خراب و از خط تولید A باشد برابر $\frac{4}{11}$ است.

سوال ۸) ثابت کنید:

1. $\mu = np$
2. $\delta^2 = npq$
3. $\delta = \sqrt{npq}$
4. $\lim\limits_{n\to\infty} C_n^x p^x q^{n-x}$

می دانیم:

احتمال عدم وقوع $= q$　　احتمال وقوع $= p$　　انحراف معیار $= \delta$　　واریانس $= \delta^2$　　میانگین $= \mu$

اثبات قسمت ۱)

همانطور که می دانید، میانگین همان امید ریاضی است. برای اثبات مقدر میانگین، از تعریف امید ریاضی استفاده می کنیم. توجه داریم که چون با یک توزیع گسسته سروکار داریم، از رابطه زیر استفاده می کنیم.

$$E(x) = \sum_{i=1}^{n} x_i . p(x_i)$$

در نتیجه داریم:

$$E(x) = \sum x_i . p = np$$

اثبات قسمت ۲)

$$\delta^2 = Var(x) = E(x^2) - E(x) = n^2p^2 + np - np^2 - n^2p^2 = np - np^2 = np(1-p) = npq$$

اثبات قسمت ۳)

انحراف معیار برابر است با جذر واریانس، لذا داریم:

$$\delta = \sqrt{npq}$$

اثبات قسمت ۴)

چون قضیه بسط دوجمله ای تضمین می کند که p یک تابع جرم احتمال است داریم:

$$\lim_{n\to\infty} C_n^x p^x q^{n-x} = \sum \binom{n}{x} C_n^x p^x (1-p)^{n-x} = (p + (1-p))^n = 1^n = 1$$