1. **برای مطالعه رابطه میان مدت زمان یک سرویس (تعمیر) و تعداد قطعات الکترونیکی که باید تعمیر شوند یا تعویض شوند، نمونه ای از رکوردهای مربوط به سرویسها انتخاب شده و در جدول زیر نشان داده شده است. این داده ها شامل طول زمان سرویس به دقیقه (متغیر پاسخ) و تعداد قطعات تعمیر شده (متغیر پیشگو) هستند.**

میزان کواریانس و همبستگی دو متغیر را محاسبه نمایید.

|  |  |  |
| --- | --- | --- |
| ردیف | تعداد قطعات | تعداد دقیقه ها |
| 1 | 1 | 23 |
| 2 | 2 | 29 |
| 3 | 3 | 49 |
| 4 | 4 | 64 |
| 5 | 4 | 74 |
| 6 | 5 | 87 |
| 7 | 6 | 96 |
| 8 | 6 | 97 |
| 9 | 7 | 109 |
| 10 | 8 | 119 |
| 11 | 9 | 149 |
| 12 | 9 | 145 |
| 13 | 10 | 154 |
| 14 | 10 | 166 |
| جدول 1  رابطه میان تعداد قطعات و زمان تعمیر | | |

ج) ابتدا کمیتهایی را که برای محاسبه همبستگی و کواریانس مورد نیاز هستند، محاسبه می شود و در جدول زیر نشان داده می‌شود:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *i* |  |  | (-) | (-) |  |  | (-)(-) |
| 1 | 23 | 1 | -74.21 | -5 | 5507.76 | 25 | 371.07 |
| 2 | 29 | 2 | -68.21 | -4 | 4653.19 | 16 | 272.86 |
| 3 | 49 | 3 | -48.21 | -3 | 2324.62 | 9 | 144.64 |
| 4 | 64 | 4 | -33.21 | -2 | 1103.19 | 4 | 66.43 |
| 5 | 74 | 4 | -23.211 | -2 | 538.90 | 4 | 46.43 |
| 6 | 87 | 5 | -10.21 | -1 | 104.33 | 1 | 10.21 |
| 7 | 96 | 6 | -1.21 | 0 | 1.47 | 0 | 0 |
| 8 | 97 | 6 | -0.21 | 0 | 0.05 | 0 | 0 |
| 9 | 109 | 7 | 11.79 | 1 | 138.90 | 1 | 11.79 |
| 10 | 119 | 8 | 21.79 | 2 | 474.62 | 4 | 43.57 |
| 11 | 149 | 9 | 51.79 | 3 | 2681.76 | 9 | 155.36 |
| 12 | 145 | 9 | 47.79 | 3 | 2283.47 | 9 | 143.36 |
| 13 | 154 | 10 | 56.79 | 4 | 3294.62 | 16 | 227.14 |
| 14 | 166 | 10 | 68.79 | 4 | 4731.47 | 16 | 275.14 |
| کل | 1361 | 84 | 0 | 0 | 27768.36 | 114 | 1768.00 |

جدول 2- محاسبات جزئی برای محاسبه کواریانس و همبستگی





Cov(Y,X)=

Cor(Y,X)=

میزان بالای ضریب همبستگی نشانگر رابطه خطی قوی بین دو متغیر X و Y است که با مشاهدات جدول داده ها سازگار است. بنابراین رابطه قوی مثبتی میان زمان تعمیر و تعداد قطعات تعمیر شده وجود دارد.

1. **رابطه میان متغیر پاسخ (زمان بر حسب دقیقه) و متغیر پیشگو (تعداد قطعات تعمیر شده) را با استفاده رگرسیون خطی توصیف نمایید.**

**ج) همانگونه که در سوال 1 نشان داده شد، تنها یک متغیر پیشگو و یک متغیر پاسخ وجود دارد. بنابراین، مدل رگرسیون از نوع رگرسیون ساده است.**

****

**بنابراین می خواهیم ضرایب معادله زیر را برآورد نماییم:**

**=دقیقه ها**

**به منظور برآورد پارامترها از روش کمترین مربعات استفاده می کنیم که به سبب ساده بودن معادله رگرسیونی به سادگی قابل محاسبه می باشد**. هریک از رکوردهای مربوط به زمان تعمیر و تعداد قطعات تعمیر شده در جدول 1 در فضای دو بعدی زیر نشان داده شده اند:

**شکل 1- نمودار تعداد قطعات بر حسب زمان مربوط به داده های جدول 1**

قطعات

دقیقه ها

میخواهیم برمبنای داده های موجود در جدول 1 پارامترهای را برآورد کنیم. به عبارت دیگر، می خواهیم خط راستی را پیدا کنیم که بهترین برازش را به نقاط نمودار پراکندگی پاسخ در مقابل پیشگو در شکل 1می دهد. پارامترها را با استفاده از روش کمترین توانهای دوم برآورد می کنیم. این روش خطی را می دهد که مجموعه توانهای دوم فواصل عمودی از هر نقطه تا خط را کمینه می سازد. فواصل عمودی نمایانگر خطاهای در تغییر پاسخ هستند. این خطاها را می توان با نوشتن به صورت زیر به دست آورد:

, *i=1,2,..,n*

مجموع توانهای دوم این فواصل به صورت زیر نوشته می شوند:



با در نظر گرفتن معادله ساده خط در فضای دو بعدی براساس شیب خط و عرض از مبدا به سادگی می توان خطی را یافت که کمترین مجموع توانهای دوم را می دهد. تمرین)نحوه برازش این خط را نشان دهید.

پس:



با قرار دادن داده های موجود در جدول 1 در معادله بالا ضرایب محاسبه می شوند:



در نتیجه معادله رگرسیونی تعمیر قطعات به شکل زیر خواهد بود.

15.509\*قطعات+4.162=دقیقه ها

3)در معادله رگرسیونی شکل 2، میزان خطاهای جزئی را محاسبه کرده و رابطه میان برازش پارامترها با ضرایب همبستگی و کواریانس را بیابید.

ج)به سادگی می توان با قرار دادن مقادیر موجود در جدول 1 در معادله رگرسیونی، میزان متغیر پاسخ برآورد شده و خطای مربوطه را محاسبه نمود. نتایج این محاسبه در جدول 3 نشان داده شده است.

|  |  |  |  |
| --- | --- | --- | --- |
| *i* | قطعات |  |  |
| 1 | 1 | 19.67 | 3.33 |
| 2 | 2 | 35.18 | 6.18- |
| 3 | 3 | 50.69 | 1.69- |
| 4 | 4 | 66.20 | 2.20- |
| 5 | 4 | 66.20 | 7.80 |
| 6 | 5 | 81.71 | 5.29 |
| 7 | 6 | 97.21 | 1.21- |
| 8 | 6 | 97.21 | 0.21- |
| 9 | 7 | 112.72 | 3.72- |
| 10 | 8 | 128.23 | 9.23- |
| 11 | 9 | 143.74 | 5.26 |
| 12 | 9 | 143.74 | 1.26 |
| 13 | 10 | 159.25 | 5.25- |
| 14 | 10 | 159.25 | 6.75 |

ارتباط مقادیر کواریانس و همبستگی با برآورد پارامترهای معادله رگرسیونی به شکل زیر است:



که در آن و انحراف معیار دو متغیر x و Y می باشند.

4. مجموعه داده های نشان داده شده در جدول زیر، تعداد اتاق خوابها، تعداد حمامها و قیمت فروش نمونه ای تصادفی از هشت خانه را نمایش می دهد. با استفاده از روش کمترین مربعات رابطه میان قیمت خانه و تعداد اتاقها و حمامها را پیدا کنید.

|  |  |  |  |
| --- | --- | --- | --- |
| شماره | قیمت(به دلار):Y | تعداد حمامها: x2 | تعداد حمامها: x2 |
| 1 | 78800 | 2 | 3 |
| 2 | 74300 | 1 | 2 |
| 3 | 83800 | 3 | 4 |
| 4 | 74200 | 1 | 2 |
| 5 | 79700 | 2 | 3 |
| 6 | 74900 | 2 | 2 |
| 7 | 88400 | 3 | 5 |
| 8 | 82900 | 2 | 4 |

ج) نمودار پراکندگی y بر حسب X1 و X2 به شکل زیر می شود:

همانگونه که در درس بیان شد با در نظر گرفتن مقادیر مربوط به حمامها و اتاقها به صورت ماتریس دو بعدی 3\*8 به همراه یک ستون با مقادیر 1 برای در نظر گرفتن ضریب عرض از مبدا معادله و قابل ضرب بودن ماتریسها، مقادیر قیمت خانه ها به صورت بردار 1\*8 و پارامترهای ضرایب رگرسیون با در نظر گرفتن ضریب عرض از مبدا به صورت بردار 3\*1 :

ماتریس به صورت زیر محاسبه می شود:



به این منظور :



با استفاده از روش مبتنی بر دترمینان ، معکوس ماتریس حاصلضرب محاسبه می شود:



که در آن 84 مقدار دترمینان  است.

و



و سر انجام:

=





که در آن ها به 0.1 گرد شده اند. بنابراین معادله نهایی رگرسیون به صورت زیر خواهد بود:

(758.3\*تعداد اتاقها) +(4133.3\*تعداد حمامها)+6519.7 =قیمت خانه

7- با فرض اینکه در برنامه ای دو تعیین کننده داشته باشیم، نشان دهید ماتریس نهایی رگرسیون ستیغی شامل مقادیر همبستگی تعیین کننده ها با یکدیگر است.

ج)

|  |  |  |  |
| --- | --- | --- | --- |
| =y1 |  | + |  |
| =y2 |  | + |  |

فرض کنید در برنامه دو تعیین‌کننده داشته باشیم. در دو اجرای مختلف، تعداد True شدن هر تعیین کننده به همراه وضعیت خاتمه برنامه (موفقیت یا عدم موفقیت) تشکیل یک معادله را می‌دهد:

(1)

اصولاً در آمار و احتمال برای نرمالسازی داده‌ها و آسان نمودن انجام محاسبات برروی آنها، داده ها را مرکزی[[1]](#footnote-2) و مقیاس بندی[[2]](#footnote-3) می ‌کنند. برای مرکزی کردن داده ها، هریک از داده ها را از میانگینشان تفریق می‌کنند و برای مقیاس‌بندی آنها (از بین بردن وابستگی داده ها به یک واحد خاص اندازه‌گیری) داده مرکزی شده را بر انحراف معیار آن تقسیم می‌کنند. به این ترتیب هر تعیین‌کننده به مقدار نرمالسازی شده  تبدیل می‌شود:

(2) =

به این ترتیب،

(3) 

و

|  |  |  |  |
| --- | --- | --- | --- |
| =y1 |  | + |  |
| =y2 |  | + |  |

(4)

با ضرب کردن دوطرف دستگاه معادلات در ترانهاده ماتریس Z ، ماتریس تعیین کننده ها به ماتریس همبستگی هر تعیین‌کننده با سایر تعیین‌کننده ها در سمت چپ معادلات و ماتریس همبستگی میان وضعیت خاتمه برنامه و سایر تعیین‌کننده‌ها تبدیل می‌شود:

با در نظر گرفتن ساختار دستگاه معادلات (4) به صورت ماتریسی

(5) B\*Z=Y

که B بردار ستونی و Z یک ماتریس 2\*2 شامل داده های نرمال شده تعیین کننده‌ها و Y بردار ستونی وضعیت خاتمه برنامه است. با ضرب (5) در ماتریس ترانهاده Z داریم:

B\*ZTZ=ZTY (6)

برای اثبات اینکه حاصلضرب ZTZ و ZTY ماتریسهای همبستگی هستند، نشان می‌دهیم:

(7)

ZTZ==

و

(8) ZTY=

اما اینکه چگونه درایه‌های این ماتریس، بیانگر همبستگی مابین تعیین‌کننده‌ها برای (7) و همبستگی بین تعیین کننده ها و وضعیت خاتمه برنامه برای (8) هستند، تعریف همبستگی میان X و Y را مرور می‌کنیم:

(9) cor(Y,X)=

درایه (1,1) ماتریس (7) به معنای وابستگی تعیین کننده اول با خودش است و بدیهی است که مقدار همبستگی، یک می‌شود (cor=1)

(10)

برای درایه (2،1) داریم:

(11)

Cor(P1,P2)

که نشان‌دهنده همبستگی میان تعیین کننده P1 و P2 است. به این ترتیب برای سایر درایه های ماتریس حاصلضرب، می توان نشان داد، درایه‌ها نقش عنصر همبستگی را ایفا می‌کنند:

(12) ZTZ=

1. Centering [↑](#footnote-ref-2)
2. Scaling [↑](#footnote-ref-3)