

تحلیل مؤلفه‌های اصلی

بخش اول: تعریف مؤلفه‌های اصلی جامعه

مقدمه

فرض کنید بردار تصادفی

$$\mathbf{X} = (X_1, X_2, \dots, X_p)'$$

دارای بردار میانگین

$$E(\mathbf{X}) = \boldsymbol{\mu}$$

و ماتریس کوواریانس

$$\text{Cov}(\mathbf{X}) = \Sigma$$

باشد. هدف تحلیل مؤلفه‌های اصلی آن است که از متغیرهای اولیه، ترکیب‌های خطی جدیدی بسازیم که بیشترین مقدار ممکن از تغییرات داده‌ها را در خود داشته باشند.

مطابق تعریف، ترکیب‌های خطی را به صورت زیر در نظر می‌گیریم:

$$Y_i = \mathbf{e}_i' \mathbf{X}, \quad i = 1, 2, \dots, p,$$

که در آن

$$\mathbf{e}_i = (e_{1i}, e_{2i}, \dots, e_{pi})'$$

بردار ضرایب مؤلفه i ام است.

۱. چرا ترکیب خطی می‌سازیم؟

برای مثال، اگر سه متغیر X_1, X_2, X_3 داشته باشیم، یک ترکیب خطی عمومی به صورت

$$Y = e_1 X_1 + e_2 X_2 + e_3 X_3$$

است.

این متغیر جدید می‌تواند اطلاعات چند متغیر اولیه را به طور هم‌زمان در خود جمع کند. اما هر ترکیب خطی مناسب

نیست. ما ترکیبی را می‌خواهیم که تا حد امکان تغییرات بیشتری از داده‌ها را منعکس کند. بنابراین معیار انتخاب، واریانس ترکیب خطی است.

۲. واریانس یک ترکیب خطی

اگر

$$Y = e'X,$$

آنگاه

$$\text{Var}(Y) = \text{Var}(e'X) = e'\Sigma e.$$

همچنین اگر دو ترکیب خطی

$$Y_i = e_i'X, \quad Y_k = e_k'X$$

داشته باشیم، کوواریانس آن‌ها برابر است با

$$\text{Cov}(Y_i, Y_k) = e_i'\Sigma e_k.$$

این دو رابطه، پایه تمام محاسبات تحلیل مؤلفه‌های اصلی هستند.

۳. مشکل نامحدود شدن واریانس

ممکن است بگوییم مؤلفه اصلی اول ترکیب خطی‌ای است که واریانس آن بیشینه باشد:

$$\max_{e \neq 0} e'\Sigma e.$$

اما این مسئله بدون قید جواب معناداری ندارد. زیرا اگر بردار e واریانس مثبتی ایجاد کند، با ضرب کردن آن در عددی مانند c ، داریم

$$\text{Var}(ce'X) = c^2 e'\Sigma e.$$

با بزرگ کردن c ، واریانس نیز بدون حد بزرگ می‌شود. بنابراین باید طول بردار ضرایب را ثابت کنیم. شرط استاندارد آن است که

$$e'e = 1.$$

یعنی بردار ضرایب باید یک بردار واحد باشد.

۴. تعریف مؤلفه اصلی اول

مؤلفه اصلی اول ترکیب خطی

$$Y_1 = e_1'X$$

است که در میان تمام ترکیب‌های خطی دارای شرط

$$e_1' e_1 = 1$$

بیشترین واریانس را داشته باشد.

پس مسئله بهینه‌سازی به صورت زیر است:

$$\max_{e_1} e_1' \Sigma e_1$$

به شرط

$$e_1' e_1 = 1.$$

۵. استخراج مؤلفه اصلی اول با روش لاگرانژ

تابع لاگرانژ را تشکیل می‌دهیم:

$$\mathcal{L}(e_1, \lambda) = e_1' \Sigma e_1 - \lambda(e_1' e_1 - 1).$$

نسبت به e_1 مشتق می‌گیریم:

$$\frac{\partial \mathcal{L}}{\partial e_1} = 2\Sigma e_1 - 2\lambda e_1.$$

با برابر صفر قرار دادن مشتق:

$$2\Sigma e_1 - 2\lambda e_1 = 0,$$

بنابراین

$$\Sigma e_1 = \lambda e_1.$$

پس e_1 باید یک بردار ویژه ماتریس کوواریانس Σ باشد و λ نیز مقدار ویژه متناظر آن است.

حال دو طرف رابطه را از چپ در e_1' ضرب می‌کنیم:

$$e_1' \Sigma e_1 = \lambda e_1' e_1.$$

چون

$$e_1' e_1 = 1,$$

داریم

$$e_1' \Sigma e_1 = \lambda.$$

ولی سمت چپ همان واریانس مؤلفه است. بنابراین

$$\text{Var}(Y_1) = \lambda.$$

برای بیشینه شدن واریانس، باید بزرگ‌ترین مقدار ویژه انتخاب شود. اگر مقادیرهای ویژه به ترتیب

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$$

باشند، آنگاه

$$\text{Var}(Y_1) = \lambda_1$$

و بردار ویژه متناظر با λ_1 است. بنابراین

$$Y_1 = e_1' X$$

مؤلفه اصلی اول جامعه است.

۶. تعریف مؤلفه اصلی دوم

مؤلفه اصلی دوم نیز باید واریانس بیشینه داشته باشد، اما نباید اطلاعات مؤلفه اول را تکرار کند. بنابراین علاوه بر شرط طول واحد، باید با مؤلفه اول ناهمبسته باشد.

پس

$$Y_2 = e_2' X$$

به گونه‌ای انتخاب می‌شود که

$$\text{Var}(Y_2) = e_2' \Sigma e_2$$

بیشینه باشد، تحت قیود

$$e_2' e_2 = 1$$

و

$$\text{Cov}(Y_1, Y_2) = 0.$$

از آنجا که

$$\text{Cov}(Y_1, Y_2) = e_1' \Sigma e_2,$$

و

$$\Sigma e_1 = \lambda_1 e_1,$$

داریم

$$\text{Cov}(Y_1, Y_2) = \lambda_1 e_1' e_2.$$

بنابراین برای صفر شدن کوواریانس کافی است

$$e_1' e_2 = 0.$$

یعنی بردارهای ضرایب مؤلفه‌های اول و دوم بر یکدیگر عمود باشند.

در نهایت، e_2 بردار ویژه متناظر با دومین مقدار ویژه بزرگ است و

$$\text{Var}(Y_2) = \lambda_2.$$

به همین ترتیب

$$Y_i = \mathbf{e}_i' \mathbf{X}, \quad \text{Var}(Y_i) = \lambda_i.$$

برای $i \neq k$ نیز

$$\text{Cov}(Y_i, Y_k) = 0.$$

مثال اول: ماتریس کوواریانس قطری

فرض کنید

$$\Sigma = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}.$$

می‌خواهیم مؤلفه‌های اصلی را به دست آوریم.

مرحله اول: مقدارهای ویژه

معادله مشخصه:

$$|\Sigma - \lambda I| = 0.$$

پس

$$\begin{vmatrix} 9 - \lambda & 0 \\ 0 & 4 - \lambda \end{vmatrix} = 0.$$

بنابراین

$$(9 - \lambda)(4 - \lambda) = 0.$$

در نتیجه

$$\lambda_1 = 9, \quad \lambda_2 = 4.$$

مرحله دوم: بردار ویژه مربوط به $\lambda_1 = 9$

باید حل کنیم:

$$(\Sigma - 9I)\mathbf{e}_1 = 0.$$

یعنی

$$\begin{pmatrix} 0 & 0 \\ 0 & -5 \end{pmatrix} \begin{pmatrix} e_{11} \\ e_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

از معادله دوم:

$$-5e_{21} = 0 \implies e_{21} = 0.$$

با شرط طول واحد

$$e_{11}^2 + e_{21}^2 = 1,$$

داریم

$$e_{11} = 1.$$

پس می‌توان انتخاب کرد:

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

بنابراین مؤلفه اصلی اول:

$$Y_1 = \mathbf{e}_1' \mathbf{X} = X_1.$$

و واریانس آن:

$$\text{Var}(Y_1) = 9.$$

مرحله سوم: بردار ویژه مربوط به $\lambda_2 = 4$

داریم

$$(\Sigma - 4I)\mathbf{e}_2 = 0,$$

یعنی

$$\begin{pmatrix} 5 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} e_{12} \\ e_{22} \end{pmatrix} = 0.$$

بنابراین

$$e_{12} = 0.$$

با شرط طول واحد می‌توان نوشت:

$$\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

پس مؤلفه اصلی دوم:

$$Y_2 = X_2$$

و

$$\text{Var}(Y_2) = 4.$$

تفسیر مثال اول

در این مثال متغیرها از ابتدا ناهمبسته‌اند و ماتریس کوواریانس قطری است. بنابراین محورهاى اصلی همان محورهاى اولیه‌اند.

چون

$$\text{Var}(X_1) = 9 > \text{Var}(X_2) = 4,$$

متغیر X_1 مؤلفه اصلی اول است.

سهم مؤلفه اول از واریانس کل برابر است با

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{9}{13} \approx 0.692.$$

یعنی حدود 69.2% واریانس کل توسط مؤلفه اول توضیح داده می‌شود.

مثال دوم: ماتریس کوواریانس غیرقطری

فرض کنید

$$\Sigma = \begin{pmatrix} 5 & 2 \\ 2 & 5 \end{pmatrix}.$$

در این حالت X_1 و X_2 با هم همبستگی مثبت دارند.

مرحله اول: محاسبه مقادارهای ویژه

معادله مشخصه:

$$|\Sigma - \lambda I| = 0.$$

بنابراین

$$\begin{vmatrix} 5 - \lambda & 2 \\ 2 & 5 - \lambda \end{vmatrix} = 0.$$

در نتیجه

$$(5 - \lambda)^2 - 4 = 0.$$

پس

$$(5 - \lambda)^2 = 4.$$

بنابراین

$$5 - \lambda = \pm 2.$$

پس مقادارهای ویژه:

$$\lambda_1 = 7, \quad \lambda_2 = 3.$$

مرحله دوم: بردار ویژه مربوط به $\lambda_1 = 7$

داریم

$$(\Sigma - 7I)e_1 = 0.$$

یعنی

$$\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} e_{11} \\ e_{21} \end{pmatrix} = 0.$$

معادله اول می‌دهد:

$$-2e_{11} + 2e_{21} = 0,$$

بنابراین

$$e_{21} = e_{11}.$$

پس بردار ویژه متناسب است با

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

برای واحد کردن بردار:

$$\sqrt{1^2 + 1^2} = \sqrt{2}.$$

پس

$$\mathbf{e}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

بنابراین مؤلفه اصلی اول:

$$Y_1 = \frac{1}{\sqrt{2}}X_1 + \frac{1}{\sqrt{2}}X_2$$

و واریانس آن:

$$\text{Var}(Y_1) = \lambda_1 = 7.$$

بررسی مستقیم واریانس مؤلفه اول

داریم

$$Y_1 = \frac{X_1 + X_2}{\sqrt{2}}.$$

پس

$$\text{Var}(Y_1) = \frac{1}{2} \text{Var}(X_1 + X_2).$$

بنابراین

$$\text{Var}(Y_1) = \frac{1}{2} [\text{Var}(X_1) + \text{Var}(X_2) + 2 \text{Cov}(X_1, X_2)].$$

با جایگذاری:

$$\text{Var}(Y_1) = \frac{1}{2}(5 + 5 + 2 \times 2) = \frac{14}{2} = 7.$$

مرحله سوم: بردار ویژه مربوط به $\lambda_2 = 3$

داریم

$$(\Sigma - 3I)\mathbf{e}_2 = 0,$$

یعنی

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} e_{12} \\ e_{22} \end{pmatrix} = 0.$$

پس

$$2e_{12} + 2e_{22} = 0,$$

در نتیجه

$$e_{22} = -e_{12}.$$

بردار ویژه متناسب با

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

است. پس بردار ویژه واحد:

$$\mathbf{e}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

بنابراین مؤلفه اصلی دوم:

$$Y_2 = \frac{1}{\sqrt{2}}X_1 - \frac{1}{\sqrt{2}}X_2$$

و

$$\text{Var}(Y_2) = 3.$$

بررسی ناهمبستگی مؤلفه‌ها

داریم

$$\text{Cov}(Y_1, Y_2) = \mathbf{e}_1' \Sigma \mathbf{e}_2.$$

از آنجا که

$$\Sigma \mathbf{e}_2 = 3\mathbf{e}_2,$$

پس

$$\text{Cov}(Y_1, Y_2) = 3\mathbf{e}_1' \mathbf{e}_2.$$

ولی

$$\mathbf{e}_1' \mathbf{e}_2 = \frac{1}{2}(1 \times 1 + 1 \times (-1)) = 0.$$

بنابراین

$$\boxed{\text{Cov}(Y_1, Y_2) = 0}.$$

تفسیر مثال دوم

در این مثال متغیرها همبستگی مثبت دارند. به همین دلیل مؤلفه اصلی اول تقریباً «میانگین» دو متغیر را نشان می‌دهد:

$$Y_1 = \frac{X_1 + X_2}{\sqrt{2}}.$$

اگر هر دو متغیر هم‌زمان زیاد یا کم شوند، Y_1 تغییر زیادی می‌کند. بنابراین این مؤلفه تغییرات مشترک دو متغیر را اندازه می‌گیرد.

مؤلفه دوم

$$Y_2 = \frac{X_1 - X_2}{\sqrt{2}}$$

تفاوت میان دو متغیر را نشان می‌دهد.

سهم مؤلفه اول از واریانس کل:

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{7}{10} = 0.7.$$

پس مؤلفه اول 70% و مؤلفه دوم 30% واریانس کل را توضیح می‌دهند.

نکات مهم این بخش

۱. مؤلفه اصلی، یک متغیر اولیه نیست؛ بلکه ترکیب خطی متغیرهای اولیه است.

۲. شرط

$$e'e = 1$$

برای جلوگیری از بزرگ شدن دلخواه واریانس ضروری است.

۳. ضرایب مؤلفه‌ها همان بردارهای ویژه ماتریس کوواریانس‌اند.

۴. واریانس هر مؤلفه برابر مقدار ویژه متناظر آن است:

$$\text{Var}(Y_i) = \lambda_i.$$

۵. مؤلفه اول متناظر با بزرگ‌ترین مقدار ویژه است.

۶. مؤلفه‌های مختلف با یکدیگر ناهمبسته‌اند:

$$\text{Cov}(Y_i, Y_j) = 0, \quad i \neq j.$$

۷. علامت بردار ویژه یکتا نیست. برای مثال

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{و} \quad -\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

هر دو بردار ویژه معتبرند. فقط علامت مؤلفه عوض می‌شود و واریانس و تفسیر اصلی تغییر نمی‌کند.

۱ مثال: محاسبه مؤلفه‌های اصلی از ماتریس کوواریانس

فرض کنید بردار تصادفی

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

دارای ماتریس کوواریانس زیر باشد:

$$\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

هدف آن است که مقادیر ویژه، بردارهای ویژه، مؤلفه‌های اصلی و سهم هر مؤلفه از واریانس کل را به دست آوریم.

۱.۱ مرحله اول: محاسبه مقادیر ویژه

مقادیر ویژه ماتریس Σ از حل معادله مشخصه زیر به دست می‌آیند:

$$|\Sigma - \lambda I| = 0.$$

داریم:

$$\Sigma - \lambda I = \begin{pmatrix} 1 - \lambda & -2 & 0 \\ -2 & 5 - \lambda & 0 \\ 0 & 0 & 2 - \lambda \end{pmatrix}.$$

بنابراین:

$$\begin{aligned} |\Sigma - \lambda I| &= \begin{vmatrix} 1 - \lambda & -2 & 0 \\ -2 & 5 - \lambda & 0 \\ 0 & 0 & 2 - \lambda \end{vmatrix} \\ &= (2 - \lambda) \begin{vmatrix} 1 - \lambda & -2 \\ -2 & 5 - \lambda \end{vmatrix}. \end{aligned}$$

دترمینان مرتبه دوم برابر است با:

$$\begin{aligned}\begin{vmatrix} 1-\lambda & -2 \\ -2 & 5-\lambda \end{vmatrix} &= (1-\lambda)(5-\lambda) - (-2)(-2) \\ &= (1-\lambda)(5-\lambda) - 4 \\ &= 5 - 6\lambda + \lambda^2 - 4 \\ &= \lambda^2 - 6\lambda + 1.\end{aligned}$$

در نتیجه معادله مشخصه به صورت زیر است:

$$(2-\lambda)(\lambda^2 - 6\lambda + 1) = 0.$$

بنابراین یک مقدار ویژه برابر است با:

$$\lambda = 2.$$

دو مقدار ویژه دیگر از حل معادله

$$\lambda^2 - 6\lambda + 1 = 0$$

به دست می‌آیند. با استفاده از فرمول حل معادله درجه دوم داریم:

$$\begin{aligned}\lambda &= \frac{6 \pm \sqrt{(-6)^2 - 4(1)(1)}}{2} \\ &= \frac{6 \pm \sqrt{36 - 4}}{2} \\ &= \frac{6 \pm \sqrt{32}}{2} \\ &= \frac{6 \pm 4\sqrt{2}}{2} \\ &= 3 \pm 2\sqrt{2}.\end{aligned}$$

پس مقادیر ویژه عبارت‌اند از:

$$3 + 2\sqrt{2}, \quad 2, \quad 3 - 2\sqrt{2}.$$

با مرتب‌کردن مقادیر ویژه از بزرگ به کوچک:

$$\boxed{\lambda_1 = 3 + 2\sqrt{2}}$$

$$\boxed{\lambda_2 = 2}$$

و

$$\boxed{\lambda_3 = 3 - 2\sqrt{2}}$$

مقادیر تقریبی آن‌ها برابرند با:

$$\lambda_1 \approx 5.828, \quad \lambda_2 = 2, \quad \lambda_3 \approx 0.172.$$

۲.۱ مرحله دوم: محاسبه بردار ویژه مربوط به λ_1

برای به دست آوردن بردار ویژه مربوط به

$$\lambda_1 = 3 + 2\sqrt{2},$$

باید دستگاه زیر را حل کنیم:

$$(\Sigma - \lambda_1 I)\mathbf{e}_1 = \mathbf{0}.$$

فرض می‌کنیم:

$$\mathbf{e}_1 = \begin{pmatrix} e_{11} \\ e_{21} \\ e_{31} \end{pmatrix}.$$

بنابراین:

$$\begin{pmatrix} 1 - \lambda_1 & -2 & 0 \\ -2 & 5 - \lambda_1 & 0 \\ 0 & 0 & 2 - \lambda_1 \end{pmatrix} \begin{pmatrix} e_{11} \\ e_{21} \\ e_{31} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

از معادله سوم داریم:

$$(2 - \lambda_1)e_{31} = 0.$$

چون

$$\lambda_1 \neq 2,$$

پس:

$$e_{31} = 0.$$

معادله اول دستگاه برابر است با:

$$(1 - \lambda_1)e_{11} - 2e_{21} = 0.$$

با توجه به اینکه

$$1 - \lambda_1 = 1 - (3 + 2\sqrt{2}) = -2 - 2\sqrt{2},$$

داریم:

$$(-2 - 2\sqrt{2})e_{11} - 2e_{21} = 0.$$

با تقسیم بر -2 :

$$(1 + \sqrt{2})e_{11} + e_{21} = 0.$$

بنابراین:

$$e_{21} = -(1 + \sqrt{2})e_{11}.$$

اگر

$$e_{11} = 1$$

در نظر گرفته شود، یک بردار ویژه غیرواحد برابر است با:

$$\tilde{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ -(1 + \sqrt{2}) \\ 0 \end{pmatrix}.$$

برای تبدیل آن به بردار واحد، طول آن را محاسبه می‌کنیم:

$$\begin{aligned} \|\tilde{\mathbf{e}}_1\|^2 &= 1^2 + (1 + \sqrt{2})^2 \\ &= 1 + 1 + 2\sqrt{2} + 2 \\ &= 4 + 2\sqrt{2}. \end{aligned}$$

بنابراین:

$$\|\tilde{\mathbf{e}}_1\| = \sqrt{4 + 2\sqrt{2}}.$$

پس بردار ویژه واحد مربوط به λ_1 برابر است با:

$$\mathbf{e}_1 = \frac{1}{\sqrt{4 + 2\sqrt{2}}} \begin{pmatrix} 1 \\ -(1 + \sqrt{2}) \\ 0 \end{pmatrix}$$

مقدار تقریبی این بردار برابر است با:

$$\mathbf{e}_1 \approx \begin{pmatrix} 0.383 \\ -0.924 \\ 0 \end{pmatrix}$$

۳.۱ مرحله سوم: محاسبه بردار ویژه مربوط به λ_2

برای مقدار ویژه

$$\lambda_2 = 2$$

باید دستگاه زیر را حل کنیم:

$$(\Sigma - 2I)\mathbf{e}_2 = \mathbf{0}.$$

داریم:

$$\Sigma - 2I = \begin{pmatrix} -1 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

اگر

$$\mathbf{e}_2 = \begin{pmatrix} e_{12} \\ e_{22} \\ e_{32} \end{pmatrix},$$

آنگاه دو معادله اول دستگاه عبارت‌اند از:

$$-e_{12} - 2e_{22} = 0,$$

و

$$-2e_{12} + 3e_{22} = 0.$$

از معادله اول:

$$e_{12} = -2e_{22}.$$

با جایگذاری در معادله دوم:

$$-2(-2e_{22}) + 3e_{22} = 0,$$

بنابراین:

$$7e_{22} = 0,$$

و در نتیجه:

$$e_{22} = 0, \quad e_{12} = 0.$$

متغیر e_{32} آزاد است. با انتخاب

$$e_{32} = 1$$

بردار ویژه واحد برابر است با:

$$\mathbf{e}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

۴.۱ مرحله چهارم: محاسبه بردار ویژه مربوط به λ_3

برای مقدار ویژه

$$\lambda_3 = 3 - 2\sqrt{2}$$

دستگاه زیر را حل می‌کنیم:

$$(\Sigma - \lambda_3 I)\mathbf{e}_3 = \mathbf{0}.$$

فرض می‌کنیم:

$$\mathbf{e}_3 = \begin{pmatrix} e_{13} \\ e_{23} \\ e_{33} \end{pmatrix}.$$

از معادله سوم داریم:

$$(2 - \lambda_3)e_{33} = 0.$$

چون

$$\lambda_3 \neq 2,$$

پس:

$$e_{33} = 0.$$

معادله اول برابر است با:

$$(1 - \lambda_3)e_{13} - 2e_{23} = 0.$$

اکنون:

$$1 - \lambda_3 = 1 - (3 - 2\sqrt{2}) = -2 + 2\sqrt{2} = 2(\sqrt{2} - 1).$$

در نتیجه:

$$2(\sqrt{2} - 1)e_{13} - 2e_{23} = 0.$$

با تقسیم بر 2:

$$(\sqrt{2} - 1)e_{13} - e_{23} = 0.$$

بنابراین:

$$e_{23} = (\sqrt{2} - 1)e_{13}.$$

اگر

$$e_{13} = 1$$

انتخاب شود، یک بردار ویژه غیرواحد برابر است با:

$$\tilde{\mathbf{e}}_3 = \begin{pmatrix} 1 \\ \sqrt{2} - 1 \\ 0 \end{pmatrix}.$$

طول این بردار برابر است با:

$$\begin{aligned} \|\tilde{\mathbf{e}}_3\|^2 &= 1^2 + (\sqrt{2} - 1)^2 \\ &= 1 + 2 - 2\sqrt{2} + 1 \\ &= 4 - 2\sqrt{2}. \end{aligned}$$

پس:

$$\|\tilde{\mathbf{e}}_3\| = \sqrt{4 - 2\sqrt{2}}.$$

بنابراین بردار ویژه واحد مربوط به λ_3 برابر است با:

$$\mathbf{e}_3 = \frac{1}{\sqrt{4 - 2\sqrt{2}}} \begin{pmatrix} 1 \\ \sqrt{2} - 1 \\ 0 \end{pmatrix}$$

مقدار تقریبی آن برابر است با:

$$\mathbf{e}_3 \approx \begin{pmatrix} 0.924 \\ 0.383 \\ 0 \end{pmatrix}$$

۵.۱ مرحله پنجم: بررسی متعامد بودن بردارهای ویژه

برای مثال:

$$\begin{aligned} \mathbf{e}'_1 \mathbf{e}_3 &\approx (0.383)(0.924) + (-0.924)(0.383) + (0)(0) \\ &\approx 0.354 - 0.354 \\ &= 0. \end{aligned}$$

همچنین:

$$\mathbf{e}'_1 \mathbf{e}_2 = 0, \quad \mathbf{e}'_2 \mathbf{e}_3 = 0.$$

بنابراین بردارهای ویژه متناظر با مقادیر ویژه متفاوت، دو به دو متعامد هستند.

۶.۱ مرحله ششم: تعیین مؤلفه‌های اصلی

مؤلفه اصلی اول از رابطه

$$Y_1 = \mathbf{e}'_1 \mathbf{X}$$

به دست می‌آید. بنابراین:

$$Y_1 = \frac{1}{\sqrt{4 + 2\sqrt{2}}} X_1 - \frac{1 + \sqrt{2}}{\sqrt{4 + 2\sqrt{2}}} X_2.$$

در نتیجه، به صورت تقریبی:

$$Y_1 = 0.383X_1 - 0.924X_2$$

و واریانس آن برابر است با:

$$\text{Var}(Y_1) = \lambda_1 = 3 + 2\sqrt{2} \approx 5.828$$

مؤلفه اصلی دوم برابر است با:

$$Y_2 = \mathbf{e}_2' \mathbf{X}.$$

پس:

$$Y_2 = X_3$$

و واریانس آن برابر است با:

$$\text{Var}(Y_2) = \lambda_2 = 2$$

مؤلفه اصلی سوم برابر است با:

$$Y_3 = \mathbf{e}_3' \mathbf{X}.$$

بنابراین:

$$Y_3 = \frac{1}{\sqrt{4 - 2\sqrt{2}}} X_1 + \frac{\sqrt{2} - 1}{\sqrt{4 - 2\sqrt{2}}} X_2.$$

در نتیجه، به صورت تقریبی:

$$Y_3 = 0.924X_1 + 0.383X_2$$

و واریانس آن برابر است با:

$$\text{Var}(Y_3) = \lambda_3 = 3 - 2\sqrt{2} \approx 0.172$$

در نتیجه مؤلفه‌های اصلی به ترتیب عبارت‌اند از:

$$\begin{aligned} Y_1 &= 0.383X_1 - 0.924X_2, \\ Y_2 &= X_3, \\ Y_3 &= 0.924X_1 + 0.383X_2. \end{aligned}$$

توجه شود که علامت یک بردار ویژه اختیاری است. برای مثال می‌توان $\mathbf{e}_1$ را در عدد -1 ضرب کرد. در این حالت مؤلفه اصلی اول به صورت

$$Y_1 = -0.383X_1 + 0.924X_2$$

نوشته می‌شود، اما واریانس و سهم آن از واریانس کل تغییر نمی‌کند.

۷.۱ مرحله هفتم: محاسبه واریانس کل

واریانس کل متغیرهای اولیه برابر مجموع عناصر قطر اصلی ماتریس کوواریانس است:

$$\text{tr}(\Sigma) = 1 + 5 + 2 = 8.$$

همچنین مجموع مقادیر ویژه برابر است با:

$$\begin{aligned}\lambda_1 + \lambda_2 + \lambda_3 &= (3 + 2\sqrt{2}) + 2 + (3 - 2\sqrt{2}) \\ &= 8.\end{aligned}$$

بنابراین:

$$\text{tr}(\Sigma) = \lambda_1 + \lambda_2 + \lambda_3 = 8$$

۸.۱ مرحله هشتم: سهم هر مؤلفه از واریانس کل

سهم مؤلفه اصلی i ام از واریانس کل از رابطه زیر محاسبه می‌شود:

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \lambda_3}.$$

از آنجا که مجموع مقادیر ویژه برابر ۸ است، سهم مؤلفه اول برابر است با:

$$\begin{aligned}\frac{\lambda_1}{8} &= \frac{3 + 2\sqrt{2}}{8} \\ &\approx \frac{5.828}{8} \\ &\approx 0.7286.\end{aligned}$$

بنابراین مؤلفه اول تقریباً

$$72.86\%$$

از واریانس کل را توضیح می‌دهد.

سهم مؤلفه دوم برابر است با:

$$\frac{\lambda_2}{8} = \frac{2}{8} = 0.25.$$

بنابراین مؤلفه دوم:

$$25\%$$

از واریانس کل را توضیح می‌دهد.

سهم مؤلفه سوم برابر است با:

$$\begin{aligned}\frac{\lambda_3}{8} &= \frac{3 - 2\sqrt{2}}{8} \\ &\approx \frac{0.172}{8} \\ &\approx 0.0214.\end{aligned}$$

بنابراین مؤلفه سوم تقریباً

$$2.14\%$$

از واریانس کل را توضیح می‌دهد.

۹.۱ جدول واریانس تبیین شده

مؤلفه	مقدار ویژه	درصد واریانس تبیین شده	درصد تجمعی
Y_1	$3 + 2\sqrt{2} \approx 5.828$	72.86%	72.86%
Y_2	2	25.00%	97.86%
Y_3	$3 - 2\sqrt{2} \approx 0.172$	2.14%	100%

۱۰.۱ نتیجه‌گیری و کاهش بعد

دو مؤلفه اصلی اول در مجموع:

$$72.86\% + 25\% = 97.86\%$$

از واریانس کل را توضیح می‌دهند.

بنابراین:

$$\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} \approx 0.9786$$

از آنجا که دو مؤلفه اول حدود 97.86% از واریانس کل را دربرمی‌گیرند، می‌توان در صورت هدف‌بودن کاهش بعد، متغیرهای اولیه را با دو مؤلفه

$$Y_1 = 0.383X_1 - 0.924X_2$$

و

$$Y_2 = X_3$$

خلاصه کرد و مؤلفه سوم را، که تنها حدود 2.14% از واریانس کل را توضیح می‌دهد، کنار گذاشت.

۲ ضریب همبستگی بین مؤلفه‌های اصلی و متغیرهای اولیه

برای تفسیر مؤلفه‌های اصلی، تنها بررسی ضرایب موجود در بردارهای ویژه کافی نیست؛ زیرا اندازه این ضرایب تحت تأثیر مقیاس متغیرها قرار می‌گیرد. یکی از معیارهای مناسب برای بررسی اهمیت هر متغیر در یک مؤلفه اصلی، ضریب همبستگی میان متغیر اولیه و مؤلفه اصلی است.

فرض کنید مؤلفه اصلی i ام به صورت

$$Y_i = e_i'X$$

تعریف شده باشد، که در آن

$$\mathbf{e}_i = \begin{pmatrix} e_{1i} \\ e_{2i} \\ \vdots \\ e_{pi} \end{pmatrix}$$

بردار ویژه واحد متناظر با مقدار ویژه λ_i است.

می‌دانیم:

$$\Sigma \mathbf{e}_i = \lambda_i \mathbf{e}_i.$$

همچنین واریانس مؤلفه اصلی i ام برابر است با:

$$\text{Var}(Y_i) = \lambda_i.$$

اکنون می‌خواهیم ضریب همبستگی میان Y_i و متغیر X_k را به دست آوریم.

۱.۲ استخراج فرمول ضریب همبستگی

طبق تعریف:

$$\rho_{Y_i, X_k} = \frac{\text{Cov}(Y_i, X_k)}{\sqrt{\text{Var}(Y_i) \text{Var}(X_k)}}.$$

از آنجا که

$$Y_i = \mathbf{e}_i' \mathbf{X},$$

داریم:

$$\text{Cov}(Y_i, X_k) = \text{Cov}(\mathbf{e}_i' \mathbf{X}, X_k).$$

اگر $\mathbf{a}_k$ بردار واحد مختصاتی باشد که درایه k ام آن برابر یک و سایر درایه‌های آن صفر است، آنگاه:

$$X_k = \mathbf{a}_k' \mathbf{X}.$$

در نتیجه:

$$\begin{aligned} \text{Cov}(Y_i, X_k) &= \text{Cov}(\mathbf{e}_i' \mathbf{X}, \mathbf{a}_k' \mathbf{X}) \\ &= \mathbf{e}_i' \Sigma \mathbf{a}_k. \end{aligned}$$

چون ماتریس Σ متقارن است، داریم:

$$\mathbf{e}_i' \Sigma = (\Sigma \mathbf{e}_i)' = (\lambda_i \mathbf{e}_i)' = \lambda_i \mathbf{e}_i'.$$

پس:

$$\begin{aligned}\text{Cov}(Y_i, X_k) &= \lambda_i \mathbf{e}_i' \mathbf{a}_k \\ &= \lambda_i e_{ki}.\end{aligned}$$

از طرف دیگر:

$$\text{Var}(Y_i) = \lambda_i$$

و

$$\text{Var}(X_k) = \sigma_{kk}.$$

بنابراین:

$$\begin{aligned}\rho_{Y_i, X_k} &= \frac{\lambda_i e_{ki}}{\sqrt{\lambda_i \sigma_{kk}}} \\ &= \frac{e_{ki} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}.\end{aligned}$$

پس فرمول کلی ضریب همبستگی میان مؤلفه اصلی i ام و متغیر اولیه X_k برابر است با:

$$\boxed{\rho_{Y_i, X_k} = \frac{e_{ki} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}}$$

که در آن:

- e_{ki} ضریب متغیر X_k در مؤلفه اصلی Y_i است؛
- λ_i واریانس مؤلفه اصلی Y_i است؛
- σ_{kk} واریانس متغیر اولیه X_k است.

۳ تکمیل مثال قبلی: همبستگی مؤلفه‌ها با متغیرهای اولیه

در مثال قبلی داشتیم:

$$\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

مقادیر ویژه مرتب شده برابر بودند با:

$$\lambda_1 = 3 + 2\sqrt{2} \approx 5.828, \quad \lambda_2 = 2, \quad \lambda_3 = 3 - 2\sqrt{2} \approx 0.172.$$

بردارهای ویژه واحد نیز به صورت زیر به دست آمدند:

$$\mathbf{e}_1 \approx \begin{pmatrix} 0.383 \\ -0.924 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{e}_3 \approx \begin{pmatrix} 0.924 \\ 0.383 \\ 0 \end{pmatrix}.$$

مؤلفه‌های اصلی عبارت بودند از:

$$Y_1 = 0.383X_1 - 0.924X_2,$$

$$Y_2 = X_3,$$

و

$$Y_3 = 0.924X_1 + 0.383X_2.$$

همچنین:

$$\sigma_{11} = 1, \quad \sigma_{22} = 5, \quad \sigma_{33} = 2.$$

۱.۳ همبستگی مؤلفه اول با X_1

با استفاده از رابطه

$$\rho_{Y_i, X_k} = \frac{e_{ki}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}},$$

داریم:

$$\begin{aligned} \rho_{Y_1, X_1} &= \frac{e_{11}\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}} \\ &= \frac{0.383\sqrt{5.828}}{\sqrt{1}}. \end{aligned}$$

از آنجا که

$$\sqrt{5.828} \approx 2.414,$$

داریم:

$$\begin{aligned} \rho_{Y_1, X_1} &\approx 0.383(2.414) \\ &\approx 0.925. \end{aligned}$$

بنابراین:

$$\boxed{\rho_{Y_1, X_1} \approx 0.925}$$

این مقدار نشان می‌دهد که میان مؤلفه اصلی اول و متغیر X_1 همبستگی مثبت و بسیار قوی وجود دارد.

۲.۳ همبستگی مؤلفه اول با X_2

داریم:

$$\begin{aligned}\rho_{Y_1, X_2} &= \frac{e_{21}\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}} \\ &= \frac{-0.924\sqrt{5.828}}{\sqrt{5}} \\ &\approx \frac{-0.924(2.414)}{2.236} \\ &\approx -0.998.\end{aligned}$$

بنابراین:

$$\boxed{\rho_{Y_1, X_2} \approx -0.998}$$

این مقدار نشان می‌دهد که مؤلفه اصلی اول با X_2 همبستگی منفی و تقریباً کامل دارد. علامت منفی ناشی از انتخاب علامت بردار ویژه است. اگر بردار ویژه e_1 در عدد -1 ضرب شود، علامت مؤلفه اصلی و تمام همبستگی‌های مربوط به آن تغییر می‌کند، ولی مقدار مطلق همبستگی و تفسیر ساختاری مؤلفه تغییر نمی‌کند.

۳.۳ همبستگی مؤلفه اول با X_3

از آنجا که ضریب X_3 در مؤلفه اول برابر صفر است، داریم:

$$e_{31} = 0.$$

پس:

$$\begin{aligned}\rho_{Y_1, X_3} &= \frac{e_{31}\sqrt{\lambda_1}}{\sqrt{\sigma_{33}}} \\ &= 0.\end{aligned}$$

بنابراین:

$$\boxed{\rho_{Y_1, X_3} = 0}$$

۴.۳ همبستگی مؤلفه دوم با متغیرهای اولیه

برای مؤلفه دوم داریم:

$$Y_2 = X_3.$$

بنابراین انتظار داریم این مؤلفه با X_3 همبستگی کامل و با X_1 و X_2 همبستگی صفر داشته باشد.

برای X_1 :

$$\rho_{Y_2, X_1} = \frac{e_{12}\sqrt{\lambda_2}}{\sqrt{\sigma_{11}}} = 0.$$

برای X_2 :

$$\rho_{Y_2, X_2} = \frac{e_{22}\sqrt{\lambda_2}}{\sqrt{\sigma_{22}}} = 0.$$

برای X_3 :

$$\begin{aligned}\rho_{Y_2, X_3} &= \frac{e_{32}\sqrt{\lambda_2}}{\sqrt{\sigma_{33}}} \\ &= \frac{1 \times \sqrt{2}}{\sqrt{2}} \\ &= 1.\end{aligned}$$

در نتیجه:

$$\boxed{\rho_{Y_2, X_1} = 0, \quad \rho_{Y_2, X_2} = 0, \quad \rho_{Y_2, X_3} = 1}$$

۵.۳ همبستگی مؤلفه سوم با متغیرهای اولیه

برای مؤلفه سوم:

$$\lambda_3 \approx 0.172$$

و

$$\mathbf{e}_3 \approx \begin{pmatrix} 0.924 \\ 0.383 \\ 0 \end{pmatrix}.$$

بنابراین:

$$\begin{aligned}\rho_{Y_3, X_1} &= \frac{0.924\sqrt{0.172}}{\sqrt{1}} \\ &\approx 0.924(0.414) \\ &\approx 0.383.\end{aligned}$$

پس:

$$\boxed{\rho_{Y_3, X_1} \approx 0.383}$$

همچنین:

$$\begin{aligned}\rho_{Y_3, X_2} &= \frac{0.383\sqrt{0.172}}{\sqrt{5}} \\ &\approx \frac{0.383(0.414)}{2.236} \\ &\approx 0.071.\end{aligned}$$

بنابراین:

$$\rho_{Y_3, X_2} \approx 0.071$$

و چون $e_{33} = 0$ ، داریم:

$$\rho_{Y_3, X_3} = 0$$

۶.۳ جدول همبستگی مؤلفه‌ها با متغیرها

مؤلفه	ρ_{Y_i, X_1}	ρ_{Y_i, X_2}	ρ_{Y_i, X_3}
Y_1	0.925	-0.998	0
Y_2	0	0	1
Y_3	0.383	0.071	0

۷.۳ تفسیر نتایج

از جدول بالا نتیجه می‌گیریم:

- مؤلفه اصلی اول با هر دو متغیر X_1 و X_2 ارتباط بسیار قوی دارد؛
- قدر مطلق همبستگی Y_1 با X_2 اندکی بیشتر از همبستگی آن با X_1 است؛
- مؤلفه دوم دقیقاً همان متغیر X_3 است؛
- مؤلفه سوم تنها سهم اندکی از واریانس کل را توضیح می‌دهد و همبستگی آن با متغیرهای اولیه نیز نسبت به مؤلفه اول ضعیف‌تر است.

از آنجا که دو مؤلفه اول در مجموع حدود

97.86%

از واریانس کل را توضیح می‌دهند، می‌توان برای کاهش بعد، مؤلفه سوم را کنار گذاشت و از Y_1 و Y_2 به جای سه متغیر اولیه استفاده کرد.

۴ تفسیر هندسی مؤلفه‌های اصلی

تفسیر هندسی مؤلفه‌های اصلی به‌ویژه در مورد بردارهای تصادفی نرمال چندمتغیره بسیار مفید است.

فرض کنید:

$$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma).$$

سطوح دارای چگالی ثابت در توزیع نرمال چندمتغیره از رابطه زیر به دست می آیند:

$$(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2,$$

که در آن c یک عدد ثابت مثبت است.

در حالت دو متغیره، این معادله یک بیضی و در ابعاد بالاتر یک بیضوی را تعریف می کند.

۱.۴ تجزیه طیفی ماتریس کوواریانس

چون Σ یک ماتریس متقارن و مثبت معین است، می توان آن را به صورت زیر تجزیه کرد:

$$\Sigma = E \Lambda E',$$

که در آن:

$$E = \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_p \end{pmatrix}$$

ماتریس بردارهای ویژه متعامد و

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$$

ماتریس قطری مقادیر ویژه است.

از آنجا که:

$$E' E = E E' = I,$$

داریم:

$$\Sigma^{-1} = E \Lambda^{-1} E'.$$

بنابراین:

$$(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu})' E \Lambda^{-1} E' (\mathbf{x} - \boldsymbol{\mu}).$$

حال مختصات جدید را به صورت

$$\mathbf{y} = E' (\mathbf{x} - \boldsymbol{\mu})$$

تعریف می کنیم.

در نتیجه:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$$

که در آن:

$$y_i = \mathbf{e}_i'(\mathbf{x} - \boldsymbol{\mu}).$$

پس:

$$(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \mathbf{y}' \Lambda^{-1} \mathbf{y}.$$

از آنجا که:

$$\Lambda^{-1} = \text{diag} \left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_p} \right),$$

داریم:

$$c^2 = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} + \dots + \frac{y_p^2}{\lambda_p}$$

این معادله نشان می‌دهد که در دستگاه مختصات مؤلفه‌های اصلی، معادله بیضوی به یک فرم قطری و ساده تبدیل می‌شود.

۲.۴ جهت و طول محورهای بیضوی

معادله

$$\frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} + \dots + \frac{y_p^2}{\lambda_p} = c^2$$

را می‌توان به صورت

$$\frac{y_1^2}{c^2 \lambda_1} + \frac{y_2^2}{c^2 \lambda_2} + \dots + \frac{y_p^2}{c^2 \lambda_p} = 1$$

نوشت.

در نتیجه طول نیم‌محور i ام بیضوی برابر است با:

$$c\sqrt{\lambda_i}$$

بنابراین:

- جهت محور اول بیضوی همان بردار ویژه $\mathbf{e}_1$ است؛
- جهت محور دوم همان بردار ویژه $\mathbf{e}_2$ است؛
- به طور کلی جهت محور i ام همان $\mathbf{e}_i$ است؛
- بزرگ‌ترین مقدار ویژه، بلندترین محور بیضوی را تعیین می‌کند؛
- کوچک‌ترین مقدار ویژه، کوتاه‌ترین محور بیضوی را تعیین می‌کند.

از این رو مؤلفه اصلی اول در جهتی قرار دارد که داده‌ها بیشترین پراکندگی را دارند و مؤلفه‌های بعدی به ترتیب بیشترین پراکندگی باقی‌مانده را نشان می‌دهند.

۳.۴ مختصات نقاط روی محورهای اصلی

روی محور i ام، تمام مختصات به جز y_i صفر هستند. بنابراین:

$$\frac{y_i^2}{\lambda_i} = c^2.$$

پس:

$$y_i = \pm c\sqrt{\lambda_i}.$$

در دستگاه مختصات اولیه، نقاط انتهایی محور i ام برابرند با:

$$\boxed{\mu \pm c\sqrt{\lambda_i} \mathbf{e}_i}$$

بنابراین مقدار ویژه طول محور و بردار ویژه جهت محور را تعیین می‌کند.

۴.۴ ارتباط هندسی با مؤلفه‌های اصلی

مختصات مؤلفه‌های اصلی عبارت‌اند از:

$$Y_i = \mathbf{e}_i'(\mathbf{X} - \boldsymbol{\mu}).$$

به عبارت دیگر، مؤلفه‌های اصلی حاصل تصویر بردار داده‌ها بر محورهای بردارهای ویژه هستند.

در حالت دو متغیره، تحلیل مؤلفه‌های اصلی را می‌توان به صورت دوران محورهای اولیه در نظر گرفت؛ به گونه‌ای که محورهای جدید بر محورهای اصلی بیضی چگالی ثابت منطبق شوند.

این دوران موجب می‌شود:

- ماتریس کوواریانس در دستگاه جدید قطری شود؛
- مؤلفه‌های جدید با یکدیگر ناهمبسته باشند؛
- واریانس هر مؤلفه برابر مقدار ویژه متناظر آن باشد.

تعبیر هندسی مؤلفه‌های اصلی

مثال: $\Sigma = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}$

مقادیر ویژه و بردارهای ویژه

$$\lambda_1 = 3 + 2\sqrt{2} \approx 5.828$$

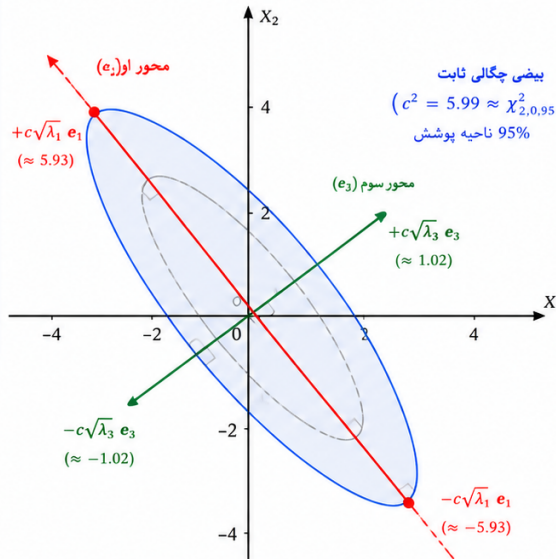
$$\mathbf{e}_1 = \frac{1}{\sqrt{4+2\sqrt{2}}} \begin{pmatrix} 1 \\ -(1+\sqrt{2}) \end{pmatrix} \approx \begin{pmatrix} 0.383 \\ -0.924 \end{pmatrix}$$

$$\lambda_2 = 2$$

$$\mathbf{e}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\lambda_3 = 3 - 2\sqrt{2} \approx 0.172$$

$$\mathbf{e}_3 = \frac{1}{\sqrt{4-2\sqrt{2}}} \begin{pmatrix} 1 \\ \sqrt{2}-1 \end{pmatrix} \approx \begin{pmatrix} 0.924 \\ 0.383 \end{pmatrix}$$



بیضی چگالی ثابت

$$(c^2 = 5.99 \approx \chi_{2,0.95}^2)$$

95% ناحیه پوشش

طول نیم‌محورها
(برای $c^2 = 5.99$)

- نیم‌محور $c\sqrt{\lambda_1}$:
• $\sqrt{5} = \sqrt{5.99 \times 5.828} \approx 5.93$
 - نیم‌محور $c\sqrt{\lambda_3}$:
• $\sqrt{5} = \sqrt{5.99 \times 0.172} \approx 1.02$
- محورها در جهت بردارهای ویژه‌اند.
($\mathbf{e}_1 \perp \mathbf{e}_3$)

فرم بیضی در دستگاه اصلی

معادله بیضی (ناحیه چگالی ثابت):

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2$$

که در آن $\mathbf{c}^2 = (0, 0)$ و ثابت است.
مرکز بیضی در $\boldsymbol{\mu} = (0, 0)$ است.

فرم بیضی در دستگاه مؤلفه‌های اصلی

$$\mathbf{y} = \mathbf{E}'(\mathbf{x} - \boldsymbol{\mu})$$

بیضی به صورت زیر ساده می‌شود:

$$\frac{y_1^2}{\lambda_1} + \frac{y_3^2}{\lambda_3} = c^2$$

که در آن $y_1 = \mathbf{e}_1'(\mathbf{x} - \boldsymbol{\mu})$ ، $y_3 = \mathbf{e}_3'(\mathbf{x} - \boldsymbol{\mu})$
(و y_2 در این مثال مستقل و عمود بر صفحه است.)

تعبیر هندسی

- مؤلفه اول (Y_1) در جهتی قرار دارد که پراکندگی داده‌ها بیشترین مقدار را دارد (بزرگترین مقدار ویژه).
- مؤلفه سوم (Y_3) عمود بر Y_1 است و بیشترین پراکندگی باقیمانده را توضیح می‌دهد.
- مؤلفه دوم (Y_2) در راستای محور X_3 قرار دارد و در این شکل نمایش داده نشده است.
- تقاطع روی بیضی با فاصله c از مرکز، یکسان چگالی دارند.

راهنما

- محور اول ($\mathbf{e}_1 - \lambda_1$) بیشترین
- محور سوم ($\mathbf{e}_3 - \lambda_3$) کمترین
- بیضی داخلی (مثلاً (68%))
- بیضی خارجی (مثلاً (95%))

شکل ۱: تعبیر هندسی مؤلفه‌های اصلی برای ماتریس کوواریانس Σ . محورهاهای اصلی بیضی در جهت بردارهای ویژه قرار دارند و طول نیم‌محورهاهای آن‌ها متناسب با جذر مقادیر ویژه است.

همان‌گونه که در شکل ۱ مشاهده می‌شود، بردار ویژه متناظر با بزرگترین مقدار ویژه، جهت بلندترین محور بیضی را مشخص می‌کند؛ بنابراین این جهت همان جهت مؤلفه اصلی اول است.

۵ مؤلفه‌های اصلی متغیرهای استاندارد شده

در بسیاری از مسائل، متغیرها دارای واحدهای اندازه‌گیری یا مقیاس‌های بسیار متفاوت‌اند. در این حالت، استفاده مستقیم از ماتریس کوواریانس ممکن است باعث شود متغیری که واریانس بیشتری دارد، صرفاً به دلیل مقیاس بزرگتر، نقش اصلی را در مؤلفه‌ها ایفا کند.

برای حذف اثر مقیاس، متغیرها را استاندارد می‌کنیم.

فرض کنید:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}, \quad E(\mathbf{X}) = \boldsymbol{\mu}, \quad \text{Cov}(\mathbf{X}) = \Sigma.$$

متغیرهای استاندارد شده به صورت زیر تعریف می‌شوند:

$$Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}, \quad i = 1, 2, \dots, p.$$

در نمایش برداری:

$$\mathbf{Z} = V^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$$

که در آن:

$$V = \text{diag}(\sigma_{11}, \sigma_{22}, \dots, \sigma_{pp}).$$

۱.۵ ماتریس کواریانس متغیرهای استاندارد شده

داریم:

$$\begin{aligned} \text{Cov}(\mathbf{Z}) &= \text{Cov}[V^{-1/2}(\mathbf{X} - \boldsymbol{\mu})] \\ &= V^{-1/2} \Sigma V^{-1/2}. \end{aligned}$$

ماتریس

$$V^{-1/2} \Sigma V^{-1/2}$$

همان ماتریس همبستگی جامعه است که آن را با ρ نشان می‌دهیم. پس:

$$\text{Cov}(\mathbf{Z}) = \rho$$

۲.۵ تعریف مؤلفه‌های اصلی استاندارد شده

فرض کنید زوج‌های مقدار ویژه و بردار ویژه ماتریس همبستگی ρ عبارت باشند از:

$$(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p),$$

به طوری که:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0.$$

مؤلفه اصلی i ام متغیرهای استاندارد شده برابر است با:

$$Y_i = \mathbf{e}_i' \mathbf{Z} = \mathbf{e}_i' V^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$$

و واریانس آن:

$$\text{Var}(Y_i) = \lambda_i$$

همچنین برای $i \neq j$:

$$\text{Cov}(Y_i, Y_j) = 0$$

۳.۵ واریانس کل متغیرهای استاندارد شده

چون واریانس هر متغیر استاندارد شده برابر یک است:

$$\text{Var}(Z_i) = 1.$$

پس واریانس کل برابر است با:

$$\sum_{i=1}^p \text{Var}(Z_i) = p.$$

از طرف دیگر مجموع مقادیر ویژه ماتریس همبستگی برابر اثر ماتریس است:

$$\sum_{i=1}^p \lambda_i = \text{tr}(\rho).$$

چون تمام عناصر قطر اصلی ماتریس همبستگی برابر یک هستند:

$$\text{tr}(\rho) = p.$$

بنابراین:

$$\sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \lambda_i = p$$

۴.۵ سهم واریانس هر مؤلفه استاندارد شده

سهم مؤلفه اصلی k ام از واریانس کل متغیرهای استاندارد شده برابر است با:

$$\frac{\lambda_k}{p}$$

و درصد واریانس تبیین شده برابر است با:

$$100 \frac{\lambda_k}{p}$$

سهم تجمعی m مؤلفه اول برابر است با:

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_m}{p}$$

۵.۵ همبستگی مؤلفه‌های استاندارد شده با متغیرها

برای متغیرهای استاندارد شده داریم:

$$\text{Var}(Z_k) = 1.$$

بنابراین فرمول همبستگی ساده‌تر می‌شود:

$$\rho_{Y_i, Z_k} = e_{ki} \sqrt{\lambda_i}$$

این ضرایب برای تفسیر مؤلفه‌ها بسیار مفیدند. هرچه قدر مطلق

$$e_{ki} \sqrt{\lambda_i}$$

بزرگ‌تر باشد، متغیر استاندارد شده Z_k ارتباط بیشتری با مؤلفه Y_i دارد.

۶ مثال: مقایسه مؤلفه‌های اصلی بر اساس کوواریانس و همبستگی

فرض کنید ماتریس کوواریانس دو متغیر X_1 و X_2 برابر باشد با:

$$\Sigma = \begin{pmatrix} 1 & 4 \\ 4 & 100 \end{pmatrix}.$$

در این مثال:

$$\text{Var}(X_1) = 1, \quad \text{Var}(X_2) = 100,$$

بنابراین مقیاس تغییرات X_2 به مراتب بزرگ‌تر از X_1 است.

۱.۶ مؤلفه‌های اصلی بر اساس ماتریس کوواریانس

معادله مشخصه عبارت است از:

$$|\Sigma - \lambda I| = 0.$$

داریم:

$$\begin{vmatrix} 1 - \lambda & 4 \\ 4 & 100 - \lambda \end{vmatrix} = 0.$$

بنابراین:

$$(1 - \lambda)(100 - \lambda) - 16 = 0,$$

$$100 - 101\lambda + \lambda^2 - 16 = 0,$$

$$\lambda^2 - 101\lambda + 84 = 0.$$

پس:

$$\lambda = \frac{101 \pm \sqrt{101^2 - 4(84)}}{2}.$$

چون:

$$101^2 - 336 = 9865,$$

داریم:

$$\lambda = \frac{101 \pm \sqrt{9865}}{2}.$$

در نتیجه مقادیر تقریبی ویژه عبارت‌اند از:

$$\lambda_1 \approx 100.16, \quad \lambda_2 \approx 0.84$$

بردارهای ویژه واحد متناظر، به طور تقریبی، عبارت‌اند از:

$$\mathbf{e}_1 \approx \begin{pmatrix} 0.040 \\ 0.999 \end{pmatrix}, \quad \mathbf{e}_2 \approx \begin{pmatrix} 0.999 \\ -0.040 \end{pmatrix}.$$

بنابراین مؤلفه‌های اصلی مبتنی بر ماتریس کوواریانس عبارت‌اند از:

$$Y_1 = 0.040X_1 + 0.999X_2$$

و

$$Y_2 = 0.999X_1 - 0.040X_2$$

واریانس کل برابر است با:

$$\text{tr}(\Sigma) = 1 + 100 = 101.$$

سهم مؤلفه اصلی اول برابر است با:

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{100.16}{101} \approx 0.992.$$

بنابراین:

$$\text{از واریانس کل را توضیح می‌دهد. } 99.2\% \text{ مؤلفه اول حدود}$$

مشاهده می‌شود که مؤلفه اول تقریباً به طور کامل تحت تأثیر متغیر X_2 قرار گرفته است. علت این امر آن است که واریانس X_2 برابر 100 و واریانس X_1 برابر 1 است.

۲.۶ ساخت ماتریس همبستگی

ضریب همبستگی میان X_1 و X_2 برابر است با:

$$\begin{aligned}\rho_{12} &= \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} \\ &= \frac{4}{\sqrt{1 \times 100}} \\ &= \frac{4}{10} \\ &= 0.4.\end{aligned}$$

بنابراین ماتریس همبستگی برابر است با:

$$\rho = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$$

۳.۶ مقادیر ویژه ماتریس همبستگی

معادله مشخصه:

$$|\rho - \lambda I| = 0.$$

داریم:

$$\begin{vmatrix} 1 - \lambda & 0.4 \\ 0.4 & 1 - \lambda \end{vmatrix} = 0.$$

در نتیجه:

$$(1 - \lambda)^2 - 0.16 = 0.$$

پس:

$$1 - \lambda = \pm 0.4.$$

بنابراین:

$$\lambda_1 = 1.4, \quad \lambda_2 = 0.6$$

بردارهای ویژه واحد متناظر عبارت‌اند از:

$$\mathbf{e}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0.707 \\ 0.707 \end{pmatrix}$$

و

$$\mathbf{e}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \approx \begin{pmatrix} 0.707 \\ -0.707 \end{pmatrix}$$

۴.۶ استانداردسازی متغیرها

متغیرهای استاندارد شده عبارت اند از:

$$Z_1 = \frac{X_1 - \mu_1}{\sqrt{1}} = X_1 - \mu_1,$$

و

$$Z_2 = \frac{X_2 - \mu_2}{\sqrt{100}} = \frac{X_2 - \mu_2}{10}.$$

۵.۶ مؤلفه‌های اصلی مبتنی بر ماتریس همبستگی

مؤلفه اصلی اول:

$$\begin{aligned} Y_1 &= 0.707Z_1 + 0.707Z_2 \\ &= 0.707(X_1 - \mu_1) + 0.707\left(\frac{X_2 - \mu_2}{10}\right) \\ &= 0.707(X_1 - \mu_1) + 0.0707(X_2 - \mu_2). \end{aligned}$$

پس:

$$Y_1 = 0.707(X_1 - \mu_1) + 0.0707(X_2 - \mu_2)$$

مؤلفه اصلی دوم:

$$\begin{aligned} Y_2 &= 0.707Z_1 - 0.707Z_2 \\ &= 0.707(X_1 - \mu_1) - 0.707\left(\frac{X_2 - \mu_2}{10}\right) \\ &= 0.707(X_1 - \mu_1) - 0.0707(X_2 - \mu_2). \end{aligned}$$

پس:

$$Y_2 = 0.707(X_1 - \mu_1) - 0.0707(X_2 - \mu_2)$$

۶.۶ سهم واریانس مؤلفه‌ها

چون تعداد متغیرهای استاندارد شده برابر $p = 2$ است، واریانس کل برابر است با:

$$p = 2.$$

پس سهم مؤلفه اول برابر است با:

$$\frac{\lambda_1}{p} = \frac{1.4}{2} = 0.7.$$

بنابراین:

از واریانس کل را توضیح می‌دهد. 70% مؤلفه اول

همچنین:

$$\frac{\lambda_2}{p} = \frac{0.6}{2} = 0.3.$$

بنابراین:

از واریانس کل را توضیح می‌دهد. 30% مؤلفه دوم

۷.۶ همبستگی مؤلفه اول با متغیرهای استاندارد شده

برای مؤلفه‌های حاصل از ماتریس همبستگی:

$$\rho_{Y_i, Z_k} = e_{ki} \sqrt{\lambda_i}.$$

بنابراین:

$$\begin{aligned}\rho_{Y_1, Z_1} &= 0.707\sqrt{1.4} \\ &\approx 0.707(1.183) \\ &\approx 0.837.\end{aligned}$$

پس:

$$\rho_{Y_1, Z_1} \approx 0.837$$

همچنین:

$$\begin{aligned}\rho_{Y_1, Z_2} &= 0.707\sqrt{1.4} \\ &\approx 0.837.\end{aligned}$$

پس:

$$\rho_{Y_1, Z_2} \approx 0.837$$

بنابراین دو متغیر استاندارد شده نقش یکسانی در مؤلفه اصلی اول دارند.

۸.۶ مقایسه دو روش

در تحلیل مبتنی بر ماتریس کوواریانس، مؤلفه اصلی اول تقریباً برابر با X_2 شد:

$$Y_1 \approx 0.040X_1 + 0.999X_2.$$

این نتیجه بیشتر ناشی از واریانس بسیار بزرگ X_2 است.
اما پس از استانداردسازی، دو متغیر در مؤلفه اصلی اول وزن یکسانی دارند:

$$Y_1 = 0.707Z_1 + 0.707Z_2.$$

بنابراین استانداردسازی سبب می‌شود اثر واحد اندازه‌گیری و تفاوت مقیاس متغیرها حذف شود.
نتیجه کلی این است که:

- اگر متغیرها واحد و مقیاس مشابه داشته باشند، استفاده از ماتریس کوواریانس مناسب است؛
- اگر متغیرها واحدهای متفاوت یا واریانس‌های بسیار نامساوی داشته باشند، استفاده از ماتریس همبستگی معمولاً مناسب‌تر است؛
- مؤلفه‌های حاصل از ماتریس کوواریانس و ماتریس همبستگی لزوماً یکسان نیستند؛
- استانداردسازی یک تبدیل ساده و بی‌اثر نیست، بلکه می‌تواند ساختار مؤلفه‌های اصلی را به طور اساسی تغییر دهد.

۷ تحلیل مؤلفه‌های اصلی مثال سه‌متغیره با نرم‌افزار R

در این مثال، ماتریس کوواریانس سه متغیر به صورت زیر است:

$$\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

ابتدا ماتریس کوواریانس را در نرم‌افزار R تعریف می‌کنیم.

Listing 1:

```

1 Sigma <- matrix(
2   c( 1, -2,  0,
3     -2,  5,  0,
4       0,  0,  2),
5   nrow = 3,
6   byrow = TRUE
7 )
8
9 rownames(Sigma) <- c("X1", "X2", "X3")
10 colnames(Sigma) <- c("X1", "X2", "X3")
11
12 Sigma

```

Listing 2:

	X1	X2	X3
X1	1	-2	0
X2	-2	5	0
X3	0	0	2

۱.۷ مقادیر ویژه و بردارهای ویژه

برای محاسبه مقادیر و بردارهای ویژه از تابع `eigen` استفاده می‌کنیم.

Listing 3:

```

1 eig <- eigen(Sigma, symmetric = TRUE)
2
3 lambda <- eig$values
4 E <- eig$vectors
5
6 names(lambda) <- c("lambda1", "lambda2", "lambda3")
7 colnames(E) <- c("e1", "e2", "e3")
8 rownames(E) <- c("X1", "X2", "X3")
9
10 round(lambda, 6)
11 round(E, 6)

```

خروجی مقادیر ویژه:

Listing 4:

lambda1	lambda2	lambda3
5.828427	2.000000	0.171573

یکی از خروجی‌های ممکن برای بردارهای ویژه برابر است با:

Listing 5:

R

	e1	e2	e3
X1	-0.382683	0.000000	-0.923880
X2	0.923880	0.000000	-0.382683
X3	0.000000	1.000000	0.000000

علامت بردارهای ویژه اختیاری است. برای هماهنگی با محاسبات دستی جزوه، علامت بردارهای اول و سوم را تغییر می‌دهیم:

Listing 6:

```

1  if (E[1, 1] < 0) {
2      E[, 1] <- -E[, 1]
3  }
4
5  if (E[1, 3] < 0) {
6      E[, 3] <- -E[, 3]
7  }
8
9  round(E, 6)

```

خروجی:

Listing 7:

	e1	e2	e3
X1	0.382683	0.000000	0.923880
X2	-0.923880	0.000000	0.382683
X3	0.000000	1.000000	0.000000

بنابراین مؤلفه‌های اصلی عبارت‌اند از:

$$Y_1 = 0.383X_1 - 0.924X_2,$$

$$Y_2 = X_3,$$

و

$$Y_3 = 0.924X_1 + 0.383X_2.$$

۲.۷ بررسی قطری شدن ماتریس کوواریانس

ماتریس کوواریانس مؤلفه‌های اصلی از رابطه

$$\text{Cov}(\mathbf{Y}) = \mathbf{E}'\Sigma\mathbf{E}$$

به دست می‌آید.

Listing 8:

```

1  Cov_Y <- t(E) %*% Sigma %*% E
2
3  round(Cov_Y, 6)

```

خروجی:

Listing 9:

	e1	e2	e3
e1	5.828427	0.000000	0.000000
e2	0.000000	2.000000	0.000000
e3	0.000000	0.000000	0.171573

درایه‌های خارج از قطر صفر هستند؛ بنابراین مؤلفه‌های اصلی با یکدیگر ناهمبسته‌اند.

۳.۷ سهم واریانس مؤلفه‌ها

Listing 10:

```

1 total_variance <- sum(lambda)
2
3 variance_share <- lambda / total_variance
4 cumulative_share <- cumsum(variance_share)
5
6 variance_table <- data.frame(
7   Component = c("Y1", "Y2", "Y3"),
8   Eigenvalue = lambda,
9   Variance_Percent = 100 * variance_share,
10  Cumulative_Percent = 100 * cumulative_share
11 )
12
13 print(
14   transform(
15     variance_table,
16     Eigenvalue = round(Eigenvalue, 3),
17     Variance_Percent = round(Variance_Percent, 2),
18     Cumulative_Percent = round(Cumulative_Percent, 2)
19   ),
20   row.names = FALSE
21 )

```

خروجی:

Listing 11:

Component	Eigenvalue	Variance_Percent	Cumulative_Percent
Y1	5.828	72.86	72.86
Y2	2.000	25.00	97.86
Y3	0.172	2.14	100.00

۴.۷ همبستگی متغیرها با مؤلفه‌های اصلی

Listing 12:

```
sd_X <- sqrt(diag(Sigma))

Cor_X_PC <- diag(1 / sd_X) %*%
            E %*%
            diag(sqrt(lambda))

rownames(Cor_X_PC) <- c("X1", "X2", "X3")
colnames(Cor_X_PC) <- c("Y1", "Y2", "Y3")

round(Cor_X_PC, 3)
```

خروجی:

Listing 13:

	Y1	Y2	Y3
X1	0.924	0.000	0.383
X2	-0.998	0.000	0.071
X3	0.000	1.000	0.000

تفاوت کوچک عدد 0.924 با مقدار 0.925 کتاب ناشی از گرد کردن بردار ویژه و مقدار ویژه است.

۸ رسم بیضی مؤلفه‌های اصلی با بسته ellipse

بسته ellipse می‌تواند بیضی متناظر با یک ماتریس کوواریانس را مستقیماً رسم کند. در این مثال، چون مؤلفه دوم دقیقاً برابر X_3 است، برای نمایش هندسی از زیرماتریس مربوط به X_1 و X_2 استفاده می‌کنیم:

$$\Sigma_{12} = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}.$$

در صورت نصب نبودن بسته، دستور زیر فقط یک بار اجرا می‌شود:

Listing 14: ellipse

```
install.packages("ellipse")
```

سپس بسته را فراخوانی می‌کنیم:

Listing 15: ellipse

```
library(ellipse)
```

۱.۸ رسم بیضی احتمال ۹۵ درصد

در تابع ellipse، اگر آرگومان $level = 95.0$ تعیین شود، بیضی احتمال ۹۵٪ متناظر با توزیع نرمال دو متغیره رسم می‌شود.

Listing 16: ellipse

```
1 library(ellipse)
2
3 Sigma_12 <- Sigma[1:2, 1:2]
4 mu_12 <- c(0, 0)
5
6 ellipse_95 <- ellipse::ellipse(
7   Sigma_12,
8   centre = mu_12,
9   level = 0.95,
10  npoints = 500
11 )
12
13 plot(
14   ellipse_95,
15   type = "l",
16   asp = 1,
17   lwd = 2,
18   xlab = expression(X[1]),
19   ylab = expression(X[2]),
20   main = "95% probability ellipse"
21 )
22
23 abline(h = 0, v = 0, lty = 3)
24 points(0, 0, pch = 19)
```

تابع ellipse یک ماتریس دو ستونی تولید می‌کند. ستون اول شامل مختصات X_1 و ستون دوم شامل مختصات X_2 نقاط روی بیضی است. برای مشاهده چند سطر نخست:

Listing 17:

```
head(ellipse_95)
```

یک خروجی نمونه به صورت زیر است؛ مقادیر دقیق ممکن است با توجه به نقطه شروع پارامتری تابع اندکی متفاوت باشند:

Listing 18:

	x	y
[1,]	-0.9360	4.9058
[2,]	-0.9973	4.9259
[3,]	-1.0584	4.9452
[4,]	-1.1192	4.9637
[5,]	-1.1798	4.9815
[6,]	-1.2401	4.9984

۲.۸ افزودن محوره‌های مؤلفه‌های اصلی به بیضی

برای رسم محوره‌های اصلی، ابتدا تجزیه ویژه زیرماتریس Σ_{12} را انجام می‌دهیم.

Listing 19:

```

1 eig_12 <- eigen(Sigma_12, symmetric = TRUE)
2
3 lambda_12 <- eig_12$values
4 E_12 <- eig_12$vectors
5
6 # Adjust signs to match the hand calculations
7 if (E_12[1, 1] < 0) {
8   E_12[, 1] <- -E_12[, 1]
9 }
10
11 if (E_12[1, 2] < 0) {
12   E_12[, 2] <- -E_12[, 2]
13 }
14
15 round(lambda_12, 6)
16 round(E_12, 6)

```

خروجی:

Listing 20:

[1]	5.828427	0.171573
	[,1]	[,2]
[1,]	0.382683	0.923880

```
[2,] -0.923880 0.382683
```

برای بیضی 95٪، مقدار ثابت برابر است با:

$$c = \sqrt{\chi_{2,0.95}^2}.$$

Listing 21:

```
1 c2 <- qchisq(0.95, df = 2)
2 c_value <- sqrt(c2)
3
4 axis_lengths <- c_value * sqrt(lambda_12)
5
6 round(c2, 6)
7 round(c_value, 6)
8 round(axis_lengths, 6)
```

خروجی:

Listing 22:

```
[1] 5.991465
[1] 2.447747
[1] 5.908846 1.013049
```

بنابراین طول نیم‌محور بزرگ تقریباً 5.91 و طول نیم‌محور کوچک تقریباً 1.01 است.

۳.۸ کد نهایی بیضی همراه با محورهای مؤلفه‌ها

Listing 23:

ellipse

```
1 library(ellipse)
2
3 # Covariance matrix of X1 and X2
4 Sigma_12 <- Sigma[1:2, 1:2]
5 mu_12 <- c(0, 0)
6
7 # Eigenvalue decomposition
8 eig_12 <- eigen(Sigma_12, symmetric = TRUE)
9 lambda_12 <- eig_12$values
10 E_12 <- eig_12$vectors
```

```

11
12 # Sign convention used in the notes
13 if (E_12[1, 1] < 0) {
14   E_12[, 1] <- -E_12[, 1]
15 }
16
17 if (E_12[1, 2] < 0) {
18   E_12[, 2] <- -E_12[, 2]
19 }
20
21 # 95% probability ellipse
22 ellipse_95 <- ellipse::ellipse(
23   Sigma_12,
24   centre = mu_12,
25   level = 0.95,
26   npoints = 500
27 )
28
29 # Constant and semi-axis lengths
30 c_value <- sqrt(qchisq(0.95, df = 2))
31
32 axis1 <- c_value * sqrt(lambda_12[1]) * E_12[, 1]
33 axis3 <- c_value * sqrt(lambda_12[2]) * E_12[, 2]
34
35 # Plot
36 plot(
37   ellipse_95,
38   type = "l",
39   lwd = 2,
40   asp = 1,
41   xlab = expression(X[1]),
42   ylab = expression(X[2]),
43   main = "Principal axes and 95% probability ellipse",
44   xlim = range(ellipse_95[, 1]) + c(-0.7, 0.7),
45   ylim = range(ellipse_95[, 2]) + c(-0.7, 0.7)
46 )
47
48 # Original coordinate axes
49 abline(h = 0, v = 0, lty = 3)
50
51 # First principal axis
52 arrows(

```

```

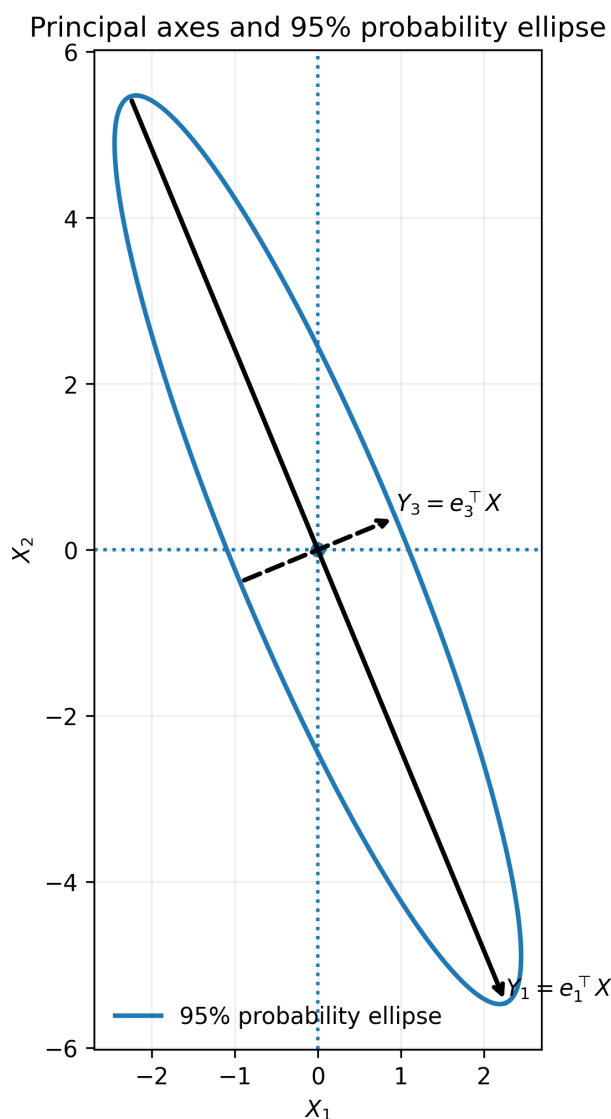
53 -axis1[1], -axis1[2],
54     axis1[1], axis1[2],
55     length = 0.10,
56     lwd = 2
57 )
58
59 # Third principal axis
60 arrows(
61     -axis3[1], -axis3[2],
62     axis3[1], axis3[2],
63     length = 0.10,
64     lwd = 2,
65     lty = 2
66 )
67
68 # Centre
69 points(mu_12[1], mu_12[2], pch = 19)
70
71 # Axis labels
72 text(
73     axis1[1],
74     axis1[2],
75     labels = expression(Y[1] == e[1]^T * X),
76     pos = 4
77 )
78
79 text(
80     axis3[1],
81     axis3[2],
82     labels = expression(Y[3] == e[3]^T * X),
83     pos = 4
84 )
85
86 legend(
87     "topright",
88     legend = c(
89         "95% ellipse",
90         "First principal axis",
91         "Third principal axis"
92     ),
93     lty = c(1, 1, 2),
94     lwd = c(2, 2, 2),

```

```

95   bty = "n"
96 )

```



شکل ۲: بیضی احتمال 95% و محورهای مؤلفه‌های اصلی برای متغیرهای X_1 و X_2 . محور بلند در جهت بردار ویژه متناظر با λ_1 و محور کوتاه در جهت بردار ویژه متناظر با λ_3 قرار دارد.

۴.۸ رسم چند بیضی هم‌چگالی

همانند کد فایل پیوست، می‌توان چند بیضی با سطوح احتمال متفاوت رسم کرد. این بیضی‌ها جهت‌های یکسانی دارند، ولی اندازه آن‌ها متفاوت است.

Listing 24:

```

1 ellipse_50 <- ellipse::ellipse(
2   Sigma_12,
3   centre = mu_12,

```

```

4   level = 0.50,
5   npoints = 500
6 )
7
8 ellipse_90 <- ellipse::ellipse(
9   Sigma_12,
10  centre = mu_12,
11  level = 0.90,
12  npoints = 500
13 )
14
15 ellipse_95 <- ellipse::ellipse(
16   Sigma_12,
17   centre = mu_12,
18   level = 0.95,
19   npoints = 500
20 )
21
22 ellipse_99 <- ellipse::ellipse(
23   Sigma_12,
24   centre = mu_12,
25   level = 0.99,
26   npoints = 500
27 )
28
29 plot(
30   ellipse_99,
31   type = "l",
32   asp = 1,
33   lwd = 2,
34   xlab = expression(X[1]),
35   ylab = expression(X[2]),
36   main = "Probability ellipses"
37 )
38
39 lines(ellipse_95, lwd = 2, lty = 2)
40 lines(ellipse_90, lwd = 2, lty = 3)
41 lines(ellipse_50, lwd = 2, lty = 4)
42
43 abline(h = 0, v = 0, lty = 3)
44
45 legend(

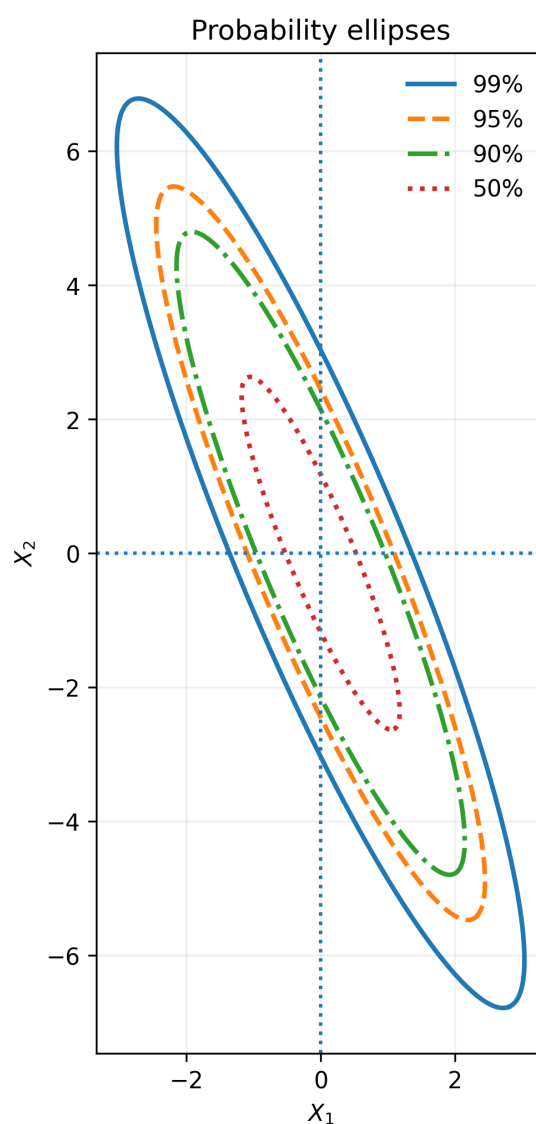
```

```

46 "topright",
47 legend = c("99%", "95%", "90%", "50%"),
48 lty = c(1, 2, 3, 4),
49 lwd = 2,
50 bty = "n"
51 )

```

تمام این بیضی‌ها در جهت بردارهای ویژه e_1 و e_3 قرار دارند. تغییر سطح احتمال تنها طول نیم‌محورها را تغییر می‌دهد.



شکل ۳: بیضی احتمال 95% و محورهای مؤلفه‌های اصلی برای متغیرهای X_1 و X_2 . محور بلند در جهت بردار ویژه متناظر با λ_1 و محور کوتاه در جهت بردار ویژه متناظر با λ_3 قرار دارد.

۹ رسم داده‌های شبیه‌سازی شده همراه با بیضی‌ها

برای نمایش ابر نقاط، ابتدا از توزیع نرمال سه‌متغیره داده تولید می‌کنیم.

Listing 25:

```
1 library(MASS)
2 library(ellipse)
3
4 set.seed(1405)
5
6 X_sim <- MASS::mvrnorm(
7   n = 500,
8   mu = c(0, 0, 0),
9   Sigma = Sigma
10 )
11
12 colnames(X_sim) <- c("X1", "X2", "X3")
13
14 head(X_sim)
```

چند سطر نخست داده‌های شبیه‌سازی شده، با توجه به بذر تصادفی تعیین شده، به صورت زیر خواهد بود:

Listing 26:

	X1	X2	X3
[1,]	0.482...	-0.914...	0.203...
[2,]	-0.316...	1.786...	-0.725...
[3,]	0.171...	-0.642...	1.388...
[4,]	-1.102...	2.417...	-1.536...
[5,]	0.626...	-1.865...	0.492...
[6,]	-0.084...	0.771...	-0.227...

نکته: اعداد بالا برای نمایش ساختار خروجی هستند. خروجی دقیق اجرای R با بذر مشخص شده باید مستقیماً از نرم‌افزار کپی شود.

۱.۹ رسم ابر نقاط و بیضی نظری

Listing 27:

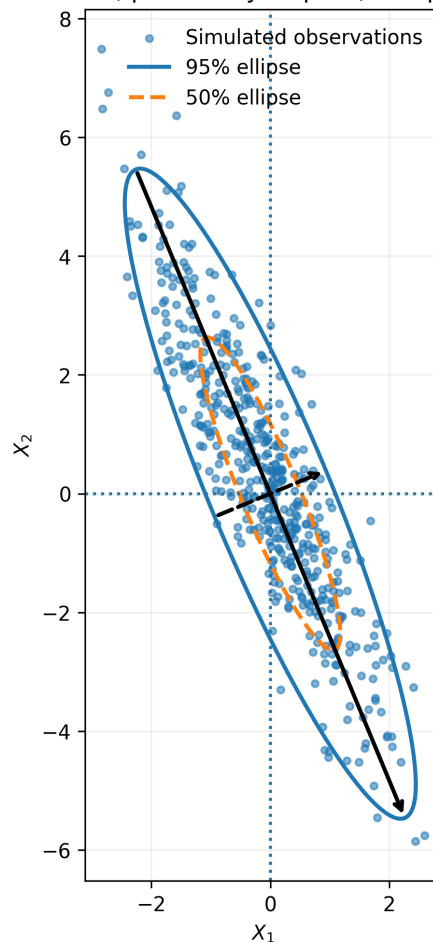
```
1 plot(
2   X_sim[, "X1"],
3   X_sim[, "X2"],
4   pch = 20,
```

```

5   cex = 0.55,
6   asp = 1,
7   xlab = expression(X[1]),
8   ylab = expression(X[2]),
9   main = "Simulated data and theoretical ellipses"
10  )
11
12  lines(ellipse_95, lwd = 2)
13  lines(ellipse_50, lwd = 2, lty = 2)
14
15  abline(h = 0, v = 0, lty = 3)
16
17  arrows(
18    -axis1[1], -axis1[2],
19     axis1[1],  axis1[2],
20    length = 0.10,
21    lwd = 2
22  )
23
24  arrows(
25    -axis3[1], -axis3[2],
26     axis3[1],  axis3[2],
27    length = 0.10,
28    lwd = 2,
29    lty = 2
30  )
31
32  legend(
33    "topright",
34    legend = c(
35      "Observations",
36      "95% ellipse",
37      "50% ellipse",
38      "First principal axis",
39      "Third principal axis"
40    ),
41    pch = c(20, NA, NA, NA, NA),
42    lty = c(NA, 1, 2, 1, 2),
43    lwd = c(NA, 2, 2, 2, 2),
44    bty = "n"
45  )

```

Simulated data, probability ellipses, and principal axes



شکل ۴: بیضی احتمال ۹۵٪ و محورهای مؤلفه‌های اصلی برای متغیرهای X_1 و X_2 . محور بلند در جهت بردار ویژه متناظر با λ_1 و محور کوتاه در جهت بردار ویژه متناظر با λ_3 قرار دارد.

۲.۹ رسم بیضی نمونه‌ای با تابع dataEllipse

در فایل پیوست، از تابع dataEllipse بسته car نیز استفاده شده است. این تابع بیضی را بر اساس داده‌های نمونه‌ای رسم می‌کند، در حالی که تابع ellipse می‌تواند مستقیماً از ماتریس کوواریانس نظری استفاده کند. در صورت نصب نبودن بسته car:

Listing 28: car

```
install.packages("car")
library(car)
```

رسم بیضی‌های نمونه‌ای:

Listing 29: dataEllipse

```
library(car)
```

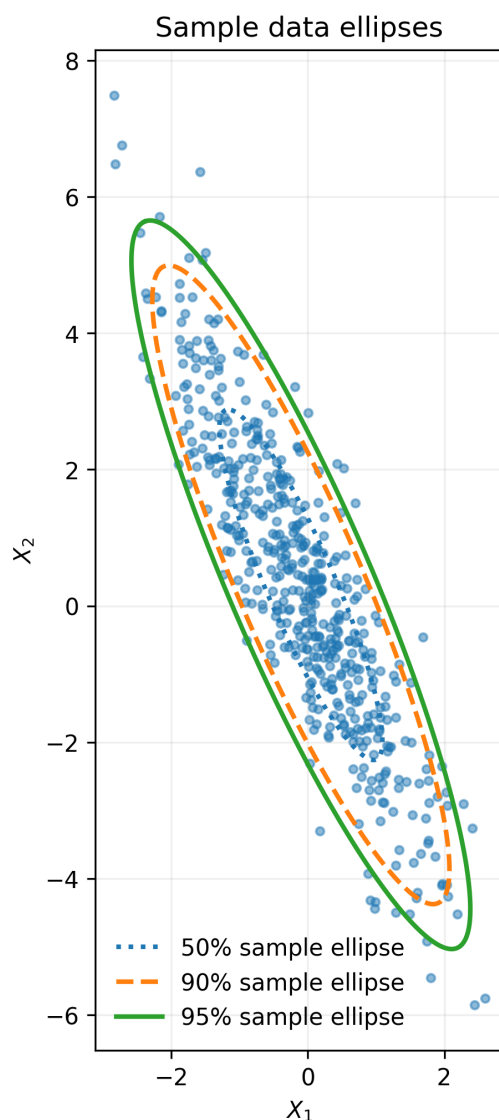
```

3 car::dataEllipse(
4   X_sim[, "X1"],
5   X_sim[, "X2"],
6   levels = c(0.50, 0.90, 0.95),
7   ellipse.label = c("50%", "90%", "95%"),
8   fill = TRUE,
9   fill.alpha = 0.08,
10  pch = 20,
11  cex = 0.5,
12  xlab = "X1",
13  ylab = "X2"
14 )
15
16 title("Sample data ellipses")

```

تفاوت دو روش به صورت زیر است:

- تابع `ellipse(Sigma_12)`, ... از ماتریس کوواریانس نظری استفاده می‌کند؛
- تابع `ydataEllipse(x)`, ... ماتریس کوواریانس و میانگین را از داده‌های نمونه‌ای برآورد می‌کند؛
- با افزایش حجم نمونه، بیضی نمونه‌ای به بیضی نظری نزدیک‌تر می‌شود.



شکل ۵: بیضی احتمال ۹۵٪ و محورهای مؤلفه‌های اصلی برای متغیرهای X_1 و X_2 . محور بلند در جهت بردار ویژه متناظر با λ_1 و محور کوتاه در جهت بردار ویژه متناظر با λ_3 قرار دارد.

۳.۹ مؤلفه‌های اصلی بر اساس ماتریس واریانس-کوواریانس نمونه

در بخش‌های قبل، مؤلفه‌های اصلی بر اساس ماتریس واریانس-کوواریانس جامعه Σ تعریف شدند. در کاربردهای واقعی معمولاً Σ ناشناخته است و تنها یک نمونه تصادفی از جامعه در اختیار داریم. بنابراین، به جای ماتریس واریانس-کوواریانس جامعه، از ماتریس واریانس-کوواریانس نمونه استفاده می‌کنیم.

فرض کنید n مشاهده از p متغیر $X_1, X_2, \dots, X_p$ در اختیار داریم و ماتریس داده‌ها به صورت

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

باشد. بردار میانگین نمونه برابر است با

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix},$$

که در آن

$$\bar{x}_j = \frac{1}{n} \sum_{r=1}^n x_{rj}, \quad j = 1, 2, \dots, p.$$

ماتریس واریانس-کوواریانس نمونه به صورت

$$\mathbf{S} = [s_{jk}]_{p \times p}$$

تعریف می‌شود که در آن

$$s_{jk} = \frac{1}{n-1} \sum_{r=1}^n (x_{rj} - \bar{x}_j)(x_{rk} - \bar{x}_k).$$

بنابراین،

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix},$$

که عناصر قطر اصلی آن یعنی $s_{11}, s_{22}, \dots, s_{pp}$ واریانس‌های نمونه‌ای متغیرها و عناصر خارج از قطر، کوواریانس‌های نمونه‌ای بین متغیرها هستند.

۱.۳.۹ ایده اصلی

از نظر محاسباتی، تحلیل مؤلفه‌های اصلی نمونه تقریباً همان تحلیل مؤلفه‌های اصلی جامعه است. تنها تفاوت اصلی این است که در حالت جامعه از Σ استفاده می‌کنیم، در حالی که در حالت نمونه ماتریس $\mathbf{S}$ جایگزین آن می‌شود.

بنابراین، اگر زوج‌های مقدار ویژه-بردار ویژه ماتریس $\mathbf{S}$ را به صورت

$$(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \dots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$$

در نظر بگیریم، داریم

$$S\hat{e}_i = \hat{\lambda}_i \hat{e}_i, \quad i = 1, 2, \dots, p,$$

و مقادیر ویژه را به ترتیب نزولی مرتب می‌کنیم:

$$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0.$$

بردارهای ویژه نیز به صورت واحد انتخاب می‌شوند، یعنی

$$\hat{e}_i' \hat{e}_i = 1,$$

و برای $i \neq j$ داریم

$$\hat{e}_i' \hat{e}_j = 0.$$

بنابراین بردارهای ویژه ماتریس S یک دستگاه متعامد یکه تشکیل می‌دهند.

۲.۳.۹ تعریف مؤلفه‌های اصلی نمونه

اگر

$$\hat{e}_i = \begin{bmatrix} \hat{e}_{1i} \\ \hat{e}_{2i} \\ \vdots \\ \hat{e}_{pi} \end{bmatrix}$$

بردار ویژه متناظر با مقدار ویژه $\hat{\lambda}_i$ باشد، مؤلفه اصلی نمونه‌ای i ام به صورت

$$\hat{Y}_i = \hat{e}_i' \mathbf{x}$$

تعریف می‌شود. در نتیجه،

$$\hat{Y}_i = \hat{e}_{1i}x_1 + \hat{e}_{2i}x_2 + \dots + \hat{e}_{pi}x_p$$

برای

$$i = 1, 2, \dots, p.$$

بنابراین مؤلفه‌های اصلی نمونه‌ای عبارت‌اند از

$$\hat{Y}_1 = \hat{\mathbf{e}}_1' \mathbf{x},$$

$$\hat{Y}_2 = \hat{\mathbf{e}}_2' \mathbf{x},$$

$\vdots$

$$\hat{Y}_p = \hat{\mathbf{e}}_p' \mathbf{x}.$$

توجه شود که $\hat{Y}_1$ در جهتی تعریف می‌شود که بیشترین تغییرپذیری نمونه را نشان می‌دهد. مؤلفه دوم $\hat{Y}_2$ بیشترین تغییرپذیری باقی‌مانده را در جهتی عمود بر مؤلفه اول نشان می‌دهد و این فرایند تا مؤلفه p ادامه می‌یابد.

۳.۳.۹ واریانس مؤلفه‌های اصلی نمونه

برای مؤلفه اصلی i ام داریم

$$\hat{Y}_i = \hat{\mathbf{e}}_i' \mathbf{x}.$$

واریانس نمونه‌ای آن برابر است با

$$s_{\hat{Y}_i}^2 = \hat{\mathbf{e}}_i' \mathbf{S} \hat{\mathbf{e}}_i.$$

ولی از تعریف مقدار ویژه داریم

$$\mathbf{S} \hat{\mathbf{e}}_i = \hat{\lambda}_i \hat{\mathbf{e}}_i.$$

بنابراین،

$$s_{\hat{Y}_i}^2 = \hat{\mathbf{e}}_i' \left(\hat{\lambda}_i \hat{\mathbf{e}}_i \right).$$

در نتیجه،

$$s_{\hat{Y}_i}^2 = \hat{\lambda}_i \hat{\mathbf{e}}_i' \hat{\mathbf{e}}_i.$$

از آنجا که بردارهای ویژه واحد هستند،

$$\hat{\mathbf{e}}_i' \hat{\mathbf{e}}_i = 1,$$

پس

$$s_{\hat{Y}_i}^2 = \hat{\lambda}_i$$

یا به عبارت دیگر،

$$\text{Var}_{\text{sample}}(\hat{Y}_i) = \hat{\lambda}_i$$

بنابراین مقدار ویژه $\hat{\lambda}_i$ دقیقاً واریانس نمونه‌ای مؤلفه اصلی i ام را نشان می‌دهد. به طور خاص،

$$s_{\hat{Y}_1}^2 = \hat{\lambda}_1, \quad s_{\hat{Y}_2}^2 = \hat{\lambda}_2, \quad \dots, \quad s_{\hat{Y}_p}^2 = \hat{\lambda}_p.$$

چون

$$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p,$$

مؤلفه اول دارای بیشترین واریانس و مؤلفه آخر دارای کمترین واریانس است.

۴.۳.۹ کوواریانس بین مؤلفه‌های اصلی نمونه

برای دو مؤلفه $\hat{Y}_i$ و $\hat{Y}_k$ داریم

$$\hat{Y}_i = \hat{\mathbf{e}}_i' \mathbf{x}, \quad \hat{Y}_k = \hat{\mathbf{e}}_k' \mathbf{x}.$$

کوواریانس نمونه‌ای آنها برابر است با

$$s_{\hat{Y}_i, \hat{Y}_k} = \hat{\mathbf{e}}_i' \mathbf{S} \hat{\mathbf{e}}_k.$$

از رابطه

$$\mathbf{S} \hat{\mathbf{e}}_k = \hat{\lambda}_k \hat{\mathbf{e}}_k$$

نتیجه می‌شود

$$s_{\hat{Y}_i, \hat{Y}_k} = \hat{\lambda}_k \hat{\mathbf{e}}_i' \hat{\mathbf{e}}_k.$$

برای $i \neq k$ ، بردارهای ویژه متعامدند، بنابراین

$$\hat{\mathbf{e}}_i' \hat{\mathbf{e}}_k = 0.$$

پس

$$s_{\hat{Y}_i, \hat{Y}_k} = 0, \quad i \neq k$$

است.

بنابراین مؤلفه‌های اصلی نمونه‌ای با یکدیگر همبستگی خطی ندارند.

این ویژگی یکی از مهم‌ترین دلایل استفاده از تحلیل مؤلفه‌های اصلی است؛ زیرا مجموعه‌ای از متغیرهای اولیه که ممکن است با یکدیگر همبستگی زیادی داشته باشند، به مجموعه‌ای از مؤلفه‌های بدون همبستگی تبدیل می‌شوند.

۵.۳.۹ نمایش ماتریسی مؤلفه‌های اصلی

ماتریس بردارهای ویژه را به صورت

$$\hat{\mathbf{E}} = \begin{bmatrix} \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 & \cdots & \hat{\mathbf{e}}_p \end{bmatrix}$$

در نظر می‌گیریم.

از متعامد بودن بردارهای ویژه داریم

$$\hat{\mathbf{E}}' \hat{\mathbf{E}} = \mathbf{I}.$$

بردار مؤلفه‌های اصلی نمونه‌ای را می‌توان به صورت

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_p \end{bmatrix}$$

نوشت. در نتیجه،

$$\hat{\mathbf{Y}} = \hat{\mathbf{E}}' \mathbf{x}$$

است.

اگر ماتریس قطری مقادیر ویژه را به صورت

$$\hat{\Lambda} = \text{diag} \left(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p \right)$$

تعریف کنیم، تجزیه طیفی ماتریس کوواریانس نمونه برابر است با

$$\mathbf{S} = \hat{\mathbf{E}} \hat{\mathbf{\Lambda}} \hat{\mathbf{E}}'$$

و همچنین

$$\hat{\mathbf{E}}' \mathbf{S} \hat{\mathbf{E}} = \hat{\mathbf{\Lambda}}.$$

بنابراین تبدیل مؤلفه‌های اصلی، ماتریس کوواریانس $\mathbf{S}$ را به یک ماتریس قطری تبدیل می‌کند.

۶.۳.۹ واریانس کل نمونه

واریانس کل داده‌های اولیه برابر مجموع واریانس‌های نمونه‌ای متغیرهای اصلی است:

$$s_{11} + s_{22} + \cdots + s_{pp}.$$

این مقدار همان اثر ماتریس $\mathbf{S}$ است:

$$\text{tr}(\mathbf{S}) = \sum_{j=1}^p s_{jj}.$$

از یکی از خواص مهم مقادیر ویژه می‌دانیم که اثر یک ماتریس برابر مجموع مقادیر ویژه آن است. بنابراین،

$$\text{tr}(\mathbf{S}) = \sum_{i=1}^p \hat{\lambda}_i.$$

در نتیجه،

$$\sum_{j=1}^p s_{jj} = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p$$

یا به عبارت دیگر،

$$\text{واریانس کل مؤلفه‌های اصلی} = \text{واریانس کل متغیرهای اولیه}$$

این رابطه نشان می‌دهد که تحلیل مؤلفه‌های اصلی واریانس کل داده‌ها را از بین نمی‌برد، بلکه فقط آن را بین محورهای جدید توزیع می‌کند.

۷.۳.۹ سهم هر مؤلفه اصلی از واریانس کل

چون واریانس مؤلفه اصلی i ام برابر $\hat{\lambda}_i$ است، سهم این مؤلفه از واریانس کل نمونه برابر است با

$$\frac{\hat{\lambda}_i}{\hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_p}$$

و درصد واریانس توضیح داده شده توسط مؤلفه i ام برابر است با

$$100 \frac{\hat{\lambda}_i}{\sum_{j=1}^p \hat{\lambda}_j}.$$

به طور خاص، برای مؤلفه اول داریم

$$\frac{\hat{\lambda}_1}{\sum_{j=1}^p \hat{\lambda}_j},$$

و برای دو مؤلفه اول داریم

$$\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{\sum_{j=1}^p \hat{\lambda}_j}.$$

به طور کلی، درصد تجمعی واریانس توضیح داده شده توسط m مؤلفه اول برابر است با

$$100 \frac{\sum_{i=1}^m \hat{\lambda}_i}{\sum_{i=1}^p \hat{\lambda}_i}.$$

این کمیت نقش مهمی در انتخاب تعداد مؤلفه‌های اصلی دارد. اگر چند مؤلفه اول درصد بزرگی از واریانس کل را توضیح دهند، می‌توان مؤلفه‌های باقی مانده را با از دست دادن مقدار اندکی از اطلاعات کنار گذاشت.

برای مثال، اگر

$$\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{\sum_{i=1}^p \hat{\lambda}_i} = 0.93,$$

آنگاه دو مؤلفه اول 93 درصد از تغییرپذیری کل داده‌ها را توضیح می‌دهند.

۸.۳.۹ همبستگی بین مؤلفه‌های اصلی و متغیرهای اولیه

تنها ضرایب موجود در بردار ویژه برای تفسیر مؤلفه‌های اصلی کافی نیستند. به‌خصوص هنگامی که واریانس متغیرهای اولیه با یکدیگر متفاوت باشد، استفاده از ضرایب همبستگی بین مؤلفه‌های اصلی و متغیرهای اولیه می‌تواند تفسیر مطمئن‌تری ارائه دهد.

برای مؤلفه اصلی i ام و متغیر X_k داریم

$$\hat{Y}_i = \hat{\mathbf{e}}'_i \mathbf{x}.$$

کوواریانس نمونه‌ای بین $\hat{Y}_i$ و X_k برابر است با

$$s_{\hat{Y}_i, X_k} = \text{Cov}_{\text{sample}}(\hat{Y}_i, X_k).$$

داریم

$$s_{\hat{Y}_i, X_k} = \hat{\mathbf{e}}'_i \mathbf{S} \mathbf{u}_k,$$

که $\mathbf{u}_k$ بردار واحد متناظر با متغیر k ام است.

از تقارن ماتریس $\mathbf{S}$ و رابطه مقدار ویژه داریم

$$\hat{\mathbf{e}}'_i \mathbf{S} = \hat{\lambda}_i \hat{\mathbf{e}}'_i.$$

بنابراین،

$$s_{\hat{Y}_i, X_k} = \hat{\lambda}_i \hat{e}_{ki}.$$

از طرف دیگر،

$$s_{\hat{Y}_i}^2 = \hat{\lambda}_i$$

و

$$s_{X_k}^2 = s_{kk}.$$

پس ضریب همبستگی نمونه‌ای بین مؤلفه اصلی i ام و متغیر X_k برابر است با

$$r_{\hat{Y}_i, X_k} = \frac{s_{\hat{Y}_i, X_k}}{s_{\hat{Y}_i} s_{X_k}}.$$

در نتیجه،

$$r_{\hat{Y}_i, X_k} = \frac{\hat{\lambda}_i \hat{e}_{ki}}{\sqrt{\hat{\lambda}_i} \sqrt{s_{kk}}}$$

و در نهایت

$$r_{\hat{Y}_i, X_k} = \frac{\hat{e}_{ki} \sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}, \quad i, k = 1, 2, \dots, p.$$

این رابطه در تفسیر مؤلفه‌های اصلی بسیار مهم است.

اگر

$$|r_{\hat{Y}_i, X_k}|$$

به یک نزدیک باشد، متغیر X_k ارتباط قوی با مؤلفه $\hat{Y}_i$ دارد.

اگر

$$r_{\hat{Y}_i, X_k} > 0,$$

رابطه در جهت مثبت است و اگر

$$r_{\hat{Y}_i, X_k} < 0,$$

رابطه در جهت منفی است.

اگر مقدار همبستگی به صفر نزدیک باشد، متغیر مورد نظر نقش چندانی در تفسیر آن مؤلفه ندارد.

۹.۳.۹ تفاوت ضریب بردار ویژه و ضریب همبستگی

در تفسیر خروجی تحلیل مؤلفه‌های اصلی باید بین دو کمیت زیر تفاوت قائل شد:

۱. ضریب متغیر در بردار ویژه، یعنی $\hat{e}_{ki}$ ؛

۲. ضریب همبستگی متغیر با مؤلفه اصلی، یعنی $r_{\hat{Y}_i, X_k}$.

ضریب $\hat{e}_{ki}$ نشان می‌دهد متغیر X_k با چه وزنی در ترکیب خطی تعریف‌کننده مؤلفه اصلی وارد شده است. برای مثال،

$$\hat{Y}_1 = 0.70X_1 + 0.10X_2 - 0.65X_3$$

نشان می‌دهد ضرایب مؤلفه اول برابر 0.70، 0.10 و -0.65 هستند.

اما برای تفسیر ارتباط مؤلفه اول با متغیرها می‌توان همبستگی‌های

$$r_{\hat{Y}_1, X_1}, \quad r_{\hat{Y}_1, X_2}, \quad r_{\hat{Y}_1, X_3}$$

را بررسی کرد.

این تمایز به‌خصوص زمانی مهم است که متغیرها دارای واریانس‌های بسیار متفاوت باشند. به همین دلیل در جدول‌های خروجی تحلیل مؤلفه‌های اصلی معمولاً علاوه بر ضرایب بردارهای ویژه، همبستگی متغیرها با مؤلفه‌ها نیز گزارش می‌شود.

۱۰.۳.۹ ارتباط تحلیل نمونه‌ای با تحلیل جامعه

در حالت جامعه، تحلیل مؤلفه‌های اصلی بر اساس زوج‌های مقدار ویژه-بردار ویژه ماتریس Σ انجام می‌شود:

$$\Sigma \mathbf{e}_i = \lambda_i \mathbf{e}_i.$$

در حالت نمونه، Σ ناشناخته است و از S استفاده می‌کنیم:

$$S \hat{\mathbf{e}}_i = \hat{\lambda}_i \hat{\mathbf{e}}_i.$$

بنابراین، به طور خلاصه می‌توان نوشت:

جامعه	نمونه
Σ	S
λ_i	$\hat{\lambda}_i$
$\mathbf{e}_i$	$\hat{\mathbf{e}}_i$
$Y_i = \mathbf{e}_i' \mathbf{X}$	$\hat{Y}_i = \hat{\mathbf{e}}_i' \mathbf{x}$
$\text{Var}(Y_i) = \lambda_i$	$s_{\hat{Y}_i}^2 = \hat{\lambda}_i$

در نتیجه، روش محاسباتی در هر دو حالت یکسان است و تنها پارامترهای ناشناخته جامعه با کمیت‌های نمونه‌ای متناظر جایگزین می‌شوند.

۱۱.۳.۹ ارتباط با برآورد درست‌نمایی بیشینه

اگر

$$\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$$

یک نمونه تصادفی از توزیع نرمال چندمتغیره

$$N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

باشند، برآورد درست‌نمایی بیشینه ماتریس کوواریانس برابر است با

$$\hat{\boldsymbol{\Sigma}}_{\text{MLE}} = \frac{1}{n} \sum_{r=1}^n (\mathbf{X}_r - \bar{\mathbf{X}})(\mathbf{X}_r - \bar{\mathbf{X}})'.$$

در حالی که ماتریس کوواریانس نمونه معمول به صورت

$$\mathbf{S} = \frac{1}{n-1} \sum_{r=1}^n (\mathbf{X}_r - \bar{\mathbf{X}})(\mathbf{X}_r - \bar{\mathbf{X}})'$$

تعریف می‌شود.

بنابراین،

$$\boxed{\hat{\boldsymbol{\Sigma}}_{\text{MLE}} = \frac{n-1}{n} \mathbf{S}.}$$

فرض کنید

$$\mathbf{S} \hat{\mathbf{e}}_i = \hat{\lambda}_i \hat{\mathbf{e}}_i.$$

آنگاه

$$\hat{\boldsymbol{\Sigma}}_{\text{MLE}} \hat{\mathbf{e}}_i = \frac{n-1}{n} \mathbf{S} \hat{\mathbf{e}}_i$$

و بنابراین

$$\hat{\boldsymbol{\Sigma}}_{\text{MLE}} \hat{\mathbf{e}}_i = \frac{n-1}{n} \hat{\lambda}_i \hat{\mathbf{e}}_i.$$

پس بردارهای ویژه $\mathbf{S}$ و $\hat{\boldsymbol{\Sigma}}_{\text{MLE}}$ دقیقاً یکسان هستند و فقط مقادیر ویژه آنها با یک ضریب ثابت تفاوت دارند:

$$\boxed{\hat{\lambda}_{i,\text{MLE}} = \frac{n-1}{n} \hat{\lambda}_i.}$$

این نکته نتیجه مهمی دارد: مؤلفه‌های اصلی حاصل از $\mathbf{S}$ و $\hat{\boldsymbol{\Sigma}}_{\text{MLE}}$ دارای جهت‌های یکسان هستند.

همچنین نسبت واریانس توضیح داده‌شده نیز تغییر نمی‌کند، زیرا

$$\frac{\frac{n-1}{n}\hat{\lambda}_i}{\sum_{j=1}^p \frac{n-1}{n}\hat{\lambda}_j} = \frac{\hat{\lambda}_i}{\sum_{j=1}^p \hat{\lambda}_j}.$$

بنابراین انتخاب S یا برآورد درست‌نمایی بیشینه ماتریس کوواریانس، جهت مؤلفه‌ها و درصد واریانس توضیح داده‌شده توسط آنها را تغییر نمی‌دهد.

۱۲.۳.۹ تعبیر هندسی مؤلفه‌های اصلی نمونه

در حالت نمونه نیز تعبیر هندسی مؤلفه‌های اصلی مشابه حالت جامعه است. فرض کنید S مثبت معین باشد. معادله

$$(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) = c^2$$

یک بیضوی p بعدی با مرکز $\bar{\mathbf{x}}$ تعریف می‌کند.

از تجزیه طیفی ماتریس S داریم

$$S = \hat{\mathbf{E}} \hat{\mathbf{\Lambda}} \hat{\mathbf{E}}'.$$

در نتیجه،

$$S^{-1} = \hat{\mathbf{E}} \hat{\mathbf{\Lambda}}^{-1} \hat{\mathbf{E}}'.$$

اگر مختصات جدید را به صورت

$$\mathbf{y} = \hat{\mathbf{E}}' (\mathbf{x} - \bar{\mathbf{x}})$$

تعریف کنیم، معادله بیضوی به صورت

$$\frac{y_1^2}{\hat{\lambda}_1} + \frac{y_2^2}{\hat{\lambda}_2} + \dots + \frac{y_p^2}{\hat{\lambda}_p} = c^2$$

درمی‌آید.

یا به صورت استانداردتر،

$$\frac{y_1^2}{c^2 \hat{\lambda}_1} + \frac{y_2^2}{c^2 \hat{\lambda}_2} + \dots + \frac{y_p^2}{c^2 \hat{\lambda}_p} = 1.$$

بنابراین طول نیم‌محورهای بیضوی متناسب با

$$c\sqrt{\hat{\lambda}_1}, \quad c\sqrt{\hat{\lambda}_2}, \quad \dots, \quad c\sqrt{\hat{\lambda}_p}$$

است و جهت محورهای آن توسط بردارهای ویژه

$$\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \dots, \hat{\mathbf{e}}_p$$

تعیین می‌شود.

از آنجا که

$$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p,$$

محور متناظر با $\hat{\mathbf{e}}_1$ بلندترین محور بیضوی و جهت بیشترین تغییرپذیری داده‌های نمونه است. به همین ترتیب، محور متناظر با $\hat{\mathbf{e}}_p$ کوتاه‌ترین محور و جهت کمترین تغییرپذیری است.

بنابراین از دیدگاه هندسی، تحلیل مؤلفه‌های اصلی را می‌توان نوعی چرخش دستگاه مختصات دانست؛ دستگاه مختصات اولیه به گونه‌ای می‌چرخد که محورهای جدید در جهت بیشترین تا کمترین تغییرپذیری داده‌ها قرار گیرند.

۱۳.۳.۹ مراحل عملی تحلیل مؤلفه‌های اصلی بر اساس ماتریس کوواریانس نمونه

برای اجرای تحلیل مؤلفه‌های اصلی نمونه‌ای می‌توان مراحل زیر را انجام داد:

۱. بردار میانگین نمونه را محاسبه کنید:

$$\bar{\mathbf{x}}.$$

۲. ماتریس واریانس-کوواریانس نمونه را تشکیل دهید:

$$\mathbf{S}.$$

۳. معادله مشخصه را حل کنید:

$$|\mathbf{S} - \lambda \mathbf{I}| = 0.$$

۴. مقادیر ویژه را به ترتیب نزولی مرتب کنید:

$$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p.$$

۵. برای هر مقدار ویژه، بردار ویژه واحد متناظر را محاسبه کنید:

$$\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \dots, \hat{\mathbf{e}}_p.$$

۶. مؤلفه‌های اصلی را تشکیل دهید:

$$\hat{Y}_i = \hat{\mathbf{e}}_i' \mathbf{x}.$$

۷. سهم واریانس هر مؤلفه را محاسبه کنید:

$$P_i = \frac{\hat{\lambda}_i}{\sum_{j=1}^p \hat{\lambda}_j}.$$

۸. درصد تجمعی واریانس را محاسبه کنید:

$$C_m = 100 \frac{\sum_{i=1}^m \hat{\lambda}_i}{\sum_{i=1}^p \hat{\lambda}_i}.$$

۹. برای تفسیر مؤلفه‌ها، همبستگی هر متغیر با مؤلفه‌ها را محاسبه کنید:

$$r_{\hat{Y}_i, X_k} = \frac{\hat{e}_{ki} \sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}.$$

۱۰. بر اساس درصد تجمعی واریانس و قابلیت تفسیر مؤلفه‌ها، تعداد مناسبی از مؤلفه‌های اصلی را نگه دارید.

۱۴.۳.۹ خلاصه روابط اصلی

برای استفاده سریع در مثال‌ها، مهم‌ترین روابط این بخش را می‌توان به صورت زیر خلاصه کرد:

$$\mathbf{S} \hat{\mathbf{e}}_i = \hat{\lambda}_i \hat{\mathbf{e}}_i$$

$$\hat{Y}_i = \hat{\mathbf{e}}_i' \mathbf{x}$$

$$s_{\hat{Y}_i}^2 = \hat{\lambda}_i$$

$$s_{\hat{Y}_i, \hat{Y}_k} = 0, \quad i \neq k$$

$$\sum_{j=1}^p s_{jj} = \sum_{i=1}^p \hat{\lambda}_i$$

$$i \text{ سهم واریانس مؤلفه } = \frac{\hat{\lambda}_i}{\sum_{j=1}^p \hat{\lambda}_j}$$

و

$$r_{\hat{Y}_i, X_k} = \frac{\hat{e}_{ki} \sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}.$$

بنابراین، از نظر ساختاری تقریباً تمام نتایجی که برای مؤلفه‌های اصلی جامعه بر اساس Σ به دست آوردیم، برای مؤلفه‌های اصلی نمونه نیز برقرار است؛ کافی است λ_i و e_i به ترتیب با $\hat{\lambda}_i$ و $\hat{e}_i$ جایگزین شوند.

۱۵.۳.۹ مثال: تحلیل مؤلفه‌های اصلی داده‌های اجتماعی-اقتصادی

در یک مطالعه، اطلاعات مربوط به ۱۴ ناحیه بر اساس پنج متغیر اجتماعی-اقتصادی گردآوری شده است. متغیرهای مورد بررسی عبارت‌اند از:

- X_1 ، کل جمعیت، بر حسب هزار نفر،
- X_2 ، میانه سال‌های تحصیل،
- X_3 ، کل اشتغال، بر حسب هزار نفر،
- X_4 ، خدمات بهداشتی-استخدامی، بر حسب صد نفر،
- X_5 ، میانه ارزش منزل، بر حسب ده‌هزار دلار.

بردار میانگین نمونه‌ای برابر است با

$$\bar{\mathbf{x}}' = [4.32 \quad 14.01 \quad 1.95 \quad 2.17 \quad 2.45],$$

و ماتریس واریانس-کوواریانس نمونه به صورت زیر گزارش شده است:

$$\mathbf{S} = \begin{pmatrix} 4.308 & 1.683 & 1.803 & 2.155 & -0.253 \\ 1.683 & 1.768 & 0.588 & 0.177 & 0.176 \\ 1.803 & 0.588 & 0.801 & 1.065 & -0.158 \\ 2.155 & 0.177 & 1.065 & 1.970 & -0.357 \\ -0.253 & 0.176 & -0.158 & -0.357 & 0.504 \end{pmatrix}.$$

هدف آن است که بررسی کنیم آیا می‌توان تغییرات موجود در پنج متغیر اولیه را با یک یا دو مؤلفه اصلی خلاصه کرد یا خیر.

مرحله اول: محاسبه واریانس کل نمونه واریانس کل نمونه برابر مجموع عناصر قطر اصلی ماتریس S است:

$$\text{tr}(\mathbf{S}) = 4.308 + 1.768 + 0.801 + 1.970 + 0.504.$$

بنابراین،

$$\text{tr}(\mathbf{S}) = 9.351.$$

این مقدار باید با مجموع پنج مقدار ویژه ماتریس S برابر باشد.

مرحله دوم: محاسبه مقادیر ویژه مقادیر ویژه از حل معادله مشخصه

$$|\mathbf{S} - \lambda \mathbf{I}| = 0$$

به دست می‌آیند. به دلیل مرتبه پنج ماتریس، بسط مستقیم دترمینان و حل چندجمله‌ای درجه پنج از نظر محاسبات دستی مناسب نیست و مقادیر ویژه به روش عددی محاسبه می‌شوند. مقادیر ویژه مرتب‌شده از بزرگ به کوچک عبارت‌اند از:

$$\begin{aligned}\hat{\lambda}_1 &= 6.9308, \\ \hat{\lambda}_2 &= 1.7867, \\ \hat{\lambda}_3 &= 0.3892, \\ \hat{\lambda}_4 &= 0.2300, \\ \hat{\lambda}_5 &= 0.0143.\end{aligned}$$

برای کنترل محاسبات داریم:

$$\begin{aligned}\sum_{i=1}^5 \hat{\lambda}_i &= 6.9308 + 1.7867 + 0.3892 + 0.2300 + 0.0143 \\ &= 9.3510.\end{aligned}$$

پس

$$\sum_{i=1}^5 \hat{\lambda}_i = \text{tr}(\mathbf{S}) = 9.351.$$

مرحله سوم: محاسبه بردارهای ویژه برای هر مقدار ویژه، دستگاه

$$(S - \hat{\lambda}_i I) \hat{e}_i = 0$$

حل و سپس بردار حاصل به طول یک نرمال می‌شود.

با انتخاب علامت‌هایی که تفسیر مؤلفه‌ها را ساده‌تر می‌کنند، بردارهای ویژه واحد به صورت تقریبی برابرند با:

$$\hat{e}_1 = \begin{pmatrix} 0.781 \\ 0.306 \\ 0.335 \\ 0.426 \\ -0.054 \end{pmatrix}, \quad \hat{e}_2 = \begin{pmatrix} -0.071 \\ -0.764 \\ 0.083 \\ 0.579 \\ -0.262 \end{pmatrix},$$

$$\hat{e}_3 = \begin{pmatrix} 0.004 \\ -0.162 \\ 0.015 \\ 0.220 \\ 0.962 \end{pmatrix}, \quad \hat{e}_4 = \begin{pmatrix} -0.542 \\ 0.545 \\ -0.050 \\ 0.636 \\ -0.050 \end{pmatrix},$$

و

$$\hat{e}_5 = \begin{pmatrix} -0.302 \\ -0.010 \\ 0.937 \\ -0.173 \\ 0.024 \end{pmatrix}.$$

توجه شود که علامت هر بردار ویژه اختیاری است. برای مثال، $\hat{e}_1$ و $-\hat{e}_1$ هر دو بردار ویژه معتبرند و تنها علامت مؤلفه اصلی متناظر را تغییر می‌دهند.

مرحله چهارم: تشکیل مؤلفه‌های اصلی مؤلفه اصلی اول برابر است با

$$\hat{Y}_1 = \hat{e}_1' \mathbf{x}.$$

بنابراین،

$$\hat{Y}_1 = 0.781X_1 + 0.306X_2 + 0.335X_3 + 0.426X_4 - 0.054X_5.$$

مؤلفه اصلی دوم برابر است با

$$\hat{Y}_2 = -0.071X_1 - 0.764X_2 + 0.083X_3 + 0.579X_4 - 0.262X_5.$$

به همین ترتیب، مؤلفه‌های سوم تا پنجم عبارت‌اند از:

$$\hat{Y}_3 = 0.004X_1 - 0.162X_2 + 0.015X_3 + 0.220X_4 + 0.962X_5,$$

$$\hat{Y}_4 = -0.542X_1 + 0.545X_2 - 0.050X_3 + 0.636X_4 - 0.050X_5,$$

و

$$\hat{Y}_5 = -0.302X_1 - 0.010X_2 + 0.937X_3 - 0.173X_4 + 0.024X_5.$$

واریانس هر مؤلفه برابر مقدار ویژه متناظر آن است:

$$\text{Var}_{\text{sample}}(\hat{Y}_i) = \hat{\lambda}_i.$$

بنابراین،

$$s_{\hat{Y}_1}^2 = 6.9308,$$

$$s_{\hat{Y}_2}^2 = 1.7867,$$

$$s_{\hat{Y}_3}^2 = 0.3892,$$

$$s_{\hat{Y}_4}^2 = 0.2300,$$

$$s_{\hat{Y}_5}^2 = 0.0143.$$

مرحله پنجم: درصد واریانس توضیح داده‌شده سهم مؤلفه i ام از واریانس کل از رابطه

$$P_i = \frac{\hat{\lambda}_i}{\sum_{j=1}^5 \hat{\lambda}_j}$$

به دست می‌آید.

برای مؤلفه اول داریم:

$$\begin{aligned} P_1 &= \frac{6.9308}{9.351} \\ &= 0.7412. \end{aligned}$$

بنابراین،

از واریانس کل را توضیح می‌دهد. 74.12% مؤلفه اول حدود

برای مؤلفه دوم:

$$P_2 = \frac{1.7867}{9.351} \\ = 0.1911.$$

پس

از واریانس کل را توضیح می‌دهد. 19.11% مؤلفه دوم حدود

برای سه مؤلفه دیگر داریم:

$$P_3 = \frac{0.3892}{9.351} = 0.0416, \\ P_4 = \frac{0.2300}{9.351} = 0.0246, \\ P_5 = \frac{0.0143}{9.351} = 0.0015.$$

بنابراین درصدهای واریانس توضیح داده شده عبارت‌اند از:

$$100P_1 = 74.12\%, \\ 100P_2 = 19.11\%, \\ 100P_3 = 4.16\%, \\ 100P_4 = 2.46\%, \\ 100P_5 = 0.15\%.$$

درصد تجمعی دو مؤلفه اول برابر است با

$$100(P_1 + P_2) = 100 \frac{6.9308 + 1.7867}{9.351} \\ = 93.23\%.$$

پس

از واریانس کل را توضیح می‌دهند. 93.23% دو مؤلفه اول در مجموع حدود

جدول کامل مقادیر ویژه و درصد واریانس به صورت زیر است:

مؤلفه	مقدار ویژه	درصد واریانس	درصد تجمعی
$\hat{Y}_1$	6.9308	74.12	74.12
$\hat{Y}_2$	1.7867	19.11	93.23
$\hat{Y}_3$	0.3892	4.16	97.39
$\hat{Y}_4$	0.2300	2.46	99.85
$\hat{Y}_5$	0.0143	0.15	100.00

از آنجا که دو مؤلفه اول بیش از 93% از واریانس کل را توضیح می‌دهند، کاهش بعد از پنج متغیر به دو مؤلفه اصلی منطقی است.

مرحله ششم: محاسبه همبستگی متغیرها با مؤلفه‌های اصلی برای تفسیر دقیق‌تر مؤلفه‌ها از رابطه

$$r_{\hat{Y}_i, X_k} = \frac{\hat{e}_{ki} \sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}$$

استفاده می‌کنیم.

ابتدا انحراف معیارهای نمونه‌ای متغیرها را محاسبه می‌کنیم:

$$s_{X_1} = \sqrt{4.308} = 2.0756,$$

$$s_{X_2} = \sqrt{1.768} = 1.3297,$$

$$s_{X_3} = \sqrt{0.801} = 0.8949,$$

$$s_{X_4} = \sqrt{1.970} = 1.4036,$$

$$s_{X_5} = \sqrt{0.504} = 0.7099.$$

همچنین،

$$\sqrt{\hat{\lambda}_1} = \sqrt{6.9308} = 2.6327.$$

بنابراین، همبستگی مؤلفه اول با X_1 برابر است با

$$r_{\hat{Y}_1, X_1} = \frac{(0.781)(2.6327)}{2.0756} \approx 0.991.$$

برای X_2 داریم:

$$r_{\hat{Y}_1, X_2} = \frac{(0.306)(2.6327)}{1.3297} \approx 0.605.$$

برای X_3 :

$$r_{\hat{Y}_1, X_3} = \frac{(0.335)(2.6327)}{0.8949} \\ \approx 0.984.$$

برای X_4 :

$$r_{\hat{Y}_1, X_4} = \frac{(0.426)(2.6327)}{1.4036} \\ \approx 0.799.$$

و برای X_5 :

$$r_{\hat{Y}_1, X_5} = \frac{(-0.054)(2.6327)}{0.7099} \\ \approx -0.201.$$

بنابراین،

$$\begin{aligned} r_{\hat{Y}_1, X_1} &= 0.991, \\ r_{\hat{Y}_1, X_2} &= 0.605, \\ r_{\hat{Y}_1, X_3} &= 0.984, \\ r_{\hat{Y}_1, X_4} &= 0.799, \\ r_{\hat{Y}_1, X_5} &= -0.201. \end{aligned}$$

برای مؤلفه دوم نیز داریم:

$$\sqrt{\hat{\lambda}_2} = \sqrt{1.7867} = 1.3367.$$

در نتیجه،

$$\begin{aligned} r_{\hat{Y}_2, X_1} &= -0.046, \\ r_{\hat{Y}_2, X_2} &= -0.768, \\ r_{\hat{Y}_2, X_3} &= 0.124, \\ r_{\hat{Y}_2, X_4} &= 0.552, \\ r_{\hat{Y}_2, X_5} &= -0.494. \end{aligned}$$

جدول ضرایب بردارهای ویژه و همبستگی‌های دو مؤلفه اول به صورت زیر است:

متغیر	$\hat{e}_{k1}$	$r_{\hat{Y}_1, X_k}$	$\hat{e}_{k2}$	$r_{\hat{Y}_2, X_k}$
کل جمعیت	0.781	0.991	-0.071	-0.046
میانۀ سال‌های تحصیل	0.306	0.605	-0.764	-0.768
کل اشتغال	0.335	0.984	0.083	0.124
خدمات بهداشتی-استخدامی	0.426	0.799	0.579	0.552
میانۀ ارزش منزل	-0.054	-0.201	-0.262	-0.494

تفسیر مؤلفه اصلی اول مؤلفه اول برابر است با

$$\hat{Y}_1 = 0.781X_1 + 0.306X_2 + 0.335X_3 + 0.426X_4 - 0.054X_5.$$

ضرایب چهار متغیر اول مثبت‌اند و همبستگی مؤلفه اول با متغیرهای کل جمعیت، کل اشتغال و خدمات بهداشتی-استخدامی بسیار زیاد است:

$$r_{\hat{Y}_1, X_1} = 0.991,$$

$$r_{\hat{Y}_1, X_3} = 0.984,$$

و

$$r_{\hat{Y}_1, X_4} = 0.799.$$

در نتیجه، مؤلفه اول را می‌توان یک شاخص کلی از اندازه و فعالیت اجتماعی-اقتصادی ناحیه در نظر گرفت. ناحیه‌ای که مقدار $\hat{Y}_1$ بزرگی دارد، معمولاً دارای جمعیت بیشتر، اشتغال بیشتر و فعالیت خدماتی بیشتر است. متغیر میانۀ ارزش منزل همبستگی نسبتاً ضعیفی با مؤلفه اول دارد، بنابراین در تعریف این بعد کلی نقش کمتری ایفا می‌کند.

تفسیر مؤلفه اصلی دوم مؤلفه دوم برابر است با

$$\hat{Y}_2 = -0.071X_1 - 0.764X_2 + 0.083X_3 + 0.579X_4 - 0.262X_5.$$

بزرگ‌ترین ضرایب این مؤلفه، از نظر قدر مطلق، مربوط به میانۀ سال‌های تحصیل و خدمات بهداشتی-استخدامی هستند. همبستگی‌ها نیز نشان می‌دهند:

$$r_{\hat{Y}_2, X_2} = -0.768,$$

$$r_{\hat{Y}_2, X_4} = 0.552,$$

و

$$r_{\hat{Y}_2, X_5} = -0.494.$$

بنابراین مؤلفه دوم نوعی تقابل میان

خدمات بهداشتی-استخدامی

و

سطح تحصیلات و ارزش منزل

را نشان می‌دهد.

به دلیل اختیاری بودن علامت بردار ویژه، می‌توان کل مؤلفه دوم را در 1- ضرب کرد و آن را به صورت

$$-\hat{Y}_2 = 0.071X_1 + 0.764X_2 - 0.083X_3 - 0.579X_4 + 0.262X_5$$

نوشت. در این نمایش، مؤلفه دوم بیشتر به عنوان شاخصی از تحصیلات و ارزش منزل در مقابل اشتغال خدماتی تفسیر می‌شود.

نتیجه کاهش بعد دو مؤلفه اول در مجموع

$$93.23\%$$

از واریانس کل را توضیح می‌دهند. بنابراین می‌توان به جای پنج متغیر اولیه، هر ناحیه را با دو نمره

$$\hat{Y}_1$$

و

$$\hat{Y}_2$$

نمایش داد.

در نتیجه، داده‌های مربوط به 14 ناحیه که در فضای پنج‌بعدی قرار داشتند، با از دست دادن کمتر از 7% از تغییرپذیری، در یک فضای دوبعدی خلاصه می‌شوند.

۱۶.۳.۹ حل نرم‌افزاری مثال در R

برای حل این مثال در نرم‌افزار R، از آنجا که ماتریس کوواریانس نمونه مستقیماً در اختیار است، می‌توان از تابع `eigen` استفاده کرد.

توجه شود که برای استفاده از تابع `prcomp` به داده‌های خام مربوط به 14 ناحیه نیاز داریم. چون در این مثال تنها بردار میانگین و ماتریس کوواریانس گزارش شده‌اند، تجزیه مستقیم ماتریس `S` روش مناسب‌تری است.

Listing 30:

R

```

1 S <- matrix(
2   c(
3     4.308, 1.683, 1.803, 2.155, -0.253,
4     1.683, 1.768, 0.588, 0.177, 0.176,
5     1.803, 0.588, 0.801, 1.065, -0.158,
6     2.155, 0.177, 1.065, 1.970, -0.357,
7     -0.253, 0.176, -0.158, -0.357, 0.504
8   ),
9   nrow = 5,
10  byrow = TRUE
11 )
12
13 variable_names <- c(
14   "Population",
15   "SchoolYears",
16   "Employment",
17   "HealthServices",
18   "HomeValue"
19 )
20
21 rownames(S) <- variable_names
22 colnames(S) <- variable_names
23
24 S

```

خروجی نرم‌افزار:

Listing 31:

	Population	SchoolYears	Employment	HealthServices	HomeValue
Population	4.308	1.683	1.803	2.155	-0.253
SchoolYears	1.683	1.768	0.588	0.177	0.176
Employment	1.803	0.588	0.801	1.065	-0.158
HealthServices	2.155	0.177	1.065	1.970	-0.357
HomeValue	-0.253	0.176	-0.158	-0.357	0.504

Listing 32:

```

1 eig <- eigen(S, symmetric = TRUE)
2
3 lambda <- eig$values
4 E <- eig$vectors
5
6 # Adjust signs to match the convention used in the notes
7 E[, 1] <- -E[, 1]
8 E[, 2] <- -E[, 2]
9
10 colnames(E) <- paste0("PC", 1:5)
11 rownames(E) <- variable_names
12
13 cat("Eigenvalues:\n")
14 print(round(lambda, 6))
15
16 cat("\nEigenvectors:\n")
17 print(round(E, 3))

```

خروجی مقادیر ویژه:

Listing 33:

```
[1] 6.930831 1.786685 0.389200 0.229995 0.014289
```

خروجی تقریبی بردارهای ویژه:

Listing 34:

	PC1	PC2	PC3	PC4	PC5
Population	0.781	-0.071	0.004	-0.542	-0.302
SchoolYears	0.306	-0.764	-0.162	0.545	-0.010
Employment	0.335	0.083	0.015	-0.050	0.937
HealthServices	0.426	0.579	0.220	0.636	-0.173
HomeValue	-0.054	-0.262	0.962	-0.050	0.024

ممکن است علامت یک یا چند ستون در اجرای مستقیم تابع eigen متفاوت باشد. این تفاوت هیچ اثری بر واریانس، درصد واریانس توضیح داده شده یا نتیجه کاهش بعد ندارد.

Listing 35:

```

1 total_variance <- sum(lambda)
2
3 variance_percent <- 100 * lambda / total_variance

```

```

4 cumulative_percent <- cumsum(variance_percent)
5
6 variance_table <- data.frame(
7   Component = paste0("PC", 1:5),
8   Eigenvalue = round(lambda, 4),
9   Variance_Percent = round(variance_percent, 2),
10  Cumulative_Percent = round(cumulative_percent, 2)
11 )
12
13 print(variance_table, row.names = FALSE)

```

خروجی:

Listing 36:

Component	Eigenvalue	Variance_Percent	Cumulative_Percent
PC1	6.9308	74.12	74.12
PC2	1.7867	19.11	93.23
PC3	0.3892	4.16	97.39
PC4	0.2300	2.46	99.85
PC5	0.0143	0.15	100.00

Listing 37:

```

1 sd_variables <- sqrt(diag(S))
2
3 cor_X_PC <- diag(1 / sd_variables) %*%
4   E %*%
5   diag(sqrt(lambda))
6
7 rownames(cor_X_PC) <- variable_names
8 colnames(cor_X_PC) <- paste0("PC", 1:5)
9
10 round(cor_X_PC, 3)

```

خروجی:

Listing 38:

	PC1	PC2	PC3	PC4	PC5
Population	0.991	-0.046	0.001	-0.125	-0.017
SchoolYears	0.605	-0.768	-0.076	0.196	-0.001
Employment	0.984	0.124	0.011	-0.027	0.125

HealthServices	0.799	0.552	0.098	0.217	-0.015
HomeValue	-0.201	-0.494	0.845	-0.034	0.004

رسم نمودار اسکری نمودار اسکری مقادیر ویژه را بر حسب شماره مؤلفه نمایش می‌دهد. افت شدید پس از مؤلفه دوم نشان می‌دهد که نگه‌داشتن دو مؤلفه اول مناسب است.

Listing 39:

```

1 plot(
2   1:5,
3   lambda,
4   type = "b",
5   pch = 19,
6   lwd = 2,
7   xlab = "Principal component",
8   ylab = "Eigenvalue",
9   main = "Scree plot",
10  xaxt = "n"
11 )
12
13 axis(1, at = 1:5, labels = paste0("PC", 1:5))
14 abline(h = 1, lty = 2)
15
16 text(
17   1:5,
18   lambda,
19   labels = round(lambda, 2),
20   pos = 3
21 )

```

Listing 40:

```

1 mid_points <- barplot(
2   variance_percent,
3   names.arg = paste0("PC", 1:5),
4   ylim = c(0, 105),
5   xlab = "Principal component",
6   ylab = "Percentage of variance",
7   main = "Explained variance"
8 )
9

```

```

10 lines(
11     mid_points,
12     cumulative_percent,
13     type = "b",
14     pch = 19,
15     lwd = 2
16 )
17
18 text(
19     mid_points,
20     variance_percent,
21     labels = paste0(round(variance_percent, 1), "%"),
22     pos = 3
23 )
24
25 legend(
26     "right",
27     legend = c("Individual percentage", "Cumulative percentage"),
28     pch = c(15, 19),
29     lty = c(NA, 1),
30     bty = "n"
31 )

```

نمودار متغیرها در فضای دو مؤلفه اول از آنجا که داده‌های خام در اختیار نیست، نمی‌توان نمره‌های مربوط به 14 ناحیه را رسم کرد. با این حال، می‌توان متغیرها را بر اساس همبستگی آنها با دو مؤلفه اول نمایش داد.

Listing 41:

```

1 plot(
2     cor_X_PC[, 1],
3     cor_X_PC[, 2],
4     xlim = c(-1, 1),
5     ylim = c(-1, 1),
6     asp = 1,
7     pch = 19,
8     xlab = "Correlation with PC1",
9     ylab = "Correlation with PC2",
10    main = "Variables in the PC1-PC2 plane"
11 )
12
13 abline(h = 0, v = 0, lty = 3)
14

```

```

15 symbols(
16   0,
17   0,
18   circles = 1,
19   inches = FALSE,
20   add = TRUE
21 )
22
23 arrows(
24   0,
25   0,
26   cor_X_PC[, 1],
27   cor_X_PC[, 2],
28   length = 0.08
29 )
30
31 text(
32   cor_X_PC[, 1],
33   cor_X_PC[, 2],
34   labels = variable_names,
35   pos = 3,
36   cex = 0.8
37 )

```

در این نمودار، متغیرهای کل جمعیت، کل اشتغال و خدمات بهداشتی—استخدامی در جهت مثبت مؤلفه اول قرار می‌گیرند. متغیر میانه سال‌های تحصیل در جهت منفی مؤلفه دوم و خدمات بهداشتی—استخدامی در جهت مثبت مؤلفه دوم قرار می‌گیرد. این وضعیت تفسیر ارائه‌شده برای دو مؤلفه اول را تأیید می‌کند.

۱۷.۳.۹ مثال: تحلیل مؤلفه‌های اصلی اندازه‌های لاک‌پشت‌ها

در یک مطالعه زیست‌سنجی، سه اندازه مربوط به لاک‌پشت‌های ماده بررسی شده است. متغیرهای مورد مطالعه عبارت‌اند از:

X_1 = لگاریتم طول لاک‌پشت،

X_2 = لگاریتم عرض لاک‌پشت،

X_3 = لگاریتم ارتفاع لاک‌پشت.

در این مثال، به دلیل آنکه اندازه‌های طول، عرض و ارتفاع معمولاً دارای روابط ضربی هستند، از تبدیل لگاریتمی استفاده شده است. بردار میانگین نمونه‌ای برای 24 لاک‌پشت ماده برابر است با

$$\bar{\mathbf{x}}' = \begin{bmatrix} 4.725 & 4.478 & 3.703 \end{bmatrix}.$$

ماتریس واریانس-کوواریانس نمونه نیز برابر است با

$$\mathbf{S} = 10^{-3} \begin{pmatrix} 11.555 & 8.367 & 8.508 \\ 8.367 & 6.697 & 6.264 \\ 8.508 & 6.264 & 7.061 \end{pmatrix}.$$

بنابراین، به صورت اعشاری داریم

$$\mathbf{S} = \begin{pmatrix} 0.011555 & 0.008367 & 0.008508 \\ 0.008367 & 0.006697 & 0.006264 \\ 0.008508 & 0.006264 & 0.007061 \end{pmatrix}.$$

هدف آن است که بررسی کنیم آیا می‌توان سه متغیر اندازه‌گیری‌شده را با یک یا دو مؤلفه اصلی خلاصه کرد یا خیر.

مرحله اول: محاسبه واریانس کل واریانس کل نمونه برابر اثر ماتریس $\mathbf{S}$ است:

$$\text{tr}(\mathbf{S}) = 0.011555 + 0.006697 + 0.007061.$$

پس

$$\boxed{\text{tr}(\mathbf{S}) = 0.025313.}$$

این مقدار باید با مجموع سه مقدار ویژه ماتریس $\mathbf{S}$ برابر باشد.

مرحله دوم: محاسبه مقادیر ویژه مقادیر ویژه از حل معادله مشخصه

$$|\mathbf{S} - \lambda \mathbf{I}| = 0$$

به دست می‌آیند.

به صورت عددی، مقادیر ویژه مرتب‌شده از بزرگ به کوچک عبارت‌اند از:

$$\boxed{\begin{aligned} \hat{\lambda}_1 &= 0.0243095, \\ \hat{\lambda}_2 &= 0.0006246, \\ \hat{\lambda}_3 &= 0.0003789. \end{aligned}}$$

با استفاده از ضریب 10^{-3} می‌توان آنها را به صورت زیر نیز نوشت:

$$\begin{aligned}\hat{\lambda}_1 &= 24.3095 \times 10^{-3}, \\ \hat{\lambda}_2 &= 0.6246 \times 10^{-3}, \\ \hat{\lambda}_3 &= 0.3789 \times 10^{-3}.\end{aligned}$$

برای کنترل محاسبات داریم:

$$\begin{aligned}\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3 &= 0.0243095 + 0.0006246 + 0.0003789 \\ &= 0.025313.\end{aligned}$$

بنابراین

$$\sum_{i=1}^3 \hat{\lambda}_i = \text{tr}(\mathbf{S}).$$

مرحله سوم: محاسبه بردارهای ویژه برای هر مقدار ویژه، دستگاه

$$(\mathbf{S} - \hat{\lambda}_i \mathbf{I}) \hat{\mathbf{e}}_i = \mathbf{0}$$

حل می‌شود.

با انتخاب علامت مناسب برای سهولت تفسیر، بردارهای ویژه واحد به صورت زیر هستند:

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 0.683 \\ 0.510 \\ 0.522 \end{pmatrix}$$

$$\hat{\mathbf{e}}_2 = \begin{pmatrix} 0.162 \\ 0.591 \\ -0.790 \end{pmatrix}$$

و

$$\hat{\mathbf{e}}_3 = \begin{pmatrix} 0.712 \\ -0.625 \\ -0.321 \end{pmatrix}.$$

می‌توان متعامد بودن بردارها را بررسی کرد. برای مثال،

$$\hat{\mathbf{e}}_1' \hat{\mathbf{e}}_2 = (0.683)(0.162) + (0.510)(0.591) + (0.522)(-0.790).$$

پس

$$\hat{\mathbf{e}}_1' \hat{\mathbf{e}}_2 \approx 0.111 + 0.301 - 0.412 \approx 0.$$

همچنین

$$\hat{\mathbf{e}}_1' \hat{\mathbf{e}}_3 \approx 0$$

و

$$\hat{\mathbf{e}}_2' \hat{\mathbf{e}}_3 \approx 0.$$

بنابراین بردارهای ویژه دو به دو متعامد هستند.

مرحله چهارم: تشکیل مؤلفه‌های اصلی مؤلفه اصلی اول برابر است با

$$\hat{Y}_1 = \hat{\mathbf{e}}_1' \mathbf{x}.$$

پس

$$\hat{Y}_1 = 0.683X_1 + 0.510X_2 + 0.522X_3.$$

با توجه به تعریف متغیرها، داریم

$$\hat{Y}_1 = 0.683 \ln(\text{طول}) + 0.510 \ln(\text{عرض}) + 0.522 \ln(\text{ارتفاع}).$$

مؤلفه اصلی دوم برابر است با

$$\hat{Y}_2 = 0.162X_1 + 0.591X_2 - 0.790X_3.$$

یعنی

$$\hat{Y}_2 = 0.162 \ln(\text{طول}) + 0.591 \ln(\text{عرض}) - 0.790 \ln(\text{ارتفاع}).$$

مؤلفه اصلی سوم برابر است با

$$\hat{Y}_3 = 0.712X_1 - 0.625X_2 - 0.321X_3.$$

یعنی

$$\hat{Y}_3 = 0.712 \ln(\text{طول}) - 0.625 \ln(\text{عرض}) - 0.321 \ln(\text{ارتفاع}).$$

واریانس مؤلفه‌ها برابر مقادیر ویژه متناظر است:

$$\begin{aligned} s_{\hat{Y}_1}^2 &= 0.0243095, \\ s_{\hat{Y}_2}^2 &= 0.0006246, \\ s_{\hat{Y}_3}^2 &= 0.0003789. \end{aligned}$$

مرحله پنجم: سهم واریانس مؤلفه‌ها سهم مؤلفه اول از واریانس کل برابر است با

$$\begin{aligned} P_1 &= \frac{0.0243095}{0.025313} \\ &= 0.9604. \end{aligned}$$

بنابراین

از واریانس کل را توضیح می‌دهد. 96.04% مؤلفه اول حدود

برای مؤلفه دوم داریم:

$$\begin{aligned} P_2 &= \frac{0.0006246}{0.025313} \\ &= 0.0247. \end{aligned}$$

پس

از واریانس کل را توضیح می‌دهد. 2.47% مؤلفه دوم حدود

برای مؤلفه سوم:

$$\begin{aligned} P_3 &= \frac{0.0003789}{0.025313} \\ &= 0.0150. \end{aligned}$$

در نتیجه

از واریانس کل را توضیح می‌دهد. 1.50% مؤلفه سوم حدود

درصد تجمعی دو مؤلفه اول برابر است با

$$96.04\% + 2.47\% = 98.51\%.$$

بنابراین

از واریانس کل را توضیح می‌دهند. 98.51% دو مؤلفه اول در مجموع حدود

جدول نتایج به صورت زیر است:

مؤلفه	مقدار ویژه	درصد واریانس	درصد تجمعی
$\hat{Y}_1$	0.0243095	96.04	96.04
$\hat{Y}_2$	0.0006246	2.47	98.51
$\hat{Y}_3$	0.0003789	1.50	100.00

مشاهده می‌شود که تنها مؤلفه اول بیش از 96% واریانس کل را توضیح می‌دهد. بنابراین، در این مثال حتی استفاده از یک مؤلفه اصلی نیز می‌تواند خلاصه بسیار مناسبی از سه اندازه اولیه ارائه دهد.

مرحله ششم: تفسیر مؤلفه اصلی اول مؤلفه اول به صورت

$$\hat{Y}_1 = 0.683 \ln(\text{طول}) + 0.510 \ln(\text{عرض}) + 0.522 \ln(\text{ارتفاع})$$

تعریف شده است.

تمام ضرایب این مؤلفه مثبت و از نظر اندازه نسبتاً نزدیک به یکدیگر هستند. بنابراین، هرگاه طول، عرض و ارتفاع لاک‌پشت هم‌زمان افزایش یابند، مقدار $\hat{Y}_1$ نیز افزایش می‌یابد.

با استفاده از خاصیت لگاریتم داریم

$$a \ln(u) + b \ln(v) + c \ln(w) = \ln(u^a v^b w^c).$$

پس

$$\begin{aligned} \hat{Y}_1 &= 0.683 \ln(\text{طول}) + 0.510 \ln(\text{عرض}) + 0.522 \ln(\text{ارتفاع}) \\ &= \ln \left[(\text{طول})^{0.683} (\text{عرض})^{0.510} (\text{ارتفاع})^{0.522} \right]. \end{aligned}$$

بنابراین

$$\hat{Y}_1 = \ln \left[(\text{ارتفاع})^{0.522} (\text{عرض})^{0.510} (\text{طول})^{0.683} \right].$$

از نظر زیستی، این مؤلفه را می‌توان یک شاخص کلی از اندازه بدن لاک‌پشت در نظر گرفت. هرچه مقدار $\hat{Y}_1$ بزرگ‌تر باشد، لاک‌پشت در مجموع بزرگ‌تر است. از آنجا که ضرایب سه متغیر تقریباً مشابه هستند، مؤلفه اول نوعی میانگین وزنی از سه اندازه لگاریتمی است. همچنین عبارت داخل لگاریتم را می‌توان نوعی حجم تعدیل‌شده یا شاخص هندسی اندازه در نظر گرفت:

$$(\text{ارتفاع})^{0.522} (\text{عرض})^{0.510} (\text{طول})^{0.683}.$$

مرحله هفتم: تفسیر مؤلفه دوم مؤلفه دوم برابر است با

$$\hat{Y}_2 = 0.162 \ln(\text{طول}) + 0.591 \ln(\text{عرض}) - 0.790 \ln(\text{ارتفاع}).$$

در این مؤلفه، ضریب عرض مثبت و ضریب ارتفاع منفی و بزرگ است. بنابراین مؤلفه دوم نوعی تقابل میان عرض و ارتفاع را نشان می‌دهد.

لاک‌پشتی با مقدار بزرگ $\hat{Y}_2$ معمولاً نسبت به ارتفاع خود عرض بیشتری دارد، در حالی که مقدار کوچک $\hat{Y}_2$ نشان‌دهنده ارتفاع بیشتر نسبت به عرض است.

با استفاده از خواص لگاریتم می‌توان نوشت:

$$\hat{Y}_2 = \ln \left[\frac{(\text{طول})^{0.162} (\text{عرض})^{0.591}}{(\text{ارتفاع})^{0.790}} \right].$$

پس

$$\hat{Y}_2 = \ln \left[\frac{(\text{طول})^{0.162} (\text{عرض})^{0.591}}{(\text{ارتفاع})^{0.790}} \right].$$

این مؤلفه بیشتر یک شاخص شکل هندسی است تا شاخص اندازه.

مرحله هشتم: تفسیر مؤلفه سوم مؤلفه سوم برابر است با

$$\hat{Y}_3 = 0.712 \ln(\text{طول}) - 0.625 \ln(\text{عرض}) - 0.321 \ln(\text{ارتفاع}).$$

در نتیجه،

$$\hat{Y}_3 = \ln \left[\frac{(\text{طول})^{0.712}}{(\text{ارتفاع})^{0.321} (\text{عرض})^{0.625}} \right].$$

این مؤلفه تفاوت میان طول و ترکیبی از عرض و ارتفاع را نشان می‌دهد و می‌تواند به عنوان شاخصی از کشیدگی لاک پشت تفسیر شود.

با این حال، مؤلفه سوم تنها حدود 1.50% از واریانس کل را توضیح می‌دهد؛ بنابراین اهمیت آن در خلاصه‌سازی داده‌ها بسیار کم است.

مرحله نهم: محاسبه همبستگی متغیرها با مؤلفه‌ها برای محاسبه همبستگی بین متغیرهای اولیه و مؤلفه‌ها از رابطه

$$r_{\hat{Y}_i, X_k} = \frac{\hat{e}_{ki} \sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}$$

استفاده می‌کنیم.

ابتدا انحراف معیارهای سه متغیر را محاسبه می‌کنیم:

$$s_{X_1} = \sqrt{0.011555} = 0.1075,$$

$$s_{X_2} = \sqrt{0.006697} = 0.0818,$$

$$s_{X_3} = \sqrt{0.007061} = 0.0840.$$

همچنین

$$\sqrt{\hat{\lambda}_1} = \sqrt{0.0243095} = 0.1559.$$

بنابراین برای متغیر اول داریم:

$$\begin{aligned} r_{\hat{Y}_1, X_1} &= \frac{(0.683)(0.1559)}{0.1075} \\ &= 0.991. \end{aligned}$$

برای متغیر دوم:

$$\begin{aligned} r_{\hat{Y}_1, X_2} &= \frac{(0.510)(0.1559)}{0.0818} \\ &= 0.972. \end{aligned}$$

و برای متغیر سوم:

$$r_{\hat{Y}_1, X_3} = \frac{(0.522)(0.1559)}{0.0840} \\ = 0.969.$$

پس

$$\begin{aligned} r_{\hat{Y}_1, X_1} &= 0.991, \\ r_{\hat{Y}_1, X_2} &= 0.972, \\ r_{\hat{Y}_1, X_3} &= 0.969. \end{aligned}$$

این همبستگی‌های بسیار بزرگ و مثبت نشان می‌دهند که مؤلفه اول تقریباً تمام اطلاعات مربوط به اندازه مشترک سه متغیر را در خود خلاصه می‌کند.

برای مؤلفه دوم داریم:

$$\sqrt{\hat{\lambda}_2} = \sqrt{0.0006246} = 0.0250.$$

بنابراین،

$$\begin{aligned} r_{\hat{Y}_2, X_1} &= 0.038, \\ r_{\hat{Y}_2, X_2} &= 0.181, \\ r_{\hat{Y}_2, X_3} &= -0.235. \end{aligned}$$

برای مؤلفه سوم نیز داریم:

$$\sqrt{\hat{\lambda}_3} = \sqrt{0.0003789} = 0.0195,$$

و در نتیجه

$$\begin{aligned} r_{\hat{Y}_3, X_1} &= 0.129, \\ r_{\hat{Y}_3, X_2} &= -0.149, \\ r_{\hat{Y}_3, X_3} &= -0.074. \end{aligned}$$

جدول کامل ضرایب بردارهای ویژه و همبستگی‌ها به صورت زیر است:

متغیر	$\hat{e}_{k1}$	$r_{\hat{Y}_1, X_k}$	$\hat{e}_{k2}$	$r_{\hat{Y}_2, X_k}$	$\hat{e}_{k3}$	$r_{\hat{Y}_3, X_k}$
لگاریتم طول	0.683	0.991	0.162	0.038	0.712	0.129
لگاریتم عرض	0.510	0.972	0.591	0.181	-0.625	-0.149
لگاریتم ارتفاع	0.522	0.969	-0.790	-0.235	-0.321	-0.074

نتیجه‌گیری آماری مثال نتایج این مثال نشان می‌دهند:

- مؤلفه اول به تنهایی حدود 96.04% از واریانس کل را توضیح می‌دهد.
- هر سه متغیر طول، عرض و ارتفاع با مؤلفه اول همبستگی بسیار بزرگی دارند.
- مؤلفه اول را می‌توان به عنوان شاخص کلی اندازه لاک‌پشت تفسیر کرد.
- مؤلفه دوم بیشتر تفاوت شکل میان عرض و ارتفاع را نشان می‌دهد.
- مؤلفه سوم نیز نوعی کشیدگی طولی را نشان می‌دهد، اما سهم واریانس آن بسیار اندک است.
- برای هدف کاهش بعد، جایگزینی سه متغیر اولیه با تنها مؤلفه اول کاملاً قابل قبول است.

در نتیجه، می‌توان داده‌های سه‌بعدی هر لاک‌پشت را با یک نمره اصلی

$$\hat{Y}_1$$

خلاصه کرد و تنها حدود 3.96% از واریانس کل را از دست داد.

۱۸.۳.۹ حل نرم‌افزاری مثال لاک‌پشت‌ها در R

در این مثال، ماتریس واریانس-کوواریانس نمونه مستقیماً در اختیار است. بنابراین تحلیل مؤلفه‌های اصلی را با تابع eigen انجام می‌دهیم.

Listing 42:

```

1 S_turtle <- 1e-3 * matrix(
2   c(
3     11.555, 8.367, 8.508,
4     8.367, 6.697, 6.264,
5     8.508, 6.264, 7.061
6   ),
7   nrow = 3,
8   byrow = TRUE
9 )
10
11 variable_names <- c(

```

```

12 "LogLength",
13 "LogWidth",
14 "LogHeight"
15 )
16
17 rownames(S_turtle) <- variable_names
18 colnames(S_turtle) <- variable_names
19
20 S_turtle

```

خروجی:

Listing 43:

	LogLength	LogWidth	LogHeight
LogLength	0.011555	0.008367	0.008508
LogWidth	0.008367	0.006697	0.006264
LogHeight	0.008508	0.006264	0.007061

Listing 44:

```

1 eig_turtle <- eigen(
2   S_turtle,
3   symmetric = TRUE
4 )
5
6 lambda_turtle <- eig_turtle$values
7 E_turtle <- eig_turtle$vectors
8
9 # Adjust signs for easier interpretation
10 if (E_turtle[1, 1] < 0) {
11   E_turtle[, 1] <- -E_turtle[, 1]
12 }
13
14 if (E_turtle[1, 2] < 0) {
15   E_turtle[, 2] <- -E_turtle[, 2]
16 }
17
18 if (E_turtle[1, 3] < 0) {
19   E_turtle[, 3] <- -E_turtle[, 3]
20 }
21

```

```

22 rownames(E_turtle) <- variable_names
23 colnames(E_turtle) <- c("PC1", "PC2", "PC3")
24
25 cat("Eigenvalues:\n")
26 print(round(lambda_turtle, 7))
27
28 cat("\nEigenvectors:\n")
29 print(round(E_turtle, 3))

```

خروجی:

Listing 45:

```
[1] 0.0243095 0.0006246 0.0003789
```

خروجی بردارهای ویژه:

Listing 46:

	PC1	PC2	PC3
LogLength	0.683	0.162	0.712
LogWidth	0.510	0.591	-0.625
LogHeight	0.522	-0.790	-0.321

Listing 47:

```

round(
  t(E_turtle) %*% E_turtle,
  6
)

```

خروجی:

Listing 48:

	PC1	PC2	PC3
PC1	1	0	0
PC2	0	1	0
PC3	0	0	1

Listing 49:

```

Cov_PC <- t(E_turtle) %*%
          S_turtle %*%

```

```

3      E_turtle
4
5 round(Cov_PC, 7)

```

خروجی:

Listing 50:

	PC1	PC2	PC3
PC1	0.0243095	0.0000000	0.0000000
PC2	0.0000000	0.0006246	0.0000000
PC3	0.0000000	0.0000000	0.0003789

Listing 51:

```

1 total_variance <- sum(lambda_turtle)
2
3 variance_percent <-
4   100 * lambda_turtle / total_variance
5
6 cumulative_percent <-
7   cumsum(variance_percent)
8
9 variance_table <- data.frame(
10   Component = c("PC1", "PC2", "PC3"),
11   Eigenvalue = round(lambda_turtle, 7),
12   Variance_Percent =
13     round(variance_percent, 2),
14   Cumulative_Percent =
15     round(cumulative_percent, 2)
16 )
17
18 print(
19   variance_table,
20   row.names = FALSE
21 )

```

خروجی:

Listing 52:

Component	Eigenvalue	Variance_Percent	Cumulative_Percent
PC1	0.0243095	96.04	96.04

PC2	0.0006246	2.47	98.51
PC3	0.0003789	1.50	100.00

Listing 53:

```

1 sd_variables <- sqrt(
2   diag(S_turtle)
3 )
4
5 Cor_X_PC <- diag(
6   1 / sd_variables
7 ) %*%
8   E_turtle %*%
9   diag(
10    sqrt(lambda_turtle)
11  )
12
13 rownames(Cor_X_PC) <- variable_names
14 colnames(Cor_X_PC) <- c(
15   "PC1",
16   "PC2",
17   "PC3"
18 )
19
20 round(Cor_X_PC, 3)

```

خروجی:

Listing 54:

	PC1	PC2	PC3
LogLength	0.991	0.038	0.129
LogWidth	0.972	0.181	-0.149
LogHeight	0.969	-0.235	-0.074

Listing 55:

```

1 plot(
2   1:3,
3   lambda_turtle,
4   type = "b",
5   pch = 19,

```

```

6   lwd = 2,
7   xaxt = "n",
8   xlab = "Principal component",
9   ylab = "Eigenvalue",
10  main = "Scree plot for turtle data"
11 )
12
13 axis(
14     1,
15     at = 1:3,
16     labels = c("PC1", "PC2", "PC3")
17 )
18
19 text(
20     1:3,
21     lambda_turtle,
22     labels = signif(
23         lambda_turtle,
24         3
25     ),
26     pos = 3
27 )

```

نمودار اسکری افت بسیار شدیدی پس از مؤلفه اول نشان می‌دهد. بنابراین انتخاب یک مؤلفه اصلی کاملاً مناسب است.

Listing 56:

```

1 mid_points <- barplot(
2     variance_percent,
3     names.arg = c(
4         "PC1",
5         "PC2",
6         "PC3"
7     ),
8     ylim = c(0, 105),
9     xlab = "Principal component",
10    ylab = "Explained variance (%)",
11    main = "Explained variance"
12 )
13
14 lines(
15     mid_points,

```

```

16 cumulative_percent,
17 type = "b",
18 pch = 19,
19 lwd = 2
20 )
21
22 text(
23   mid_points,
24   variance_percent,
25   labels = paste0(
26     round(variance_percent, 2),
27     "%"
28   ),
29   pos = 3
30 )
31
32 legend(
33   "right",
34   legend = c(
35     "Individual variance",
36     "Cumulative variance"
37   ),
38   pch = c(15, 19),
39   lty = c(NA, 1),
40   bty = "n"
41 )

```

Listing 57:

```

1 plot(
2   Cor_X_PC[, 1],
3   Cor_X_PC[, 2],
4   xlim = c(-1, 1),
5   ylim = c(-1, 1),
6   asp = 1,
7   pch = 19,
8   xlab = "Correlation with PC1",
9   ylab = "Correlation with PC2",
10  main = "Variables in the PC1-PC2 plane"
11 )
12

```

```

13 abline(
14     h = 0,
15     v = 0,
16     lty = 3
17 )
18
19 symbols(
20     0,
21     0,
22     circles = 1,
23     inches = FALSE,
24     add = TRUE
25 )
26
27 arrows(
28     0,
29     0,
30     Cor_X_PC[, 1],
31     Cor_X_PC[, 2],
32     length = 0.08
33 )
34
35 text(
36     Cor_X_PC[, 1],
37     Cor_X_PC[, 2],
38     labels = variable_names,
39     pos = 3
40 )

```

در این نمودار، هر سه متغیر در نزدیکی جهت مثبت مؤلفه اول قرار می‌گیرند. این موضوع نشان می‌دهد که مؤلفه اول یک بعد مشترک مربوط به اندازه کلی لاک‌پشت را اندازه‌گیری می‌کند.

مقایسه با حالت داده‌های خام اگر داده‌های خام 24 لاک‌پشت در اختیار باشند، می‌توان تحلیل را با تابع `prcomp` نیز انجام داد:

Listing 58: PCA

```

1 # Suppose turtle_data contains:
2 # LogLength, LogWidth, LogHeight
3
4 pca_turtle <- prcomp(
5     turtle_data,

```

```

6   center = TRUE,
7   scale. = FALSE
8 )
9
10 summary(pca_turtle)
11
12 pca_turtle$rotation
13
14 head(pca_turtle$x)

```

در این حالت:

- عبارت `pca_turtle$rotation` بردارهای ویژه را نشان می‌دهد؛
- عبارت `pca_turtle$sdev^2` مقادیر ویژه را ارائه می‌کند؛
- عبارت `pca_turtle$x` شامل نمره‌های مؤلفه‌های اصلی برای هر لاک‌پشت است.

اما در مثال حاضر، چون تنها ماتریس کوواریانس نمونه در اختیار است، تابع `eigen` روش مستقیم و مناسب تحلیل است.

تمرین‌های پیشنهادی

برای تمرین و تثبیت مطالب این فصل، تمرین‌های 1 تا 7 فصل هشتم کتاب چند متغیره جانسون - Applied Multivariate Statistical Analysis تألیف Johnson and Wichern را با استفاده از مفاهیم و روش‌های ارائه‌شده در این فصل حل کنید.

۱۹.۳.۹ نمونه تمرین: تمرین ۸-۶: تحلیل مؤلفه‌های اصلی برای دو متغیر مالی

صورت تمرین داده‌های مربوط به دارایی X_1 و درآمد خالص X_2 برای تعدادی از بزرگ‌ترین شرکت‌های صنعتی در اختیار است. بردار میانگین و ماتریس واریانس-کوواریانس نمونه به صورت زیر داده شده‌اند:

$$\bar{\mathbf{x}} = \begin{pmatrix} 19.32 \\ 1.51 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} 70.41 & 5.87 \\ 5.87 & 0.97 \end{pmatrix}.$$

۱. مؤلفه‌های اصلی نمونه و واریانس آنها را به دست آورید.

۲. نسبت واریانس کل نمونه که توسط مؤلفه اول توضیح داده می‌شود را محاسبه کنید.

۳. بیضی

$$(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) = 1.4$$

را بررسی کرده و جهت مؤلفه‌های اصلی را مشخص کنید.

۴. ضرایب همبستگی بین مؤلفه اول و دو متغیر اولیه را محاسبه و تفسیر کنید.

حل الف) محاسبه مقادیر ویژه
مقادیر ویژه از حل معادله

$$|\mathbf{S} - \lambda \mathbf{I}| = 0$$

به دست می‌آیند:

$$\begin{vmatrix} 70.41 - \lambda & 5.87 \\ 5.87 & 0.97 - \lambda \end{vmatrix} = 0.$$

پس

$$(70.41 - \lambda)(0.97 - \lambda) - 5.87^2 = 0.$$

با بسط داریم

$$\lambda^2 - 71.38\lambda + 33.8408 = 0.$$

بنابراین

$$\lambda = \frac{71.38 \pm \sqrt{71.38^2 - 4(33.8408)}}{2},$$

و در نتیجه

$$\boxed{\hat{\lambda}_1 = 70.9027, \quad \hat{\lambda}_2 = 0.4773.}$$

برای کنترل:

$$\hat{\lambda}_1 + \hat{\lambda}_2 = 71.38 = 70.41 + 0.97 = \text{tr}(\mathbf{S}).$$

بردار ویژه مربوط به $\hat{\lambda}_1$
از

$$(\mathbf{S} - 70.9027\mathbf{I})\hat{\mathbf{e}}_1 = 0$$

داریم

$$\begin{pmatrix} -0.4927 & 5.87 \\ 5.87 & -69.9327 \end{pmatrix} \begin{pmatrix} e_{11} \\ e_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

از معادله اول:

$$-0.4927e_{11} + 5.87e_{21} = 0,$$

پس

$$e_{21} = 0.0839e_{11}.$$

پس از واحد کردن بردار:

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 0.9965 \\ 0.0836 \end{pmatrix}.$$

بنابراین مؤلفه اصلی اول برابر است با

$$\hat{Y}_1 = 0.9965X_1 + 0.0836X_2.$$

و واریانس آن برابر است با

$$s_{\hat{Y}_1}^2 = 70.9027.$$

بردار ویژه مربوط به $\hat{\lambda}_2$

برای مقدار ویژه دوم، بردار ویژه واحد عمود بر بردار اول است. بنابراین می‌توان نوشت

$$\hat{\mathbf{e}}_2 = \begin{pmatrix} 0.0836 \\ -0.9965 \end{pmatrix}.$$

پس مؤلفه اصلی دوم برابر است با

$$\hat{Y}_2 = 0.0836X_1 - 0.9965X_2.$$

و

$$s_{\hat{Y}_2}^2 = 0.4773.$$

ب) سهم واریانس مؤلفه اول واریانس کل نمونه برابر است با

$$70.9027 + 0.4773 = 71.38.$$

سهم مؤلفه اول از واریانس کل برابر است با

$$\frac{\hat{\lambda}_1}{\hat{\lambda}_1 + \hat{\lambda}_2} = \frac{70.9027}{71.38} = 0.9933.$$

بنابراین

از واریانس کل را توضیح می‌دهد. 99.33% مؤلفه اول حدود

و مؤلفه دوم فقط

$$0.67\%$$

از واریانس کل را توضیح می‌دهد.

بنابراین برای کاهش بعد، استفاده از تنها مؤلفه اول تقریباً تمام اطلاعات تغییرپذیری داده‌ها را حفظ می‌کند.

ج) تعبیر هندسی بیضی بیضی نمونه‌ای از رابطه

$$(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) = 1.4$$

به دست می‌آید.

ابتدا معکوس ماتریس کوواریانس برابر است با

$$\mathbf{S}^{-1} \approx \begin{pmatrix} 0.02866 & -0.17346 \\ -0.17346 & 2.08062 \end{pmatrix}.$$

مرکز بیضی برابر بردار میانگین است:

$$(19.32, 1.51).$$

جهت محور بزرگ بیضی توسط بردار ویژه اول

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 0.9965 \\ 0.0836 \end{pmatrix}$$

و جهت محور کوچک توسط

$$\hat{\mathbf{e}}_2 = \begin{pmatrix} 0.0836 \\ -0.9965 \end{pmatrix}$$

مشخص می‌شود.

چون

$$c^2 = 1.4, \quad c = \sqrt{1.4},$$

طول نیم‌محورهای بیضی برابر است با

$$a_1 = \sqrt{1.4\hat{\lambda}_1} = \sqrt{1.4(70.9027)} \approx 9.963,$$

و

$$a_2 = \sqrt{1.4\hat{\lambda}_2} = \sqrt{1.4(0.4773)} \approx 0.817.$$

بنابراین

$$a_1 \approx 9.963, \quad a_2 \approx 0.817.$$

زاویه محور اصلی اول نسبت به محور X_1 نیز تقریباً برابر است با

$$\theta = \tan^{-1} \left(\frac{0.0836}{0.9965} \right) \approx 4.8^\circ.$$

پس محور بزرگ بیضی تقریباً در راستای محور دارایی قرار دارد.

(د) همبستگی مؤلفه اول با متغیرهای اولیه از رابطه

$$r_{\hat{Y}_i, X_k} = \frac{\hat{e}_{ki} \sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}$$

استفاده می‌کنیم.

برای X_1 داریم

$$r_{\hat{Y}_1, X_1} = \frac{0.9965 \sqrt{70.9027}}{\sqrt{70.41}} \approx 0.99998.$$

پس

$$r_{\hat{Y}_1, X_1} \approx 1.000.$$

برای X_2 داریم

$$r_{\hat{Y}_1, X_2} = \frac{0.0836\sqrt{70.9027}}{\sqrt{0.97}} \approx 0.715.$$

پس

$$r_{\hat{Y}_1, X_2} \approx 0.715.$$

بنابراین مؤلفه اصلی اول تقریباً همبستگی کامل با متغیر دارایی دارد و همبستگی مثبت نسبتاً قوی نیز با درآمد خالص دارد.

از طرف دیگر، چون ضریب X_1 در مؤلفه اول بسیار بزرگتر از ضریب X_2 است،

$$\hat{Y}_1 = 0.9965X_1 + 0.0836X_2,$$

مؤلفه اول عمدتاً تحت تأثیر متغیر دارایی قرار دارد.

در نتیجه می‌توان مؤلفه اول را به طور تقریبی یک شاخص اندازه مالی شرکت تفسیر کرد؛ شرکت‌هایی که دارایی بیشتری دارند معمولاً مقدار بزرگتری از مؤلفه اول نیز خواهند داشت.

جمع‌بندی در این مثال

$$\hat{Y}_1 = 0.9965X_1 + 0.0836X_2$$

به تنهایی حدود

$$99.33\%$$

از واریانس کل را توضیح می‌دهد. بنابراین دو متغیر دارایی و درآمد خالص را می‌توان، با از دست دادن بسیار اندکی از تغییرپذیری، با یک مؤلفه اصلی خلاصه کرد.

حل نرم‌افزاری تمرین ۸-۶ در **R** ماتریس واریانس-کوواریانس نمونه و بردار میانگین را در نرم‌افزار **R** به صورت زیر تعریف می‌کنیم:

Listing 59:

```
xbar <- c(19.32, 1.51)
```

```

2
3 S <- matrix(
4   c(
5     70.41, 5.87,
6     5.87, 0.97
7   ),
8   nrow = 2,
9   byrow = TRUE
10 )
11
12 rownames(S) <- c("X1", "X2")
13 colnames(S) <- c("X1", "X2")
14
15 xbar
16 S

```

خروجی:

Listing 60:

```

[1] 19.32  1.51

      X1  X2
X1 70.41 5.87
X2  5.87 0.97

```

Listing 61:

```

1 eig <- eigen(S, symmetric = TRUE)
2
3 lambda <- eig$values
4 E <- eig$vectors
5
6 # Change signs for consistency with the hand solution
7 if (E[1, 1] < 0) {
8   E[, 1] <- -E[, 1]
9 }
10
11 if (E[2, 2] > 0) {
12   E[, 2] <- -E[, 2]
13 }
14

```

```

15 colnames(E) <- c("PC1", "PC2")
16 rownames(E) <- c("X1", "X2")
17
18 cat("Eigenvalues:\n")
19 print(round(lambda, 6))
20
21 cat("\nEigenvectors:\n")
22 print(round(E, 6))

```

خروجی:

Listing 62:

```
[1] 70.902715  0.477285
```

و بردارهای ویژه:

Listing 63:

	PC1	PC2
X1	0.996496	0.083644
X2	0.083644	-0.996496

بنابراین مؤلفه‌های اصلی نرم‌افزار نیز همان نتایج محاسبات دستی را تأیید می‌کنند:

$$\hat{Y}_1 = 0.9965X_1 + 0.0836X_2$$

و

$$\hat{Y}_2 = 0.0836X_1 - 0.9965X_2.$$

Listing 64:

```

1 for (i in 1:2) {
2
3   left_side <- S %*% E[, i]
4   right_side <- lambda[i] * E[, i]
5
6   cat("\nComponent", i, "\n")
7   print(
8     round(
9       cbind(left_side, right_side),
10      6

```

```

11 )
12 )
13 }

```

خروجی دو ستون برای هر مؤلفه تقریباً یکسان است و رابطه

$$S\hat{e}_i = \hat{\lambda}_i \hat{e}_i$$

را تأیید می‌کند.

Listing 65:

```

round(
  t(E) %*% E,
  6
)

```

خروجی:

Listing 66:

	PC1	PC2
PC1	1	0
PC2	0	1

بنابراین دو بردار ویژه واحد و متعامد هستند.

بررسی ماتریس کوواریانس مؤلفه‌های اصلی از نظر نظری داریم

$$\text{Cov}(\hat{\mathbf{Y}}) = \mathbf{E}' \mathbf{S} \mathbf{E}.$$

در R:

Listing 67:

```

Cov_PC <- t(E) %*% S %*% E

round(
  Cov_PC,
  6
)

```

خروجی:

Listing 68:

	PC1	PC2
PC1	70.902715	0.000000
PC2	0.000000	0.477285

پس مؤلفه‌ها ناهمبسته‌اند و واریانس هر مؤلفه برابر مقدار ویژه متناظر است.

Listing 69:

```
1 variance_percent <-  
2   100 * lambda / sum(lambda)  
3  
4 cumulative_percent <-  
5   cumsum(variance_percent)  
6  
7 variance_table <- data.frame(  
8   Component = c("PC1", "PC2"),  
9   Eigenvalue = round(lambda, 6),  
10  Variance_Percent =  
11    round(variance_percent, 2),  
12  Cumulative_Percent =  
13    round(cumulative_percent, 2)  
14 )  
15  
16 print(  
17   variance_table,  
18   row.names = FALSE  
19 )
```

خروجی:

Listing 70:

Component	Eigenvalue	Variance_Percent	Cumulative_Percent
PC1	70.902715	99.33	99.33
PC2	0.477285	0.67	100.00

این خروجی نشان می‌دهد مؤلفه اول به تنهایی حدود

99.33%

از واریانس کل را توضیح می‌دهد.

محاسبه همبستگی متغیرها با مؤلفه‌های اصلی همبستگی‌ها با استفاده از رابطه ماتریسی

$$R_{X,Y} = D^{-1} E \Lambda^{1/2}$$

محاسبه می‌شوند، که در آن D ماتریس قطری انحراف معیار متغیرهای اولیه است.

Listing 71:

```
sd_X <- sqrt(diag(S))  
  
Cor_X_PC <- diag(  
  1 / sd_X  
) %*%  
  E %*%  
  diag(  
    sqrt(lambda)  
  )  
  
rownames(Cor_X_PC) <- c("X1", "X2")  
colnames(Cor_X_PC) <- c("PC1", "PC2")  
  
round(  
  Cor_X_PC,  
  3  
)
```

خروجی:

Listing 72:

	PC1	PC2
X1	1.000	0.007
X2	0.715	-0.699

پس

$$r_{\hat{Y}_1, X_1} \approx 1.000$$

و

$$r_{\hat{Y}_1, X_2} \approx 0.715.$$

همچنین مؤلفه دوم تقریباً هیچ ارتباطی با X_1 ندارد و بیشتر در جهت تغییرات باقی مانده X_2 قرار می‌گیرد.

محاسبه مشخصات هندسی بیضی برای بیضی

$$(\mathbf{x} - \bar{\mathbf{x}})' S^{-1} (\mathbf{x} - \bar{\mathbf{x}}) = 1.4$$

ابتدا معکوس ماتریس و سپس طول نیم‌محورها را محاسبه می‌کنیم.

Listing 73:

```

1 S_inv <- solve(S)
2
3 c2 <- 1.4
4 c_value <- sqrt(c2)
5
6 axis_lengths <-
7   sqrt(c2 * lambda)
8
9 angle_pc1 <-
10  atan2(
11    E[2, 1],
12    E[1, 1]
13  ) * 180 / pi
14
15 cat("Inverse covariance matrix:\n")
16 print(round(S_inv, 6))
17
18 cat("\nSemi-axis lengths:\n")
19 print(round(axis_lengths, 6))
20
21 cat("\nAngle of PC1:\n")
22 print(round(angle_pc1, 3))

```

خروجی:

Listing 74:

```

Inverse covariance matrix:
      X1      X2
X1  0.028664 -0.173459

```

```
X2 -0.173459  2.080625
```

```
Semi-axis lengths:
```

```
[1] 9.963122 0.817434
```

```
Angle of PC1:
```

```
[1] 4.798
```

بنابراین نیم‌محور بزرگ تقریباً برابر

9.963

و نیم‌محور کوچک تقریباً برابر

0.817

است. همچنین محور اصلی اول حدود

4.8°

با محور X_1 زاویه دارد.

رسم بیضی و محورهای مؤلفه‌های اصلی با بسته ellipse برای رسم مستقیم بیضی می‌توان از بسته ellipse استفاده کرد.

Listing 75:

```
1 # install.packages("ellipse")
2 library(ellipse)
3
4 ellipse_points <- ellipse::ellipse(
5   S,
6   centre = xbar,
7   t = 1.4,
8   npoints = 500
9 )
10
11 axis1 <-
12   sqrt(1.4 * lambda[1]) *
13   E[, 1]
```

```

15 axis2 <-
16   sqrt(1.4 * lambda[2]) *
17   E[, 2]
18
19 plot(
20   ellipse_points,
21   type = "l",
22   lwd = 2,
23   asp = 1,
24   xlab = expression(X[1]),
25   ylab = expression(X[2]),
26   main = "PCA ellipse and principal axes"
27 )
28
29 points(
30   xbar[1],
31   xbar[2],
32   pch = 19
33 )
34
35 arrows(
36   xbar[1] - axis1[1],
37   xbar[2] - axis1[2],
38   xbar[1] + axis1[1],
39   xbar[2] + axis1[2],
40   length = 0.08,
41   lwd = 2
42 )
43
44 arrows(
45   xbar[1] - axis2[1],
46   xbar[2] - axis2[2],
47   xbar[1] + axis2[1],
48   xbar[2] + axis2[2],
49   length = 0.08,
50   lwd = 2,
51   lty = 2
52 )
53
54 text(
55   xbar[1] + axis1[1],
56   xbar[2] + axis1[2],

```

```

57 labels = "PC1",
58 pos = 4
59 )
60
61 text(
62 xbar[1] + axis2[1],
63 xbar[2] + axis2[2],
64 labels = "PC2",
65 pos = 4
66 )
67
68 legend(
69 "topright",
70 legend = c(
71 "Ellipse",
72 "PC1",
73 "PC2"
74 ),
75 lty = c(1, 1, 2),
76 lwd = c(2, 2, 2),
77 bty = "n"
78 )

```

در شکل حاصل، محور بلند بیضی تقریباً در جهت متغیر X_1 قرار می‌گیرد؛ این نتیجه با مقدار بسیار بزرگ ضریب

0.9965

برای X_1 در مؤلفه اول و همچنین همبستگی تقریباً یک بین X_1 و $\hat{Y}_1$ کاملاً سازگار است.