آمار و احتمال ۱

فصل چهارم

توصیف مقداری مشاهدات طبقه‌بندی‌شده

میانگین و پراکندگی طبقات · چولگی · کشیدگی

هدف فصل

محاسبه شاخص‌ها از جدول فراوانی طبقه‌بندی‌شده

پیوند با فصول قبل

هنگامی که تعداد مشاهدات زیاد باشد، داده‌ها را در طبقات یا بازه‌ها خلاصه می‌کنیم.

در این حالت:

- تمام مقادیر خام را در اختیار نداریم
- شاخص‌ها را با **فراوانی طبقات** و **نماینده هر طبقه** محاسبه می‌کنیم

> شاخص‌های به‌دست‌آمده معمولاً **تقریبی**اند.

در پایان این فصل باید بتوانید

- اجزای جدول توزیع فراوانی طبقه‌بندی‌شده را تشخیص دهید
- میانگین، میانه و نمای تقریبی را محاسبه کنید
- واریانس و انحراف معیار طبقه‌بندی‌شده را به دست آورید
- از روش‌های ساده‌شده جبری استفاده کنید
- انحراف معیار و CV را در تصمیم‌های مدیریتی تفسیر کنید
- جهت و شدت چولگی را تشخیص دهید
- کشیدگی را به‌صورت مقدماتی تفسیر کنید

بخش اول: چیت‌شیت فصل

طبقات · شاخص‌های مرکزی و پراکندگی · شکل توزیع

دادهٔ طبقه‌بندی‌شده چیست؟

مشاهدات در چند بازه قرار می‌گیرند و به‌جای تک‌تک داده‌ها، تعداد مشاهدات هر بازه گزارش می‌شود.

مبلغ خرید	تعداد مشتریان
۰ تا کمتر از ۲	۵
۲ تا کمتر از ۴	۹
۴ تا کمتر از ۶	۱۲
۶ تا کمتر از ۸	۴

مقدار دقیق خرید هر مشتری مشخص نیست؛ فقط طبقهٔ او معلوم است.

اجزای یک طبقه

هر طبقه دارای این اجزاست:

- حد پایین طبقه
- حد بالای طبقه
- طول یا عرض طبقه
- مرکز طبقه
- فراوانی طبقه

برای طبقهٔ «۲۰ تا کمتر از ۳۰»:

حد پایین $= $20	حد بالا $= $30
طول $= $10	مرکز $= $25

مرکز طبقه

مرکز طبقه نمایندهٔ تمام مشاهدات آن طبقه در محاسبات است:

$$m_i = \frac{L_i + U_i}{2}$$

- L_i: حد پایین
- U_i: حد بالا
- m_i: مرکز طبقه

مرکز طبقهٔ ۴۰ تا کمتر از ۵۰ برابر **۴۵** است.

فرض: مشاهدات هر طبقه تقریباً اطراف مرکز پخش شده‌اند → پاسخ ممکن است با دادهٔ خام کمی فرق کند.

میانگین داده‌های طبقه‌بندی‌شده

$$\bar{x} = \frac{\sum f_i m_i}{\sum f_i} = \frac{\sum f_i m_i}{n}$$

مبلغ خرید	m_i	f_i	$f_i m_i$
۰ تا کمتر از ۱۰	۵	۲	۱۰
۱۰ تا کمتر از ۲۰	۱۵	۵	۷۵
۲۰ تا کمتر از ۳۰	۲۵	۳	۷۵
جمع		۱۰	۱۶۰

$$\bar{x} = \frac{160}{10} = 16$$

میانگین طبقه‌بندی‌شده = میانگین موزون

هر مرکز طبقه به اندازهٔ **فراوانی** آن در میانگین اثر دارد.

طبقه با فراوانی بیشتر ← اثر بیشتر بر میانگین نهایی

فراموش‌کردن فراوانی‌ها و میانگین سادهٔ مراکز طبقات معمولاً **نادرست** است.

طبقهٔ میانه

ابتدا $\frac{n}{2}$ را محاسبه کنید.

سپس اولین طبقه‌ای را بیابید که فراوانی تجمعی آن $\geq n/2$ باشد – این **طبقهٔ میانه** است.

$$Me = L + \left(\frac{\frac{n}{2} - F_{\text{قبل}}}{f_{Me}} \right) h$$

- L: حد پایین طبقهٔ میانه
- $F_{\text{قبل}}$: فراوانی تجمعی قبل از طبقهٔ میانه
- f_{Me}: فراوانی طبقهٔ میانه
- h: طول طبقه

فرض فرمول: توزیع یکنواخت داده‌ها داخل طبقهٔ میانه

طبقهٔ نما

طبقه‌ای که بیشترین فراوانی را دارد، **طبقهٔ نما** است.

$$Mo = L + \left(\frac{d_1}{d_1 + d_2}\right) h$$

$$d_1 = f_m - f_{m-1}, \qquad d_2 = f_m - f_{m+1}$$

- L: حد پایین طبقهٔ نما
- f_m: فراوانی طبقهٔ نما
- f_{m-1} و f_{m+1}: فراوانی طبقهٔ قبل و بعد
- h: طول طبقه

اگر فقط شناسایی محدودهٔ پرتکرار هدف باشد، معرفی طبقهٔ نما کافی است.

مقایسه شاخص‌های مرکزی

شاخص	روش تعیین	کاربرد
میانگین	مرکز تمام طبقات × فراوانی	سطح متوسط
میانه	طبقهٔ شامل مشاهدهٔ میانی	جداکنندهٔ نیمهٔ پایین و بالا
نما	طبقه با بیشترین فراوانی	پرتکرارترین محدوده

دامنهٔ تغییرات تقریبی

اگر داده‌های خام در دسترس نباشند:

$$R \approx U_{\text{آخر}} - L_{\text{اول}}$$

این مقدار تقریبی است؛ ممکن است کوچک‌ترین و بزرگ‌ترین مشاهده دقیقاً روی حدود طبقات نباشند.

واریانس داده‌های طبقه‌بندی‌شده

مرکز طبقه نمایندهٔ مشاهدات آن طبقه است.

جامعه

$$\sigma^2 = \frac{\sum f_i (m_i - \mu)^2}{N}$$

نمونه

$$s^2 = \frac{\sum f_i (m_i - \bar{x})^2}{n - 1}$$

مراحل: مرکز → میانگین → فاصله از میانگین → توان دو → ضرب در f_i → جمع و تقسیم

انحراف معیار طبقه‌بندی‌شده

$$s = \sqrt{s^2}$$

- انحراف معیار بیشتر ← پخش گسترده‌تر
- انحراف معیار کوچک‌تر ← معمولاً ثبات بیشتر

نتیجه از جدول طبقه‌بندی‌شده **تقریبی** است؛ چون مرکز طبقه جای مقادیر واقعی می‌نشیند.

روش کوتاه محاسبهٔ واریانس

جامعه:

$$\sigma^2 = \frac{\sum f_i m_i^2}{N} - \mu^2$$

نمونه:

$$s^2 = \frac{\sum f_i m_i^2 - \frac{(\sum f_i m_i)^2}{n}}{n-1}$$

روش مستقیم و کوتاه باید (با خطای گردکردن) پاسخ یکسانی بدهند.

میانگین فرضی و کدگذاری

اگر مراکز بزرگ باشند، مقدار نزدیک به مرکز را A می‌گیریم:

$$d_i = m_i - A \implies \bar{x} = A + \frac{\sum f_i d_i}{n}$$

اگر طول همهٔ طبقات برابر h باشد:

$$u_i = \frac{m_i - A}{h} \implies \bar{x} = A + h\left(\frac{\sum f_i u_i}{n}\right)$$

مقدار واقعی میانگین تغییر نمی‌کند؛ فقط محاسبات ساده‌تر می‌شود.

اثر تغییر واحد و مبدأ

اگر $Y = a + bX$:

$$\bar{Y} = a + b\bar{X}$$
$$Var(Y) = b^2 Var(X)$$
$$SD(Y) = |b| SD(X)$$

افزودن ثابت

میانگین تغییر می‌کند

واریانس و SD ثابت

ضرب در عدد

میانگین × همان عدد

واریانس × مجذور آن

اهمیت کاربردی انحراف معیار

میانگین به‌تنهایی برای تصمیم‌گیری کافی نیست.

دو سرمایه‌گذاری ممکن است میانگین بازده مشابه داشته باشند، اما یکی نوسان بسیار بیشتری داشته باشد.

کاربردها: نوسان بازده · ثبات فروش · تغییرات هزینه · یکنواختی تحویل · پایداری شعب

> انحراف معیار بالا همیشه بد نیست — تفسیر به **هدف تصمیم** بستگی دارد.

ضریب تغییرات

$$CV = \frac{s}{\bar{x}} \times 100$$

- CV کمتر ← ثبات نسبی بیشتر
- مفید برای مقایسهٔ ریسک نسبی سرمایه‌گذاری‌ها یا نوسان فروش شعب با اندازهٔ متفاوت

> وقتی میانگین‌ها متفاوت‌اند، CV را به‌جای مقایسهٔ مستقیم s به‌کار ببرید.

شکل توزیع

شاخص‌های مرکزی و پراکندگی تمام ویژگی‌ها را نشان نمی‌دهند.

دو مجموعه ممکن است میانگین و SD مشابه داشته باشند، اما شکل متفاوتی داشته باشند.

چولگی

میزان و جهت عدم تقارن

کشیدگی

تمرکز در مرکز و سنگینی دنباله‌ها

چولگی چیست؟

چولگی نشان می‌دهد توزیع تا چه اندازه متقارن است و دنبالهٔ بلندتر به کدام سمت قرار دارد.

توزیع متقارن

میانگین = میانه = نما

نام چولگی بر اساس **جهت دنبالهٔ بلندتر** تعیین می‌شود، نه محل تجمع بیشتر داده‌ها.

راست‌چوله و چپ‌چوله

راست‌چوله (مثبت)

دنبالهٔ بلند در سمت مقادیر بزرگ

نما < میانه < میانگین

مثال: توزیع درآمد

چپ‌چوله (منفی)

دنبالهٔ بلند در سمت مقادیر کوچک

میانگین < میانه < نما

مثال: چند شعبه با فروش خیلی پایین

ضریب چولگی پیرسون

$$Sk = \frac{\bar{x} - Mo}{s}$$

اگر نما قابل‌اعتماد نباشد:

$$Sk = \frac{3(\bar{x} - Me)}{s}$$

علامت	تفسیر
$Sk > 0$	چولگی مثبت
$Sk < 0$	چولگی منفی
$Sk \approx 0$	تقریباً متقارن

قدر مطلق بزرگ‌تر ← عدم تقارن بیشتر

چولگی گشتاوری

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

μ_3 گشتاور مرکزی مرتبهٔ سوم است.

- به تمام مشاهدات توجه می‌کند
- علامت آن جهت چولگی را نشان می‌دهد

برای این فصل: **مفهوم و تفسیر علامت** مهم‌تر از محاسبات طولانی است.

کشیدگی چیست؟

کشیدگی شکل قله و به‌ویژه میزان سنگینی دنباله‌ها را نسبت به نرمال توصیف می‌کند.

$$\beta_2 = \frac{\mu_4}{\sigma^4}$$

برای توزیع نرمال: $\beta_2 = 3$

کشیدگی اضافی:

$$\gamma_2 = \beta_2 - 3$$

برای نرمال: $\gamma_2 = 0$

انواع کشیدگی

لپتوکورتیک (کشیده) – $\beta_2 > 3$ یا $\gamma_2 > 0$
دنباله‌های سنگین‌تر؛ احتمال مقادیر بسیار دور از مرکز بیشتر

مزوکورتیک – $\beta_2 \approx 3$ یا $\gamma_2 \approx 0$
مشابه نرمال

پلاتیکورتیک (تخت) – $\beta_2 < 3$ یا $\gamma_2 < 0$
تخت‌تر؛ دنباله‌های سبک‌تر

در تحلیل مالی، کشیدگی بالا می‌تواند هشدار بازده یا زیان شدید باشد.

جمع‌بندی شکل توزیع

وضعیت	نشانهٔ اصلی	رابطهٔ معمول
متقارن	دو دنباله مشابه	میانگین = میانه = نما
راست‌چوله	دنبالهٔ بلند در سمت بزرگ	نما < میانه < میانگین
چپ‌چوله	دنبالهٔ بلند در سمت کوچک	میانگین < میانه < نما

چک‌لیست حل مسائل طبقه‌بندی‌شده

1. حدود و طول طبقات را بررسی کنید
2. مرکز هر طبقه را محاسبه کنید
3. $\sum f_i$ را به دست آورید
4. برای میانگین: ستون $f_i m_i$
5. برای میانه: فراوانی تجمعی
6. برای نما: بیشترین فراوانی
7. برای واریانس: $f_i(m_i - \bar{x})^2$ یا $f_i m_i^2$
8. s و در صورت نیاز CV
9. جهت چولگی را با مقایسهٔ میانگین، میانه و نما بررسی کنید
10. نتیجه را در زمینهٔ مسئلهٔ مدیریتی تفسیر کنید

بخش دوم: سؤالات چهارگزینه‌ای

ارزیابی محاسبات طبقه‌بندی‌شده، چولگی و کشیدگی

سؤال ۱

مرکز طبقهٔ ۲۰ تا کمتر از ۳۰ کدام است؟

الف) ۱۰

ب) ۲۰

ج) ۲۵

د) ۳۰

سؤال ۲

برای محاسبهٔ میانگین داده‌های طبقه‌بندی‌شده، هر مرکز طبقه در چه مقداری ضرب می‌شود؟

الف) حد پایین طبقه

ب) فراوانی طبقه

ج) طول طبقه

د) فراوانی تجمعی قبلی

سؤال ۳

اگر مراکز دو طبقه ۱۰ و ۲۰ و فراوانی آن‌ها به ترتیب ۲ و ۳ باشد، میانگین تقریبی کدام است؟

الف) ۱۴

ب) ۱۵

ج) ۱۶

د) ۳۰

سؤال ۴

طبقهٔ میانه چگونه تعیین می‌شود؟

الف) طبقه‌ای که کمترین فراوانی را دارد.

ب) اولین طبقه‌ای که فراوانی تجمعی ان بزرگتر یا مساوی $n/2$ است.

ج) آخرین طبقهٔ جدول.

د) طبقه‌ای که مرکز آن با میانگین برابر است.

سؤال ۵

طبقهٔ نما کدام است؟

الف) طبقه‌ای با بیشترین فراوانی

ب) طبقه‌ای با بیشترین حد بالا

ج) طبقه‌ای با کمترین طول

د) طبقه شامل میانگین

سؤال ۶

چرا میانگین محاسبه‌شده از داده‌های طبقه‌بندی‌شده معمولاً تقریبی است؟

الف) زیرا فراوانی‌ها استفاده نمی‌شوند.

ب) زیرا مرکز طبقه نمایندهٔ تمام داده‌های آن طبقه در نظر گرفته می‌شود.

ج) زیرا میانگین همیشه از میانه کمتر است.

د) زیرا تعداد مشاهدات مشخص نیست.

سؤال ۷

اگر به تمام مشاهدات یک مجموعه عدد ۵ اضافه شود، کدام گزینه درست است؟

الف) میانگین ۵ واحد افزایش می‌یابد و واریانس ثابت می‌ماند.

ب) میانگین ثابت می‌ماند و واریانس ۵ واحد افزایش می‌یابد.

ج) میانگین و واریانس هر دو ۵ برابر می‌شوند.

د) میانگین و واریانس هر دو ثابت می‌مانند.

سؤال ۸

دو شعبه میانگین فروش متفاوت دارند. برای مقایسهٔ پراکندگی نسبی فروش آن‌ها کدام شاخص مناسب‌تر است؟

الف) نما

ب) ضریب تغییرات

ج) میانه

د) فراوانی تجمعی

سؤال ۹

در یک توزیع راست‌چوله، کدام رابطه معمولاً برقرار است؟

الف) میانگین < میانه < نما

ب) نما < میانه < میانگین

ج) میانگین = میانه = نما

د) میانه < میانگین < نما

سؤال ۱۰

اگر ضریب چولگی یک توزیع منفی باشد، چه نتیجه‌ای می‌گیریم؟

الف) توزیع به سمت راست چوله است.

ب) توزیع به سمت چپ چوله است.

ج) انحراف معیار صفر است.

د) توزیع حتماً نرمال است.

سؤال ۱۱

کدام حالت نشان‌دهندهٔ توزیع با دنباله‌های سنگین‌تر از نرمال است؟

الف) $\beta_2 > 3$

ب) $\beta_2 = 0$

ج) $\beta_2 < 0$

د) $\gamma_2 < 0$

سؤال ۱۲

کدام گزینه اهمیت انحراف معیار را بهتر بیان می‌کند؟

الف) فقط بزرگ‌ترین مشاهده را نشان می‌دهد.

ب) تعداد طبقات را مشخص می‌کند.

ج) میزان پراکندگی داده‌ها در اطراف میانگین را نشان می‌دهد.

د) همیشه با میانگین برابر است.

بخش سوم: سؤال تشریحی چندبخشی

تحلیل مبلغ خرید مشتریان یک فروشگاه اینترنتی

صورت مسئله

مبلغ آخرین خرید ۸۰ مشتری (میلیون تومان):

مبلغ خرید	تعداد مشتریان
۰ تا کمتر از ۲	۸
۲ تا کمتر از ۴	۱۶
۴ تا کمتر از ۶	۲۴
۶ تا کمتر از ۸	۱۸
۸ تا کمتر از ۱۰	۱۰
۱۰ تا کمتر از ۱۲	۴
جمع	**۸۰**

اهداف مدیران

- شناسایی سطح معمول خرید مشتریان
- سنجش پراکندگی مبلغ خرید
- دسته‌بندی مشتریان برای طرح‌های وفاداری
- تفسیر شکل توزیع برای پیش‌بینی رفتار

از جدول طبقه‌بندی‌شده تا تصمیم سطح ویژه باشگاه مشتریان

بخش الف: تکمیل جدول محاسباتی

جدول را با این ستون‌ها کامل کنید:

1. حد پایین و حد بالای هر طبقه
2. مرکز هر طبقه m_i
3. فراوانی تجمعی
4. حاصل‌ضرب $f_i m_i$
5. حاصل‌ضرب $f_i m_i^2$

سپس بررسی کنید که $\sum f_i = 80$ باشد.

بخش ب: شاخص‌های مرکزی

1. میانگین تقریبی مبلغ خرید
2. طبقهٔ میانه
3. مقدار تقریبی میانه (فرمول طبقه‌بندی‌شده)
4. طبقهٔ نما
5. مقدار تقریبی نما
6. تفسیر مدیریتی میانگین، میانه و نما

> هر سه شاخص را کنار هم ببینید — نه فقط یکی را.

بخش ج: شاخص‌های پراکندگی

فرض: این ۸۰ مشتری نمونه‌ای از تمام مشتریان هستند.

1. دامنهٔ تغییرات تقریبی
2. واریانس نمونه
3. انحراف معیار نمونه
4. ضریب تغییرات
5. توضیح: مبلغ خرید تا چه اندازه باثبات یا پراکنده است؟

بخش د: مقایسه با فروشگاه رقیب

رقیب گزارش کرده:

- میانگین $6 =$ میلیون تومان
- انحراف معیار $2.1 =$ میلیون تومان

برای هر دو فروشگاه:

1. CV را محاسبه کنید
2. پراکندگی مطلق را با s مقایسه کنید
3. پراکندگی نسبی را با CV مقایسه کنید
4. کدام فروشگاه ثبات نسبی بیشتری دارد؟
5. چرا مقایسهٔ s به‌تنهایی ممکن است کافی نباشد؟

بخش هـ: بررسی اثر تغییر واحد

مبلغ‌ها به‌جای «میلیون تومان» بر حسب «هزار تومان» گزارش شوند.

1. ضریب تبدیل را مشخص کنید
2. میانگین چگونه تغییر می‌کند؟
3. انحراف معیار چگونه تغییر می‌کند؟
4. واریانس چگونه تغییر می‌کند؟
5. آیا CV تغییر خواهد کرد؟

یادآوری: $Y = a + bX$ و اثر بر میانگین، واریانس و CV

بخش و: تحلیل چولگی

1. با مقایسهٔ میانگین، میانه و نما، جهت احتمالی چولگی را مشخص کنید
2. ضریب چولگی پیرسون را محاسبه کنید
3. علامت ضریب را تفسیر کنید
4. اثر خریدهای بزرگ محدود بر میانگین چیست؟
5. برای «خرید معمول»، میانگین مناسب‌تر است یا میانه؟

بخش ز: تحلیل کشیدگی و ریسک

تحلیل‌گر اعلام کرده: $\beta_2 = 4.2$

1. کشیدگی اضافی γ_2 را محاسبه کنید
2. نوع کشیدگی را مشخص کنید
3. دربارهٔ احتمال خریدهای بسیار دور از میانگین چه برداشتی ایجاد می‌کند؟
4. چرا برای برنامه‌ریزی موجودی یا سقف اعتبار مهم است؟

$\beta_2 > 3$ ← دنباله‌های سنگین‌تر از نرمال

بخش ح: تصمیم مدیریتی

پیشنهاد: مشتریان با خرید **۸ میلیون یا بیشتر** وارد سطح ویژه شوند.

1. تعداد این مشتریان را از جدول تعیین کنید
2. درصد آن‌ها از کل را محاسبه کنید
3. آیا این گروه برای برنامهٔ ویژه بیش از حد کوچک یا بزرگ است؟
4. گزارش کوتاه: سطح مرکزی · پراکندگی · شکل توزیع · پیشنهاد سطح ویژه

> استدلال را با شاخص‌های محاسبه‌شده پشتیبانی کنید.

سپاس از توجه شما