به نام خدا

**مفاهیم آماری مورد نیاز برای کار با داده و هوش مصنوعی**

**📘 آموزش فشرده مفاهیم آماری برای تحلیل داده (ویژه پایتون‌کارها)**

**✅ بخش ۱: آمار توصیفی (Descriptive Statistics)**

|  |  |  |
| --- | --- | --- |
| مفهوم | توضیح ساده | تابع در پایتون |
| میانگین (Mean) | جمع مقادیر ÷ تعداد | np.mean(data) |
| میانه (Median) | عدد وسط بعد از مرتب‌سازی | np.median(data) |
| مد (Mode) | پرتکرارترین مقدار | stats.mode(data)  از کتابخونه scipy |
| واریانس (Variance) | پراکندگی داده‌ها | np.var(data) |
| انحراف معیار (Std) | پراکندگی بر حسب واحد | np.std(data) |
| دامنه (Range) | بیشینه - کمینه | np.ptp(data) |

🎯 *کاربرد:* بررسی اولیه داده‌ها — مثلاً ببینی درآمدها به چه سمتی متمایل هستن یا چقدر پراکندگی دارن.

**✅ بخش ۲: احتمال پایه (Basic Probability)**

|  |  |  |
| --- | --- | --- |
| مفهوم | توضیح ساده | مثال |
| احتمال یک رویداد | عددی بین ۰ تا ۱ که احتمال وقوع رو نشون می‌ده | 0.75 یعنی ۷۵٪ احتمال |
| P(A|B)  احتمال شرطی | احتمال A وقتی B رخ داده | احتمال اینکه علی شاگرد اول بشه به شرط آنکه حسین قبول بشه |
| P(A ∩ B)  احتمال مشترک | احتمال وقوع همزمان دو رویداد | احتمال باران و ترافیک |
| استقلال | وقتی وقوع یکی ربطی به دیگری نداره | پرتاب دو سکه |

🎯 *کاربرد:* در مدل‌سازی، به‌ویژه در یادگیری ماشین Naive Bayes و غیره

**✅ بخش ۳: توزیع‌های آماری مهم**

|  |  |  |
| --- | --- | --- |
| توزیع | ویژگی | مثال |
| نرمال (Normal) | زنگوله‌ای، متقارن | قد انسان‌ها، وزن |
| پواسون (Poisson) | شمارش رخدادها در زمان ثابت | تعداد تماس در ۱ ساعت |
| باینومی (Binomial) | تعداد موفقیت‌ها در N بار | پرتاب سکه |

📌 در پایتون با scipy.stats می‌تونی توزیع‌ها رو رسم و تحلیل کنی.

**✅ بخش ۴: آزمون فرضیه (Hypothesis Testing)**

|  |  |  |
| --- | --- | --- |
| مفهوم | توضیح ساده | مثال |
| فرض صفر (H0) | چیزی تغییر نکرده | میانگین حقوق = ۵ میلیون |
| فرض جایگزین (H1) | چیزی تغییر کرده | میانگین حقوق ≠ ۵ میلیون |
| t-test | مقایسه میانگین‌ها | قبل و بعد از کمپین تبلیغاتی |
| p-value | احتمال دیدن این نتیجه اگر H0 درست باشه | اگر p < 0.05 ⇒ H0 رد می‌شه |

🎯 *کاربرد:* مثلاً بررسی اینکه کمپین تبلیغاتی واقعاً فروش رو بالا برده یا نه

**✅ بخش ۵: رابطه بین متغیرها**

|  |  |  |
| --- | --- | --- |
| مفهوم | کاربرد | پایتون |
| ضریب همبستگی (Correlation) | بررسی شدت رابطه بین دو عدد | data.corr() در Pandas |
| رگرسیون خطی ساده | پیش‌بینی یک متغیر بر اساس دیگری | LinearRegression از  Scikit-learn |

**📊 عنوان تمرین:**

**تحلیل آماری حقوق کارمندان یک شرکت فرضی**

**شامل موارد زیر:**

* خواندن داده با Pandas
* محاسبه میانگین، میانه، مد، واریانس و انحراف معیار
* بررسی همبستگی حقوق با سابقه کاری
* آزمون فرضیه: آیا میانگین حقوق مردان و زنان برابر است؟
* رسم نمودارهای کاربردی