

فاز سوم پروژه

تحلیل‌های آماری و تست فرضیه‌ها در داده‌های بسکتبال NBA

۱. مقدمه

در فاز سوم پروژه، هدف اصلی تحلیل آماری داده‌های ساخت‌یافته لیگ NBA بود. این تحلیل‌ها در دو بخش انجام شدند:

- بخش اول:** تحلیل‌های آماری توصیفی برای بررسی ویژگی‌های بازیکنان، تیم‌های قهرمان و برندگان جایزه Michael Jordan Trophy
 - بخش دوم:** تست فرضیه‌های کارشناسی با استفاده از روش‌های آماری برای اعتبارسنجی علمی ادعاها
- تمام داده‌ها از پایگاه داده MySQL استخراج شدند و کوئری‌های SQL مورد استفاده در این مرحله، در فاز دوم پروژه طراحی و آماده‌سازی شده بودند.

۲. ابزارها و کتابخانه‌های استفاده‌شده

برای انجام تحلیل‌ها، از ترکیبی از کتابخانه‌های قدرتمند پایتون استفاده شد:

کتابخانه	کاربرد
pandas	خواندن داده‌ها از دیتابیس، پردازش و ساخت DataFrame
numpy	محاسبات عددی و برداری
matplotlib	رسم نمودارهای هیستوگرام و باکس‌پلات
seaborn	تجسم آماری پیشرفته با ظاهر حرفه‌ای
scipy.stats	انجام آزمون‌های آماری مانند Shapiro-Wilk ، t-test و Mann-Whitney U
mysql.connector و sqlalchemy	اتصال به پایگاه داده MySQL و اجرای کوئری‌ها

۳. بخش اول: تحلیل‌های آماری توصیفی

۳.۱ مقایسه قد بازیکنان MVP با ۵ بازیکن برتر فصل

- میانگین قد MVP ها 200.30 :ساتی‌متر
 - میانگین قد ۵ بازیکن برتر فصل 198.48 :ساتی‌متر
 - اختلاف میانگین 1.82 :ساتی‌متر
- 📌 تحلیل تصویری:
- MVP ها بیشتر در بازه 201–211 ساتی‌متر قرار دارند، در حالی که بازیکنان برتر در بازه 198–206 ساتی‌متر متمرکزند.
 - باکس‌پلات‌ها نشان دادند که ربع بالایی قد MVP ها تا 211 ساتی‌متر کشیده شده، در حالی که در گروه برترها این مقدار تا 206 ساتی‌متر است.
- 📌 نتیجه: قد بلندتر می‌تواند یکی از عوامل برجستگی در انتخاب MVP باشد. این تحلیل نشان داد که فاکتور فیزیکی مثل قد، نقش مهم‌تری نسبت به صرفاً امتیازدهی یا رتبه‌بندی دارد.

۳.۲ تحلیل قد و تجربه بازیکنان تیم‌های قهرمان و ۱۵ بازیکن برتر (دو فصل آخر)

📌 تحلیل قد

- تیم‌های قهرمان: تمرکز در بازه 200–205 ساتی‌متر
- بازیکنان برتر: تمرکز در بازه‌های 185–190 و 195–200 ساتی‌متر
- میانه قد قهرمان‌ها 201 :ساتی‌متر
- میانه قد برترها 198 :ساتی‌متر
- میانگین قد قهرمان‌ها 200.43 :ساتی‌متر
- میانگین قد برترها 198.63 :ساتی‌متر

📌 نتیجه: تیم‌های قهرمان از ترکیب قد بلندتر و یکنواخت‌تری بهره می‌برند، در حالی که بازیکنان برتر پراکندگی بیشتری در قد دارند که به نقش‌های تخصصی‌تر مربوط است.

🔹 تحلیل تجربه

- تیم‌های قهرمان: تمرکز در بازه ۲ تا ۷.۵ سال
- بازیکنان برتر: تمرکز در بازه ۷ تا ۱۲.۵ سال
- میانه تجربه قهرمان‌ها: ۶ سال
- میانه تجربه برترها: ۹ سال
- میانگین تجربه قهرمان‌ها: 7.48 سال
- میانگین تجربه برترها: 10.10 سال

📌 نتیجه: تیم‌های قهرمان بیشتر از بازیکنان جوان‌تر با تجربه متمرکز استفاده می‌کنند، در حالی که بازیکنان برتر فصل به سابقه و مهارت‌های انفرادی وابسته‌اند.

۳.۳ پیشنهاد خرید بازیکن Point Guard برای باشگاه


📌 معیار انتخاب:

- تعداد حضور در لیست جایزه [Michael Jordan Trophy](#)
- کمترین سن
- بیشترین میانگین امتیاز (PTS)
- بیشترین تجربه

📌 سه گزینه پیشنهادی:

نام بازیکن	Trophy	سن	PTS	تجربه
Luka Dončić	۵	۲۶	۲۸.۶	۷ فصل

نام بازیکن	Trophy تعداد	سن	PTS	تجربه
Giannis Antetokounmpo	۵	۳۰	۲۳.۹	۱۲ فصل
Stephen Curry	۳	۳۷	۲۴.۷	۱۶ فصل

نتیجه: 

- سرمایه‌گذاری بلندمدت **Luka Dončić** :


- تجربه و فیزیک **Giannis** :

- رهبری و ثبات **Stephen Curry** :

۴. بخش دوم: تست فرضیه‌ها

۴.۱ فرضیه اول: افزایش چابکی بازیکنان برتر


 تعریف چابکی: نسبت قد به وزن

 مقایسه: بازیکنان رتبه ۱ تا ۲۰ در فصل‌های جدید (۲۲-۲۰۲۳ و ۲۳-۲۰۲۴) با قدیم (۲۰-۲۰۲۱ و ۲۱-۲۰۲۲)

 آزمون‌های آماری:

- Shapiro-Wilk Test هر دو دوره نرمال ($p > 0.05$)

- $p\text{-value} = 0.4870 \rightarrow t\text{-test}$ تفاوت معنادار وجود ندارد

 نتایج عددی:

- میانگین چابکی جدید **2.0394** :

- میانگین چابکی قدیم **2.0378** :

- اختلاف: **0.0016** (تقریباً ۰.۰۸٪)

 تحلیل تصویری:

- هیستوگرامها و باکس پلاتها نشان دهنده توزیعهای مشابه و عدم تفاوت چشمگیر

📌 نتیجه: فرضیه افزایش چابکی رد می‌شود. اختلاف عددی بسیار ناچیز و فاقد اعتبار آماری است.

۴.۲ فرضیه دوم: افزایش توانایی ذاتی بازیکنان تیمهای قهرمان

📌 تعریف توانایی ذاتی: (Skill) نسبت تجربه به سن در زمان فصل

📌 مقایسه: بازیکنان تیمهای قهرمان در فصلهای جدید (۲۰۲۲-۲۰۲۳ و ۲۰۲۳-۲۰۲۴) با قدیم (۲۰۲۰-۲۰۲۱ و ۲۰۲۱-۲۰۲۲)

📌 آزمونهای آماری:

- دوره قدیم غیرنرمال، دوره جدید نرمال → Shapiro-Wilk Test
- تفاوت معنادار وجود ندارد → $p\text{-value} = 0.2070$ Mann-Whitney U Test

📌 نتایج عددی:

- میانگین Skill جدید 0.203 :
- میانگین Skill قدیم 0.181 :
- اختلاف 0.022 :

📌 تحلیل تصویری:

- هیستوگرامها و باکس پلاتها نشان دهنده همپوشانی بالا و پراکندگی مشابه

📌 نتیجه: فرضیه افزایش توانایی ذاتی رد می‌شود. اختلاف عددی کوچک و فاقد اعتبار آماری است.

۵. نتیجه‌گیری نهایی

فاز سوم پروژه نشان داد که تحلیل‌های آماری می‌توانند دیدگاه‌های شهودی را به صورت علمی اعتبارسنجی کنند. در بخش اول، ویژگی‌های فیزیکی و تجربی بازیکنان و تیم‌ها بررسی شدند و در بخش دوم، دو فرضیه کارشناسی با روش‌های آماری تست شدند.

✅ نکات کلیدی:

- قد بلندتر در موفقیت تیمی و انتخاب MVP نقش دارد.
- تجربه بالا در عملکرد انفرادی مؤثر است، اما تیم‌های قهرمان ترکیب جوان‌تری دارند.
- هیچ‌یک از دو فرضیه تست‌شده در بخش دوم از نظر آماری تأیید نشدند، که نشان می‌دهد تحلیل‌های عددی می‌توانند از ادعاهای ذهنی جلوگیری کنند.