

Fundamentals of Statistical Hypothesis Testing

Dr. Fatemeh Jahanjoo
Assistant Professor,
TBZMED

Email: f.jahanjou@yahoo.com

آزمون فرضیه

تعریف آزمون فرضیه

- آزمون فرضیه یک روش آماری است که برای تصمیم‌گیری در مورد یک ادعا (فرضیه) بر اساس داده‌های نمونه‌ای استفاده می‌شود.
- هدف، بررسی این است که آیا شواهد کافی برای رد کردن یک فرضیه پیش‌فرض وجود دارد یا خیر.
- این فرآیند بر پایه احتمال بنا شده و به ما کمک می‌کند تا از تعمیم نتایج نمونه به جامعه اطمینان حاصل کنیم.



انواع فرضیه‌ها

- فرضیه صفر (Null Hypothesis - H_0): فرضیه پیش فرض که معمولاً بیانگر عدم وجود تفاوت، رابطه یا اثر است. مثلاً "هیچ تفاوتی بین دو گروه وجود ندارد" یا "متغیرها مستقل هستند". ما سعی می‌کنیم این فرضیه را رد کنیم.

- فرضیه مخالف (Alternative Hypothesis - H_1 یا H_a): بیانگر وجود تفاوت، رابطه یا اثر است. این فرضیه می‌تواند یک طرفه (مثلاً "گروه A بزرگ‌تر از گروه B است") یا دو طرفه (مثلاً "گروه A با گروه B متفاوت است") باشد.



مراحل آزمون فرضیه

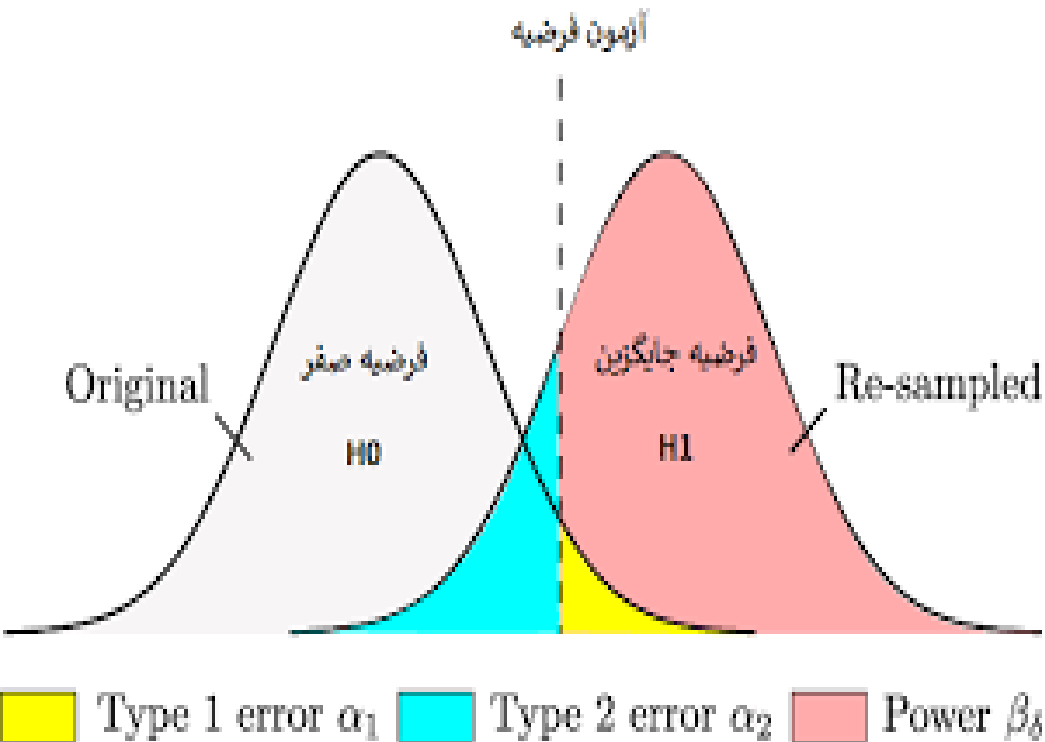
۱. بیان فرضیه‌ها: H_0 و H_1 را به طور واضح تعریف کنید.
 ۲. انتخاب آزمون آماری مناسب: بر اساس نوع داده‌ها (پیوسته، گسسته)، تعداد گروه‌ها و فرضیات توزیع (نرمال یا غیرنرمال).
 ۳. محاسبه آمار آزمون: مانند t ، F ، χ^2 بر اساس داده‌ها.
 ۴. تعیین سطح اهمیت (α – Significance Level): معمولاً ۰.۰۵ (۵٪)، که احتمال خطای نوع اول (رد اشتباه H_0) را نشان می‌دهد.
 ۵. محاسبه p -value یا مقایسه با مقدار بحرانی: اگر $p\text{-value} < \alpha$ ، H_0 را رد می‌کنیم.
 ۶. نتیجه‌گیری: بر اساس رد یا عدم رد H_0 ، نتیجه را تفسیر کنید.
- توجه: عدم رد H_0 به معنای اثبات آن نیست، فقط شواهد کافی برای رد وجود ندارد.

خطاهای آزمون فرضیه

- خطای نوع اول (Type I Error): رد اشتباه H_0 وقتی که درست است (False Positive). احتمال آن α است.

- خطای نوع دوم (Type II Error): عدم رد H_0 وقتی که نادرست است (False Negative). احتمال آن β است، و قدرت آزمون (Power) برابر $1 - \beta$ است.

- قدرت آزمون: احتمال رد H_0 وقتی که نادرست است. برای افزایش قدرت، حجم نمونه را افزایش دهید یا α را بزرگتر کنید.



فرضیات پیش نیاز (Assumptions)

- هر آزمون فرضیه فرضیاتی دارد، مانند نرمال بودن توزیع داده‌ها، همسانی واریانس‌ها (Homoscedasticity) یا استقلال مشاهدات. اگر این فرضیات نقض شوند، از آزمون‌های غیرپارامتریک (مانند Mann-Whitney به جای t-test) استفاده کنید.

Normality, Homogeneity of Variance, and Independence

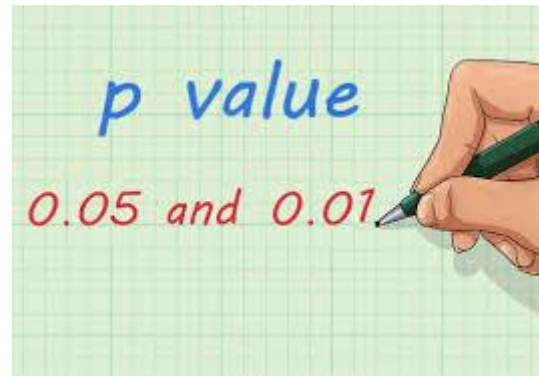
Normality



Independence

Homogeneity of Variance

p-value مفهوم



- تعریف p-value

اگر فرضیه صفر درست باشد، چقدر احتمال دارد داده‌هایی مانند آنچه مشاهده کردیم، ببینیم؟"

- تفسیر p-value

- p-value کوچک (معمولاً $0.05 >$): شواهد قوی علیه H_0 وجود دارد، پس آن را رد می‌کنیم. نتیجه "معنی‌دار آماری" (Statistically Significant) است.

- p-value بزرگ ($0.05 \leq$): شواهد کافی برای رد H_0 نیست، اما این به معنای اثبات H_0 نیست.



-نکته مهم p-value: احتمال درست بودن H_0 نیست! همچنین، اندازه اثر (Effect Size) را نشان نمی‌دهد؛ فقط اهمیت آماری را.

مثال ساده

فرض کنید H_0 : میانگین قد مردان و زنان برابر است.

اگر $p\text{-value} = 0.03$ ، یعنی اگر واقعاً تفاوتی نباشد، احتمال مشاهده چنین تفاوتی در داده‌ها تنها ۳٪ است. پس H_0 را رد می‌کنیم.

اشتباهات رایج در تفسیر p-value

- اشتباه ۱: $p\text{-value} = 0.05$ به معنای ۵٪ احتمال اشتباه بودن نتیجه نیست. (این α است، نه $p\text{-value}$)
- اشتباه ۲: $p\text{-value}$ کوچک به معنای اثر بزرگ نیست. مثلاً در نمونه‌های بزرگ، تفاوت‌های کوچک هم معنی‌دار می‌شوند.
- اشتباه ۳: $p\text{-value} > 0.05$ به معنای عدم وجود اثر نیست؛ ممکن است حجم نمونه کم باشد یا قدرت آزمون پایین.
- نکته اپیدمیولوژیک: در مطالعات اپیدمیولوژی، $p\text{-value}$ را با Confidence Interval ترکیب کنید تا اندازه اثر (مانند Odds Ratio) را ارزیابی کنید.

محدودیت‌ها و جایگزین‌ها

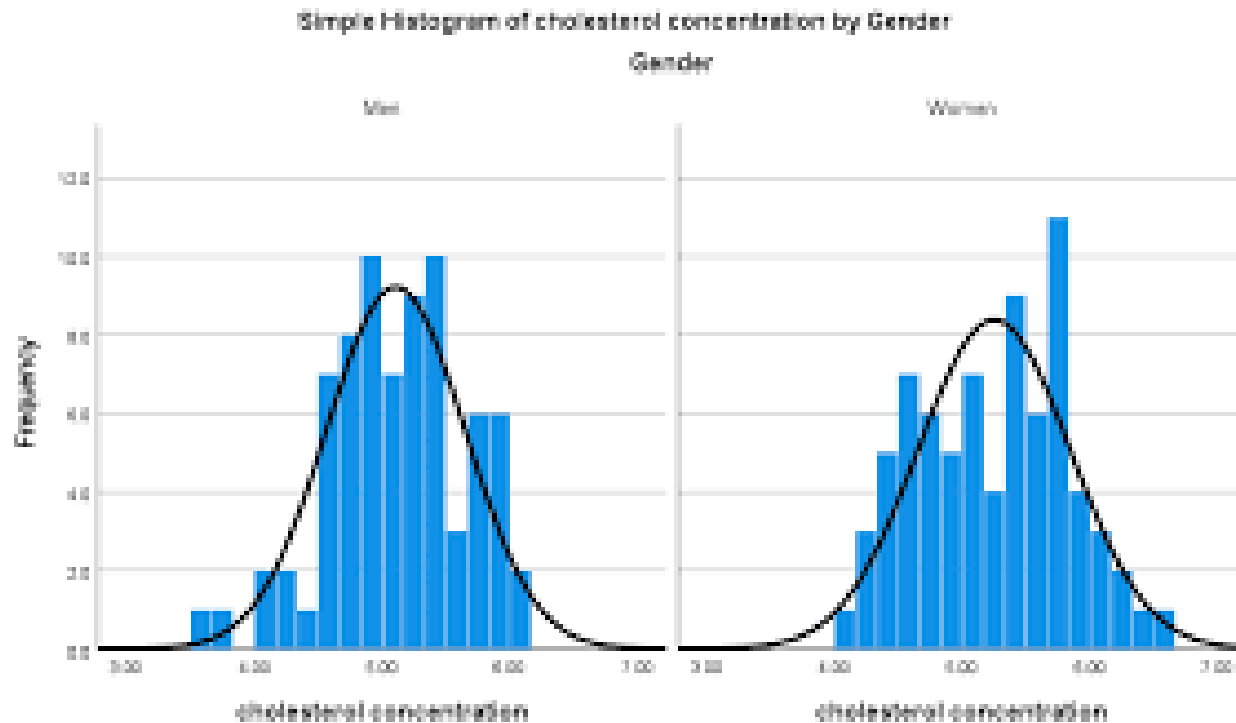
- p-value تحت تأثیر حجم نمونه است (نمونه بزرگ‌تر \rightarrow p-value کوچک‌تر).
- انتقادها: منجر به "p-hacking" (دستکاری داده‌ها برای رسیدن به $p < 0.05$) می‌شود.

- جایگزین: استفاده از Bayes Factor، Effect Size (مانند Cohen's d) یا گزارش Confidence Intervals.

فرضیه‌های آماری برای آزمون‌های خاص

آزمون‌های نرمالیتی (Normality Tests)

آزمون‌های نرمالیتی بررسی می‌کنند که آیا داده‌ها از توزیع نرمال پیروی می‌کنند یا خیر. این موضوع در انتخاب آزمون‌های پارامتریک یا غیرپارامتریک مهم است.



آزمون‌های نرمالیتی (Normality Tests)

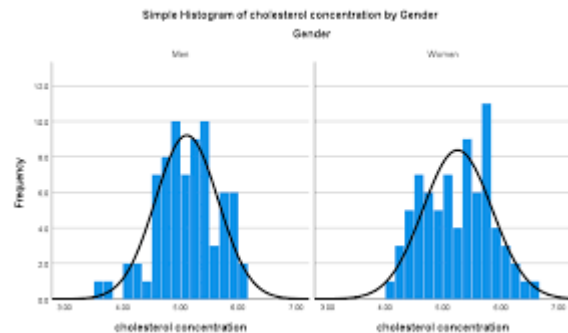
آزمون	فرضیه صفر (H_0)	فرضیه مخالف (H_1)	فرضیات پیش‌نیاز	توضیحات
Shapiro-Wilk	داده‌ها از توزیع نرمال پیروی می‌کنند	داده‌ها از توزیع نرمال پیروی نمی‌کنند	استقلال مشاهدات، داده‌ها پیوسته	حساس‌ترین آزمون برای نمونه‌های کوچک ($n < 50$)، اما به حجم نمونه حساس است
Kolmogorov-Smirnov (K-S)	داده‌ها با توزیع نرمال مطابقت دارند	داده‌ها با توزیع نرمال مطابقت ندارند	داده‌ها پیوسته، استقلال مشاهدات	برای نمونه‌های بزرگ‌تر مناسب‌تر است، اما کمتر قوی
Anderson-Darling	داده‌ها از توزیع نرمال پیروی می‌کنند	داده‌ها از توزیع نرمال پیروی نمی‌کنند	داده‌ها پیوسته، استقلال	روی دم‌های توزیع حساس‌تر است
Jarque-Bera	داده‌ها نرمال هستند (چولگی و کشیدگی صفر)	داده‌ها نرمال نیستند	داده‌ها پیوسته، استقلال، حجم نمونه بزرگ	بر اساس چولگی (Skewness) و کشیدگی (Kurtosis)

آزمون‌های نرمالیتی (Normality Tests)

p-value تفسیر: اگر $p\text{-value} < 0.05$ ، H_0 رد می‌شود و داده‌ها غیرنرمال تلقی می‌شوند.

کاربرد در اپیدمیولوژی: نرمال بودن داده‌ها پیش‌نیاز آزمون‌هایی مانند **t-test** و **ANOVA** است. اگر داده‌ها غیرنرمال باشند، از آزمون‌های غیرپارامتریک (مانند **Mann-Whitney** یا **Kruskal-Wallis**) استفاده کنید.

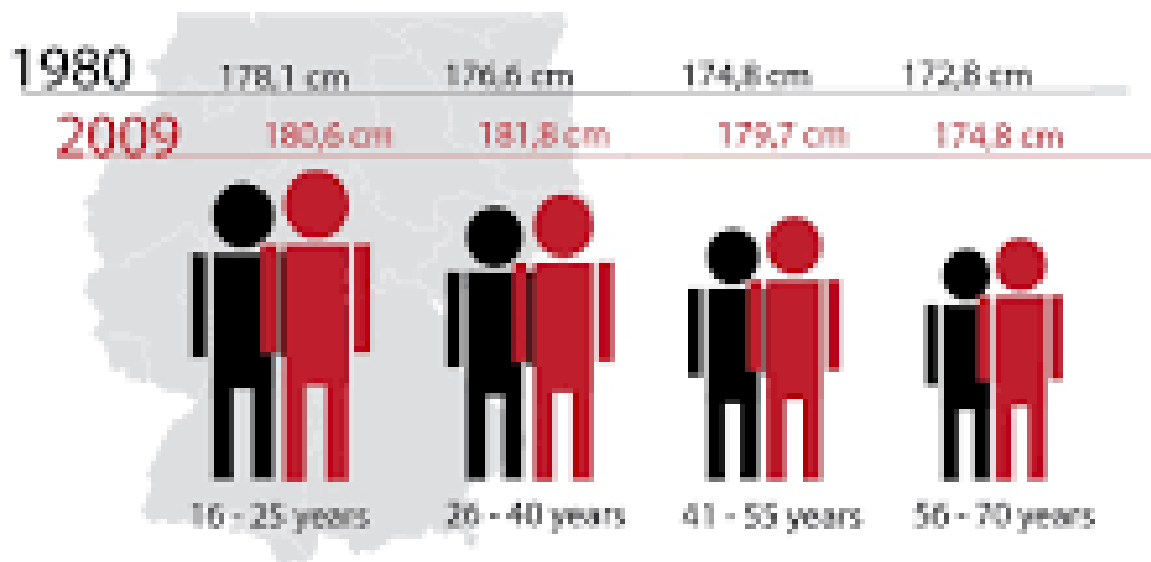
روش‌های بصری: علاوه بر آزمون‌ها، از نمودارهای **Q-Q Plot** یا هیستوگرام برای بررسی بصری نرمالیتی استفاده کنید.



آزمون‌های برابری واریانس‌ها (Homogeneity of Variance Tests)

این آزمون‌ها بررسی می‌کنند که آیا واریانس گروه‌های مختلف برابر است یا خیر. برابری واریانس‌ها (Homoscedasticity) پیش‌نیاز بسیاری از آزمون‌های پارامتریک است.

Compare Variances Between Groups



آزمون‌های برابری واریانس‌ها (Homogeneity of Variance Tests)

آزمون	فرضیه صفر (H_0)	فرضیه مخالف (H_1)	فرضیات پیش‌نیاز	توضیحات
Levene's Test	واریانس همه گروه‌ها برابر است ($\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$)	حداقل واریانس یکی از گروه‌ها متفاوت است	استقلال مشاهدات، داده‌ها پیوسته	قوی در برابر نقض نرمالیتی، برای ANOVA مناسب است
Bartlett's Test	واریانس همه گروه‌ها برابر است	حداقل واریانس یکی از گروه‌ها متفاوت است	داده‌ها نرمال، استقلال مشاهدات	به نرمالیتی حساس است، برای داده‌های نرمال دقیق‌تر است
F-test (برای دو گروه)	واریانس دو گروه برابر است ($\sigma_1^2 = \sigma_2^2$)	واریانس دو گروه برابر نیست ($\sigma_1^2 \neq \sigma_2^2$)	داده‌ها نرمال، استقلال	برای مقایسه واریانس دو گروه، حساس به غیرنرمالیتی
Brown-Forsythe	واریانس همه گروه‌ها برابر است	حداقل واریانس یکی از گروه‌ها متفاوت است	استقلال مشاهدات، داده‌ها پیوسته	نسخه مقاوم‌تر Levene، برای داده‌های غیرنرمال مناسب‌تر

آزمون‌های برابری واریانس‌ها (Homogeneity of Variance Tests)

p-value تفسیر: اگر $p\text{-value} < 0.05$ ، H_0 رد می‌شود و واریانس‌ها نابرابر (Heteroscedasticity) تلقی می‌شوند.

کاربرد در اپیدمیولوژی: برابری واریانس‌ها برای $t\text{-test}$ مستقل و ANOVA ضروری است. اگر واریانس‌ها نابرابر باشند، از نسخه‌های تعدیل‌شده آزمون‌ها (مانند Welch's t-test) استفاده کنید.

حساسیت: Bartlett به نرمالیتی حساس است، در حالی که Levene و Brown-Forsythe مقاوم‌تر هستند.

آزمون t-test

این آزمون برای مقایسه میانگین دو گروه پیوسته استفاده می‌شود. انواع: مستقل (Independent)، جفتی (Paired)، و تک‌نمونه‌ای (One-sample).

نوع آزمون	فرضیه صفر (H_0)	فرضیه مخالف (H_1) - دوطرفه	فرضیه مخالف (H_1) - یک‌طرفه	فرضیات پیش‌نیاز
t-test تک‌نمونه‌ای (مقایسه میانگین نمونه با مقدار ثابت μ_0)	$\mu = \mu_0$ (میانگین جامعه برابر μ_0 است)	$\mu \neq \mu_0$	$\mu > \mu_0$ یا $\mu < \mu_0$	داده‌ها نرمال، استقلال مشاهدات
t-test مستقل (مقایسه دو گروه مستقل)	$\mu_1 = \mu_2$ (میانگین دو گروه برابر است)	$\mu_1 \neq \mu_2$	$\mu_1 > \mu_2$ یا $\mu_1 < \mu_2$	نرمال بودن، هم‌سانی واریانس‌ها (Homoscedasticity)، استقلال
t-test جفتی (مقایسه دو گروه وابسته، مانند قبل/بعد)	$\mu_d = 0$ (میانگین تفاوت‌ها صفر است)	$\mu_d \neq 0$	$\mu_d > 0$ یا $\mu_d < 0$	تفاوت‌ها نرمال، جفت بودن مشاهدات

آنالیز واریانس (ANOVA)

برای مقایسه میانگین بیش از دو گروه پیوسته. انواع: یک طرفه (One-way)، دو طرفه (Two-way). پس از ANOVA معنی دار، از آزمون‌های پس‌تستی مانند Tukey برای

شناسایی تفاوت‌ها استفاده کنید.

نوع آزمون	فرضیه صفر (H_0)	فرضیه مخالف (H_1)	فرضیات پیش‌نیاز
One-way ANOVA (مقایسه k گروه)	$\mu_1 = \mu_2 = \dots = \mu_k$ (میانگین همه گروه‌ها برابر است)	حداقل دو میانگین متفاوت هستند	نرمال بودن، همسانی واریانس‌ها، استقلال
Two-way ANOVA (با دو عامل)	برای عامل A: هیچ اثری ندارد برای عامل B: هیچ اثری ندارد برای تعامل: هیچ تعاملی ندارد	حداقل یکی از عوامل یا تعامل اثر دارد	همانند One-way، به علاوه عدم هم‌خطی عوامل

پس از ANOVA معنی دار، از آزمون‌های پس‌تستی مانند Tukey برای شناسایی تفاوت‌ها استفاده کنید.

آزمون کای دو (χ^2 - Chi-square)

برای داده‌های گسسته (دسته‌ای). انواع: استقلال (Independence) و -goodness-of-fit.

نوع آزمون	فرضیه صفر (H_0)	فرضیه مخالف (H_1)	فرضیات پیش‌نیاز
Chi-square Goodness-of-Fit (تطابق) توزیع مشاهده‌شده با مورد انتظار	توزیع مشاهده‌شده با انتظار برابر است ($O = E$ برای همه دسته‌ها)	حداقل یک دسته متفاوت است	داده‌ها گسسته، حداقل ۵ مورد انتظار در هر سلول، استقلال
Chi-square Test of Independence (بررسی استقلال دو متغیر دسته‌ای)	دو متغیر مستقل هستند (هیچ ارتباطی ندارند)	دو متغیر وابسته هستند (ارتباط وجود دارد)	همانند Goodness-of-Fit، جدول احتمالات (Contingency Table)

در اپیدمیولوژی، Chi-square برای بررسی ارتباط بین عوامل خطر (مانند جنسیت و بیماری) مفید است. اگر سلول‌ها کم باشند، از Fisher's Exact Test استفاده کنید.



For further Questions:

f.jahanjou@yahoo.com

Thank you

