
تحلیل واریانس و آزمون‌های تعقیبی در پژوهش‌های پزشکی

راهنمای کاربردی One-Way ANOVA، انتخاب آزمون Post Hoc
و گزارش نتایج در مقاله پزشکی

این متن شامل مثال‌های آموزشی کامل برای آزمون‌های تعقیبی Games-Howell، Tukey، Bonferroni و Dunnett است. تمام داده‌ها ساختگی اند و فقط برای آموزش روش آماری طراحی شده‌اند.

نسخه آموزشی - مناسب آمار زیستی و نگارش مقاله پزشکی

فهرست مطالب

1	تعریف و جایگاه تحلیل واریانس در پزشکی	1
1	1.1 چه زمانی ANOVA مناسب نیست؟	1
2	2 پیش‌فرض‌های مورد نیاز	2
2	3 مثال پایه: مقایسه چهار گروه درمانی	2
2	3.1 سناریوی پزشکی	2
3	3.2 جدول تحلیل واریانس	3
3	4 حالت اول: آزمون تعقیبی Tukey	3
3	4.1 چه زمانی انتخاب می‌شود؟	3
4	4.2 خروجی Tukey	4
4	5 حالت دوم: Games-Howell در واریانس‌های نابرابر	4
4	5.1 سناریوی پزشکی	4
5	6 حالت سوم: Dunnett برای مقایسه با گروه کنترل	5
6	7 حالت چهارم: اصلاح Bonferroni برای مقایسه‌های محدود	6
7	8 راهنمای انتخاب سریع آزمون تعقیبی	7
7	9 تفسیر پزشکی: معناداری آماری در برابر اهمیت بالینی	7
8	10 الگوی نهایی گزارش در بخش نتایج مقاله	8
8	11 چک‌لیست کوتاه پیش از انتشار	8

1 تعریف و جایگاه تحلیل واریانس در پزشکی

تحلیل واریانس یک طرفه یا One-Way ANOVA برای مقایسه میانگین یک پیامد کمی در سه گروه مستقل یا بیشتر استفاده می‌شود. برای مثال، پژوهشگر می‌تواند میانگین کاهش فشار خون سیستمیک را بین گروه کنترل و سه گروه دارویی مقایسه کند.

اگر k گروه داشته باشیم، فرض صفر و فرض مقابل به صورت زیر تعریف می‌شوند:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k, \quad (1)$$

$$H_1 : \text{حداقل یکی از میانگین‌ها با بقیه متفاوت است.} \quad (2)$$

آماره آزمون از نسبت تغییرات بین گروه‌ها به تغییرات درون گروه‌ها ساخته می‌شود:

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}}. \quad (3)$$

نکته اصلی

معنادار شدن آزمون ANOVA فقط نشان می‌دهد حداقل یک تفاوت میانگین وجود دارد؛ اما مشخص نمی‌کند کدام گروه‌ها با یکدیگر متفاوت‌اند. برای پاسخ به این سؤال باید مقایسه‌های ازپیش‌تعیین‌شده یا آزمون‌های تعقیبی انجام شوند.

1.1 چه زمانی ANOVA مناسب نیست؟

جدول 1: انتخاب آزمون بر اساس نوع پیامد و طراحی مطالعه

وضعیت پژوهش	مثال پزشکی	روش مناسب
سه گروه مستقل، پیامد کمی	فشار خون در سه روش درمان	One-Way ANOVA
واریانس‌های نابرابر	کراتینین با پراکندگی بسیار متفاوت گروه‌ها	Welch ANOVA
اندازه‌گیری چندباره از یک بیمار	فشار خون قبل، یک ماه و سه ماه بعد	Repeated Measures ANOVA یا مدل آمیخته
پیامد رتبه‌ای یا غیرنرمال شدید	امتیاز رتبه‌ای شدت علائم	Kruskal-Wallis
پیامد دوحالتی	وقوع یا عدم وقوع عارضه	رگرسیون لجستیک یا آزمون کای‌دو
زمان تا رویداد همراه با سانسور	زمان تا شکست پیوند	Kaplan-Meier، Log-rank یا رگرسیون Cox

2 پیش‌فرض‌های مورد نیاز

1. استقلال مشاهدات: هر بیمار باید فقط به یک گروه مستقل تعلق داشته باشد، مگر اینکه طراحی مطالعه از نوع اندازه‌گیری تکراری باشد.
2. کمی بودن پیامد: متغیر وابسته باید کمی باشد؛ مانند فشار خون، هموگلوبین، کراتینین یا مدت بستری.
3. نرمال بودن تقریبی باقیمانده‌ها: این موضوع بهتر است با نمودار Q-Q و بررسی باقیمانده‌ها ارزیابی شود؛ اتکای صرف به آزمون شاپیرو-ویلک توصیه نمی‌شود.
4. همگنی واریانس‌ها: پراکندگی گروه‌ها باید تقریباً برابر باشد. آزمون Levene یکی از روش‌های متداول بررسی این فرض است.

تفسیر صحیح آزمون Levene

اگر $p > 0.05$ باشد، شواهد کافی برای نابرابری واریانس‌ها مشاهده نشده است. این نتیجه، برابری قطعی واریانس‌ها را اثبات نمی‌کند. علاوه بر مقدار p ، باید انحراف معیار گروه‌ها و نمودار داده‌ها نیز بررسی شود.

3 مثال پایه: مقایسه چهار گروه درمانی

3.1 سناریوی پزشکی

در یک مطالعه آموزشی، 60 بیمار به چهار گروه مساوی تخصیص یافته‌اند: کنترل، داروی A، داروی B و داروی C. پیامد اصلی، میزان کاهش فشار خون سیستولیک پس از هشت هفته، بر حسب میلی‌متر جیوه است. مقدار بزرگ‌تر نشان‌دهنده کاهش بیشتر است.

جدول 2: آمار توصیفی کاهش فشار خون سیستولیک

گروه	تعداد	میانگین	انحراف معیار
کنترل	15	0.800	1.474
داروی A	15	5.800	1.146
داروی B	15	3.267	1.580
داروی C	15	2.267	1.580

آزمون لون معنادار نبود:

$$F(3, 56) = 1.198, \quad p = 0.319.$$

بنابراین در این مثال شواهدی علیه فرض همگنی واریانس‌ها وجود ندارد و استفاده از ANOVA معمولی و آزمون تعقیبی Tukey قابل قبول است.

3.2 جدول تحلیل واریانس

جدول 3: جدول تحلیل واریانس یک طرفه

منبع تغییرات	مجموع مربعات	درجه آزادی	میانگین مربعات	F	p
بین گروه‌ها	199.267	3	66.422	31.345	< 0.001
درون گروه‌ها	118.667	56	2.119	—	—
کل	317.933	59	—	—	—

اندازه اثر اتا اسکوتر برابر است با:

$$\eta^2 = \frac{SS_{\text{Between}}}{SS_{\text{Total}}} = \frac{199.267}{317.933} = 0.627. \quad (4)$$

همچنین برآورد کم‌سوگیرتر امگا اسکوتر برابر $\omega^2 = 0.603$ است.

نتیجه آزمون کلی

میانگین کاهش فشار خون سیستمیک بین چهار گروه تفاوت آماری معناداری داشت، $F(3, 56) = 31.345$ ، $p < 0.001$ ، $\eta^2 = 0.627$. با این حال، برای تعیین محل تفاوت‌ها باید آزمون تعقیبی انجام شود.

4 حالت اول: آزمون تعقیبی Tukey

4.1 چه زمانی انتخاب می‌شود؟

آزمون Tukey HSD زمانی مناسب است که:

- هدف، مقایسه همه زوج‌های ممکن گروه‌ها باشد؛
 - واریانس گروه‌ها تقریباً برابر باشد؛
 - مقایسه‌ها پس از مشاهده نتیجه آزمون کلی انجام شوند؛
 - پژوهشگر بخواهد خطای نوع اول خانوادگی را در سطح تعیین شده کنترل کند.
- در مثال حاضر چهار گروه وجود دارد؛ بنابراین تعداد مقایسه‌های دوتایی برابر است با:

$$\binom{4}{2} = \frac{4(4-1)}{2} = 6. \quad (5)$$

4.2 خروجی Tukey

جدول 4: مقایسه‌های زوجی با آزمون تعقیبی Tukey

مقایسه	اختلاف میانگین	حد پایین 95%	حد بالا 95%	p تعدیل شده	نتیجه
A منهای کنترل	5.000	3.593	6.407	< 0.001	معنادار
B منهای کنترل	2.467	1.059	3.874	< 0.001	معنادار
C منهای کنترل	1.467	0.059	2.874	0.038	معنادار
A منهای B	2.533	1.126	3.941	< 0.001	معنادار
A منهای C	3.533	2.126	4.941	< 0.001	معنادار
B منهای C	1.000	-0.407	2.407	0.248	غیرمعنادار

برای مقایسه B و C، فاصله اطمینان شامل صفر است و مقدار $p = 0.248$ به دست آمده است؛ بنابراین شواهد کافی برای تفاوت میانگین این دو گروه وجود ندارد.

متن آماده برای مقاله - Tukey

تحلیل واریانس یک‌طرفه تفاوت معناداری را بین گروه‌ها نشان داد ($F(3, 56) = 31.345, p < 0.001$). آزمون تعقیبی Tukey نشان داد کاهش فشار خون در گروه داروی A به‌طور معناداری بیشتر از گروه کنترل بود (اختلاف میانگین $= 5.00$ ، فاصله اطمینان 95%: 3.59 تا 6.41، $p < 0.001$). بااین‌حال، تفاوت بین داروهای B و C معنادار نبود (اختلاف میانگین $= 1.00$ ، فاصله اطمینان 95%: -0.41 تا 2.41، $p = 0.248$).

5 حالت دوم: Games-Howell در واریانس‌های نابرابر

5.1 سناریوی پزشکی

در یک مثال مستقل، کاهش کلسترول LDL پس از 12 هفته در سه گروه بررسی شده است. حجم نمونه و پراکندگی گروه‌ها برابر نیست:

جدول 5: آمار توصیفی مثال واریانس‌های نابرابر

گروه	تعداد	میانگین کاهش LDL	انحراف معیار
درمان استاندارد	24	8.071	1.510
داروی کم‌دوز	15	12.060	4.084
داروی پردوز	10	13.800	7.230

آزمون لون معنادار است:

$$F(2, 46) = 18.276, \quad p < 0.001.$$

در نتیجه، فرض همگنی واریانس‌ها قابل دفاع نیست. به جای ANOVA معمولی از Welch ANOVA استفاده می‌شود:

$$F_{\text{Welch}}(2, 15.811) = 8.972, \quad p = 0.002.$$

پس از معنادار شدن آزمون ولچ، مقایسه‌های زوجی با روش Games–Howell انجام می‌شوند؛ زیرا این روش به برابری واریانس‌ها و برابر بودن حجم نمونه‌ها نیاز ندارد.

جدول 6: مقایسه‌های زوجی Games–Howell

نتیجه	p تعدیل شده	حد بالا 95%	حد پایین 95%	اختلاف میانگین	مقایسه
معنادار	0.006	6.817	1.161	3.989	کم‌دوز منهای استاندارد
غیرمعنادار	0.079	12.129	-0.671	5.729	پردوز منهای استاندارد
غیرمعنادار	0.773	8.397	-4.917	1.740	پردوز منهای کم‌دوز

اگرچه میانگین گروه پردوز از گروه استاندارد بیشتر است، پراکندگی زیاد و حجم نمونه کم باعث شده است فاصله اطمینان وسیع باشد و صفر را دربرگیرد. بنابراین بزرگ‌تر بودن عددی میانگین به‌تنهایی برای نتیجه‌گیری کافی نیست.

متن آماده برای مقاله - Games–Howell

به دلیل نقض فرض همگنی واریانس‌ها در آزمون لون ($p < 0.001$)، از تحلیل واریانس ولچ استفاده شد. نتیجه آزمون ولچ معنادار بود ($F(2, 15.811) = 8.972$, $p = 0.002$). مقایسه‌های Games–Howell نشان داد کاهش LDL در گروه کم‌دوز نسبت به درمان استاندارد بیشتر بود (اختلاف میانگین 3.99 =، فاصله اطمینان 95%: 1.16 تا 6.82، $p = 0.006$)، اما تفاوت گروه پردوز با درمان استاندارد معنادار نبود ($p = 0.079$).

6 حالت سوم: Dunnett برای مقایسه با گروه کنترل

آزمون Dunnett زمانی استفاده می‌شود که هدف اصلی مقایسه هر درمان با یک گروه کنترل مشترک باشد و مقایسه مستقیم تمام درمان‌ها با یکدیگر ضرورت نداشته باشد. در چنین طراحی‌ای، Dunnett معمولاً از Tukey یا اجرای چند آزمون مستقل کارآمدتر است؛ زیرا فقط خانواده مقایسه‌های مرتبط با کنترل را تعدیل می‌کند.

در مثال کاهش فشار خون، سه دارو با گروه کنترل مقایسه می‌شوند:

جدول 7: مقایسه درمان‌ها با کنترل با استفاده از آزمون Dunnett

مقایسه	اختلاف میانگین	حد پایین 95%	حد بالا 95%	p تعدیل شده	نتیجه
A منهای کنترل	5.000	3.717	6.283	< 0.001	معنادار
B منهای کنترل	2.467	1.183	3.750	< 0.001	معنادار
C منهای کنترل	1.467	0.183	2.750	0.021	معنادار

متن آماده برای مقاله - Dunnett

در مقایسه‌های تعدیل شده Dunnett، هر سه دارو نسبت به گروه کنترل کاهش بیشتری در فشار خون ایجاد کردند. اختلاف میانگین داروی A با کنترل برابر 5 میلی‌متر جیوه بود (فاصله اطمینان 95%: 3.72 تا 6.28، $p < 0.001$).

7 حالت چهارم: اصلاح Bonferroni برای مقایسه‌های محدود

روش Bonferroni برای تعداد محدودی مقایسه از پیش تعیین شده ساده و قابل دفاع است. اگر m مقایسه انجام شود، مقدار احتمال تعدیل شده از رابطه زیر به دست می‌آید:

$$p_{\text{adjusted}} = \min(mp_{\text{raw}}, 1). \quad (6)$$

روش معادل دیگر این است که سطح معناداری هر مقایسه برابر α/m در نظر گرفته شود. در مثال زیر، چهار مقایسه از قبل در پروتکل مشخص شده‌اند؛ بنابراین:

$$\alpha_{\text{per comparison}} = \frac{0.05}{4} = 0.0125.$$

جدول 8: مقایسه‌های برنامه‌ریزی شده با اصلاح Bonferroni

مقایسه	اختلاف میانگین	t	p خام	p تعدیل شده	نتیجه نهایی
A منهای کنترل	5.000	10.372	< 0.001	< 0.001	معنادار
B منهای کنترل	2.467	4.422	< 0.001	< 0.001	معنادار
C منهای کنترل	1.467	2.630	0.014	0.055	غیرمعنادار
A منهای B	2.533	5.027	< 0.001	< 0.001	معنادار

مقایسه داروی C با کنترل در نگاه اول دارای p خام برابر 0.014 است؛ اما پس از اصلاح چهار مقایسه، $p_{\text{adjusted}} = 0.055$ می‌شود. بنابراین این مقایسه در سطح خطای خانوادگی 5 درصد معنادار نیست.

خطای رایج

نباید مقدار p خام را گزارش کرد و سپس ادعا کرد خطای چندگانه کنترل شده است. در جدول اصلی باید مقدار p تعدیل‌شده، روش تعدیل و تعداد مقایسه‌ها مشخص باشند.

متن آماده برای مقاله - Bonferroni

چهار مقایسه ازپیش‌تعیین‌شده با اصلاح Bonferroni انجام شد. پس از تعدیل، تفاوت داروی C با کنترل از نظر آماری معنادار نبود (اختلاف میانگین $= 1.47$ ، $p_{\text{adjusted}} = 0.055$)، درحالی‌که تفاوت داروهای A و B با کنترل معنادار باقی ماند ($p_{\text{adjusted}} < 0.001$).

8 راهنمای انتخاب سریع آزمون تعقیبی

جدول 9: انتخاب روش مقایسه پس از آزمون کلی

روش	سؤال پژوهشی و شرایط	نمونه کاربرد پزشکی
Tukey HSD	همه زوج‌های گروهی مهم‌اند و همگنی واریانس‌ها قابل قبول است.	مقایسه همه رژیم‌های درمانی با یکدیگر
Games-Howell	همه زوج‌ها مهم‌اند، اما واریانس‌ها یا حجم نمونه‌ها نابرابرند.	مقایسه درمان‌ها با پراکندگی بسیار متفاوت آزمایشگاهی
Dunnett	هر درمان فقط با یک کنترل مشترک مقایسه می‌شود.	چند دوز دارو در برابر دارونما
Bonferroni	تعداد کمی مقایسه از قبل در پروتکل تعیین شده است.	سه مقایسه اصلی ازپیش‌تعریف‌شده

قاعده عملی

انتخاب آزمون تعقیبی باید بر اساس سؤال ازپیش‌تعریف‌شده پژوهش و وضعیت پیش‌فرض‌ها باشد، نه بر اساس اینکه کدام روش مقدار p کوچک‌تری تولید می‌کند.

9 تفسیر پزشکی: معناداری آماری در برابر اهمیت بالینی

مقدار p فقط میزان ناسازگاری داده‌ها با فرض صفر را نشان می‌دهد و اندازه یا اهمیت بالینی اثر را مشخص نمی‌کند. برای مثال، کاهش متوسط 1.5 میلی‌متر جیوه ممکن است از نظر آماری معنادار باشد، ولی اهمیت بالینی آن باید بر اساس شواهد تخصصی و آستانه تفاوت بالینی مهم تعیین شود. برای گزارش کامل، موارد زیر ارائه شوند:

- تعداد، میانگین و انحراف معیار هر گروه؛
- آماره آزمون کلی، درجات آزادی و مقدار p ؛
- اندازه اثر مانند η^2 یا ω^2 ؛
- اختلاف میانگین هر مقایسه همراه با فاصله اطمینان 95 درصد؛
- مقدار p تعدیل‌شده و نام روش تعدیل؛
- تفسیر جداگانه اهمیت بالینی و آماری.

10 الگوی نهایی گزارش در بخش نتایج مقاله

الگوی قابل ویرایش

میانگین [نام پیامد و واحد اندازه‌گیری] در گروه‌های [نام گروه‌ها] با استفاده از [Welch/One-Way ANOVA] مقایسه شد. [فرض همگنی واریانس‌ها بر اساس آزمون لون قابل قبول بود/فرض همگنی واریانس‌ها نقض شد]. نتیجه آزمون کلی نشان داد تفاوت بین گروه‌ها [معنادار/غیرمعنادار] بود ($F(df_1, df_2) = [F], p = [p]$). مقایسه‌های [نام آزمون تعقیبی] نشان داد اختلاف میانگین گروه [الف] و [ب] برابر [مقدار با واحد] بود (فاصله اطمینان 95%: [حد پایین] تا [حد بالا]، $p_{adjusted} = [p]$). این اختلاف از نظر [بالینی/آماري] [قابل توجه/غیرقابل توجه] ارزیابی شد.

11 چک‌لیست کوتاه پیش از انتشار

1. آیا پیامد کمی و گروه‌ها مستقل‌اند؟
2. آیا نمودار داده‌ها، نقاط پرت و باقیمانده‌ها بررسی شده‌اند؟
3. آیا برابری واریانس‌ها ارزیابی شده است؟
4. اگر واریانس‌ها نابرابرند، آیا از ولج و Games-Howell استفاده شده است؟
5. آیا آزمون کلی پیش از مقایسه‌های پسینی گزارش شده است؟
6. آیا روش کنترل خطای چندگانه مشخص شده است؟
7. آیا اختلاف میانگین و فاصله اطمینان 95 درصد گزارش شده‌اند؟
8. آیا اندازه اثر و اهمیت بالینی جدا از معناداری آماری تفسیر شده‌اند؟

تمام اعداد این سند متعلق به مثال‌های آموزشی ساختگی هستند و نباید به‌عنوان نتیجه یک مطالعه واقعی پزشکی نقل شوند.