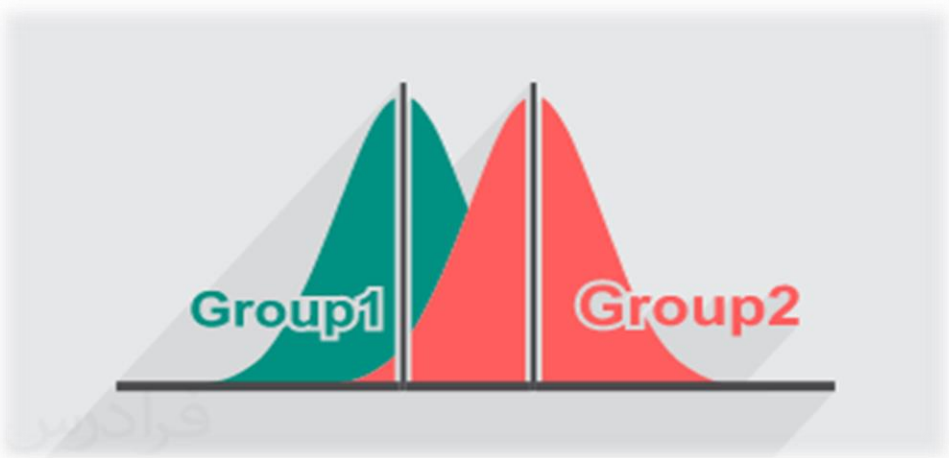
بسمه تعالی

موضوع

تحلیل واریانس



گردآورندگان

سارا معصومی ، محدثه حیدری ، زهرا رسولی

استاد راهنما

سرکار خانم زهرا صمدی

دی ۱۴۰۰-دانشگاه قم

فهرست

فصل اول

فصل دوم

فصل سوم : انجام پروژه

پیوست

چکیده

آنالیزواریانس به محقق این امکان را می‌دهد که بتواند میانگین چند جامعه را به طور همزمان با یکدیگر مقایسه کند و همچنین بداند میانگین جوامع مورد نظر با یکدیگر برابر هستند یا خیر.

آنالیزواریانس در بسیاری از علوم مانند: پزشکی، اقتصاد، کشاورزی و ... کاربرد دارد .

همچنین بسته به تعداد عامل‌های اثرگذار شناسایی شده و سطوح هر عامل دارای انواع مختلفی است. برای مثال بسته به تعداد عامل‌ها، آنالیزواریانس یک‌راهه، دوراهه، یا چندراهه وجود دارد. و یا بسته به تعداد سطوح موجود در هر جامعه می‌توان از آنالیزواریانس با اثرات ثابت، تصادفی و یا آمیخته استفاده نمود.

ما در این مقاله قصد داریم به مباحث ذکر شده در رابطه با آنالیزواریانس به‌طور کامل بپردازیم و در انتها مثالی کاربردی را با موضوع بررسی تاثیر کودهای مختلف(عامل اول) و انواع بذرگندم(عامل دوم) بر روی طول خوشه‌های گندم(متغیر پاسخ)شرح خواهیم‌داد.

طرح ما یک طرح متعادل است چرا که تعداد مشاهدات در هر سطح با یکدیگر برابر است.

نکته‌ی بسیار مهم بررسی مفروضات طرح آنالیزواریانس می‌باشد که باید مدل نهایی فرضیات طرح آنالیزواریانس را دارا باشد تا نتایج به‌دست‌آمده از اعتبار خوبی برخوردار باشند. البته که آنالیزواریانس در برابر عدم برقراری فرضیات، مقاوم است. اما در ادامه به بیان فرضیات و بررسی آن‌ها نیز خواهیم پرداخت.

مروری بر تاریخچه آنالیز واریانس

تحلیل واریانس به انگلیسی Analysis of variance و به اختصار ANOVA میباشد. این روش توسط رونالد فیشر [1] زیست شناس و آمارشناس مشهور، ابداع شده است. سال ۱۹۱۹ در پی پیشنهاد هم‌زمان دو پُست، فیشر از شغل معلمی دست کشید.«کارل پیرسون» به او پیشنهاد کرد تا به عنوان کارشناس ارشد آمار در لابراتوار گالتون مشغول به کار شود. همچنین پست مشابهی در ایستگاه آزمایش‌های کشاورزی «روتهامستد» [2] به وی پیشنهاد شد که یکی از قدیمی‌ترین مؤسسات تحقیقات کشاورزی در انگلستان بود و در سال ۱۸۳۷ به منظور مطالعه بر روی اثرات تغذیه‌ی خاک و انواع تیپ خاک بر روی باروری گیاهان تأسیس شده بود. علاقه فیشر به کشاورزی سبب شد تا پست پیشنهادی «روتهامستد» را بپذیرد. جایی که او به واسطه ارائه روش‌های آنالیز و تجزیه و تحلیل نتایج آزمایش‌ها، خدمات زیادی هم به علم آمار و هم به علم ژنتیک کرد.

در آنجا بود که او بر طراحی آزمایش‌هایی بوسیله معرفی مفهوم انتخاب تصادفی و آنالیزواریانس مطالعاتی انجام داد، روش‌هایی که هم‌اکنون نیز در تمام دنیا مورد استفاده قرار می‌گیرند. فیشر عقیده داشت که چیدمان و تنظیم یک آزمون به صورت مجموعه‌ای از زیرآزمایش‌های مجزا که در داشتن یک یا چندین فاکتور یا رفتار، با یکدیگر متفاوت هستند برای آن‌ها کارایی بیشتری دارد. این زیر آزمایش‌ها در چنین روشی به گونه‌ای طراحی می‌شدند که تفاوت خروجی‌های آن‌ها به فاکتورها یا ترکیبی از فاکتورها به وسیله آنالیز آماری ربط داده شود. این پیشرفت قابل توجهی نسبت به روش‌های موجود بود که تنها یک فاکتور را در یک زمان در یک آزمایش بررسی می‌کرد، که روش ناکارآمدی بود.

او در کتاب معروف خود به نام «روش‌های آماری برای محققین» [3] به بررسی و شیوه تفکیک واریانس پرداخت و به کمک آن بسیاری از آزمون‌های فرض آماری را تشکیل داد. اساس همه این روش‌ها، تفکیک واریانس یا پراکندگی داده‌ها به چند جزء بود.

---

[1] R. A. Fisher

[2] Rothamsted

[3] Statistical Methods for Research Workers

تعریف آنالیز واریانس

ANOVA یک آزمون آماری برای تعیین تفاوت میانگین‌های دو یا چند جامعه آماری مستقل است. به عبارت دیگر، ANOVA آزمون آماری ارائه می‌کند که به وسیله آن می‌توان برابر بودن یا نبودن میانگین‌های بین دو یا تعداد بیشتری از جوامع را تعیین کرد.

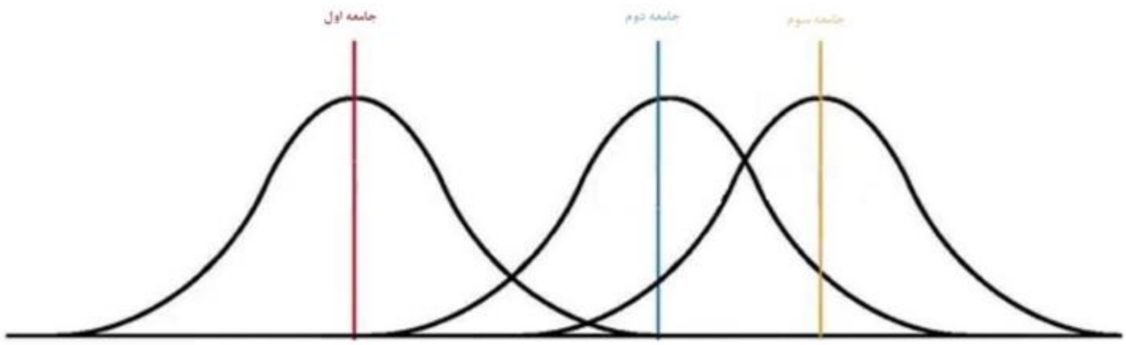
در حقیقت با استفاده از تحلیل واریانس می‌توان بررسی کرد که آیا میانگین متغیر پاسخ در سطوح عامل مورد نظر دارای تفاوت هستند یا خیر. اگر بخواهیم تاثیر سطوح متغیر مستقل که یک متغیر کیفی است را بر متغیر وابسته که کمی است بررسی کنیم از تحلیل واریانس استفاده می‌کنیم.

مهم‌ترین کاربرد آنالیزواریانس در شناخت روابط موجود بین عوامل مختلف است. شناخت روابط موجود بین عوامل مختلف یکی از مباحثی است که می توان گفت در تمامی علوم از قبیل مدیریت، روانشناسی، زیست شناسی، مهندسی، اقتصاد و..... بسیار مورد نیاز است و کاربردهای فراوانی دارد.

ANOVA تعمیمی از آزمون T-Student است

از آزمون T-Student برای مقایسه دو گروه استفاده می‌شود، در حالی که در آزمون ANOVA برای مقایسه ۳ گروه یا بیشتر، کاربرد دارد.

برای نشان دادن این موضوع سه جامعه را در نظر بگیرید:



با توجه به ترکیب‌های دوتایی سه میانگین این جوامع، آزمون‌های فرض به صورت زیر درخواهند آمد:

$$\begin{cases} H_\circ: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

$$\begin{cases} H_\circ: \mu_2 = \mu_3 \\ H_1: \mu_1 \neq \mu_3 \end{cases}$$

$$\begin{cases} H_\circ: \mu_2 = \mu_3 \\ H_1: \mu_2 \neq \mu_3 \end{cases}$$

خطای نوع اول برای هر یک از آزمون‌ها را به صورت زیر محاسبه کنیم اگر A_i را پیشامد «عدم رد فرض صفر آزمون i ام با توجه به درست بودن آن» در نظر بگیریم، می‌توان احتمال

$$\alpha_i = 1 - P(A_i)$$

حال برای محاسبه احتمال خطای نوع اول همه این آزمون‌ها به طور همزمان، باید احتمال اینکه هیچ یک از Ai ها رخ ندهد را بدست آوریم. بنابراین اگر ∩Ai را پیشامد رخداد همه آن‌ها در نظر بگیریم، کافی است احتمال متمم آن‌ها را محاسبه کنیم.

$$\alpha_t = 1 - P(\cap A_i) = 1 - \prod_{i=1}^{k} p(A_i) = 1 - \prod_{i=1}^{k} (1-\alpha)$$

در این حالت اگر احتمال خطای نوع اول را برای همه آزمون‌ها یکسان و برابر با α در نظر بگیریم، رابطه بالا ساده‌تر شده و به صورت زیر درخواهدآمد:

$$\alpha_t = 1 - \prod_{I=1}^{K} (1-\alpha) = 1 - (1-\alpha)K$$

در نتیجه اگر خطای نوع اول برای هریک از آزمون‌هابرای مثال α=0.05 باشد، خطای انجام آزمون همزمان آن‌ها در صورت استفاده از ترکیب‌های دوتایی و انجام آزمون T برابر است با:

$$\alpha_t = 1-(1-\alpha)\ k = 1-(1-0.05)3 \approx 0.1$$

تحلیل واریانس، روش مناسبی برای ارزیابی برابری چندین میانگین محسوب می‌شود. با این وجود، کاربرد تحلیل واریانس به مراتب وسیع‌تر از مثال فوق است. شاید به جرأت بتوان از تحلیل واریانس به عنوان مفیدترین روش در زمینه استنباط آماری نام برد.

شرایط انجام آزمون ANOVA

- نوع متغیرها: آزمون ANOVA به یک متغیر وابسته کمی و پیوسته(مربوط به اندازه‌گیری‌های سوال مدنظر) و یک متغیر مستقل کیفی(با حداقل ۲ سطح که گروه‌ها را برای مقایسه تعیین می‌کند) نیاز دارد.
- استقلال: داده‌هایی به عنوان نمونه ازهر گروه یا کل جامعه باید به تصادف انتخاب شده باشند، باید مستقل باشند. فرض استقلال اغلب بر اساس طراحی آزمایش و کنترل کامل شرایط تجربی، در نظر گرفته می‌شود. اگر بر اساس طرح آزمایش هنوز درباره استقلال اطمینان ندارید، از خود بپرسید که آیا یک مشاهده به مشاهدات دیگر ارتباطی دارد؟ اگر پاسخ، منفی است، به احتمال زیاد شما نمونه‌های مستقلی دارید.
- نرمال بودن: متغیر وابسته یا نمونه های حاصل از هر گروه یا جامعه باید دارای توزیع نرمال باشند.
- برابری واریانس‌ها: واریانس گروه‌های مختلف در جامعه، باید با یکدیگر برابر باشند(این فرض با نام همگن بودن واریانس‌ها نیز شناخته می‌شود).

نکته ۱ : تصادفی سازی در ترتیب انجام آزمایش با هدف پیشگیری از تأثیر متغیرهای اغتشاش ناشناخته که احیاناً در حین انجام آزمایش از حالت کنترل خارج شده و باعث مخدوش شدن نتایج می‌شود، ضروری است.

نکته ۲ : واریانس (σ^2) برای تمام سطوح عامل مورد نظر ثابت فرض می‌شود. این بدین معنا است که مشاهدات از توزیع نرمال با میانگین ($\mu + \tau i$) و واریانس (σ^2)که پیروی می‌کنند .

$$\mathrm{y_{ij}} \sim N(\mu + \ \ \tau_i \ \ \sigma^2)$$

نکته ۳ : به این موضوع توجه داشته باشد که تصادفی بودن نمونه‌ها و استقلال آنها از یکدیگر باید در زمان نمونه گیری کنترل شود و نمونه گیری بدون دخالت سلیقه شخصی یا عوامل دیگر باشد. نمونه‌ها باید از جامعه‌ی نرمال باشند که این شرط در زمان انجام پروژه بررسی می‌شود و اگر برقرار نباشد می‌توان از روش‌های ناپارامتری برای تحلیل استفاده نمود و یا اینکه داده ها را به توزیع نرمال تبدیل کنیم.

شیوه‌ی عملکرد آنالیزواریانس:

با توجه به تعریف آنالیزواریانس که هدف آن مقایسه‌ی میانگین دو یا چند جامعه است.

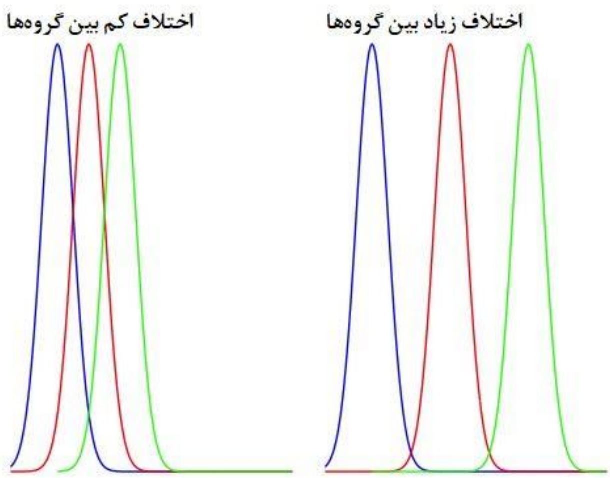
آزمون فرض :

$$\begin{cases} H_0 \quad : \quad \mu_1 = \mu_2 = \dots = \mu_a \\ H_1 \quad : \quad \mu_i \neq \mu_j \quad i \neq j \end{cases}$$

حال به شیوه‌ی عملکرد آنالیزواریانس می‌پردازیم.

مانند هر آزمون دیگر، آنالیز واریانس نیز احتیاج به یک آماره آزمون دارد. آماره آزمون برای ANOVA دارای توزیع F است. این آماره نسبت تغییرات «بین گروه‌ها»[۱] را به تغییرات «درون گروهی»[۲] اندازه‌گیری می‌کند.

«تغییرات بین گروه‌ها» ، بیانگر اختلافات بین گروه‌ها است. تصویر زیر برای درک بهتر این مفهوم مناسب است. همانطور که دیده می‌شود، در نمودارهای سمت چپ، بین میانگین گروه‌ها اختلاف زیادی وجود ندارد. در حالیکه در سمت راست، میانگین گروه‌ها دارای اختلاف محسوسی است.

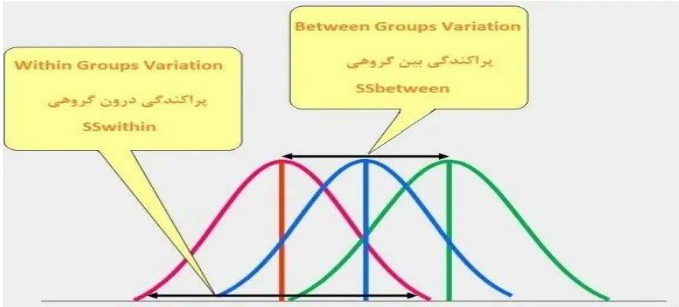


از طرف دیگر تغییرات «درون گروهی» تحت تاثیر پراکندگی اعضای هر گروه قرار دارد. در حقیقت مجموع مربعات اختلاف داده‌های هر گروه نسبت به مقدار میانگین آن گروه محاسبه شده و

---

۱ Between Groups
۲ Within Groups

حاصل برای همه گروه‌ها با یکدیگر جمع می‌شود. در تصویر زیر این مفهوم مشخص شده و پراکندگی درون گروهی و بین گروهی بطور کامل نشان داده شده است.



مقدار F براساس این دو خصوصیات محاسبه می‌شود. اگر میانگین پراکندگی بین گروهی را با MS Between و میانگین پراکندگی درون گروهی را با MSwithin نشان دهیم، مقدار آماره F را به صورت زیر محاسبه می‌کنیم.

$$F = \frac{MS_{\ between}}{MS_{\ within}}$$

بزرگ بودن مقدار F نشانه‌ای برای ردفرض صفر است، زیرا مشخص است که صورت بزرگ‌تر از مخرج است. در نتیجه گروه‌ها دارای پراکندگی بین گروهی بیشتری نسبت به پراکندگی درون گروه‌ها هستند. به این ترتیب متوجه می‌شویم که جوامعی که این گروه‌ها را تشکیل می‌دهند، یکسان نیستند. از آن‌جایی که توزیع نرمال و واریانس نیز ثابت در نظر گرفته شده است، تنها عاملی که باعث تفاوت بین جامعه‌ها است، میانگین است. پس فرض صفر که برابری میانگین گروه‌ها را نشان می‌هد، ردخواهدشد.

همچنین کوچک بودن مقدار F بیان‌گر معنا دار نبودن اختلاف بین میانگین گروه‌ها است. در نتیجه به نظر می‌رسد که همه گروه‌ها از یک جامعه آماری هستند، پس میانگین‌شان با هم برابر است.

درمورد آزمون فرض ها و مدل ها در فصل بعد بیشتر به آن میپردازیم.

اگر پراکندگی کل را با «مجموع مربعات کل»[۱] (SST)، پراکندگی بین گروهی را با «مجموع مربعات بین گروه‌ها»[۲] (SSB) و پراکندگی درون گروهی را با «مجموع مربعات درون گروهی»[۳] (SSW) نشان دهیم، خواهیم داشت:

$$SST=SSB+SSW$$

انواع داده‌ها

۱) داده‌های نامتعادل: **زمانی که تعداد مشاهدات تهیه شده برای هر تیمار(عامل)[۴] متفاوت باشد به چنین داده‌هایی، داده‌های نامتعادل گفته می شود؛ که باعث ایجاد طرح نامتعادل می‌شود .**

۲) داده‌های متعادل: **زمانی که تعداد مشاهدات تهیه شده برای هر تیمار(عامل) برابر باشد به چنین داده‌هایی، داده‌های متعادل گفته می‌شود.**

«توجه داشته‌باشید که داده‌های ما در پروژه که در انتها می‌خواهیم انجام دهیم متعادل است»

---

۱ Total Sum of Squares
۲ Between Sum of Squares
۳ Within sum of squares
۴ عامل (factor): متغیری که باعث اختلاف در جوامع می‌شود.

فصل دوم

اهداف :

انواع آنالیزواریانس

- **یک‌راهه (یک طرفه)**
  - **اثر ثابت**
  - **اثر تصادفی**
- **دوراهه (دوطرفه)**
  - **اثرثابت**
  - **تصادفی**
  - **آمیخته**

انواع آنالیزواریانس:

آنالیزواریانس یک‌راهه: در آنالیزواریانس یک‌راهه تنها به بررسی وجود یا عدم وجود اختلاف میان میانگین جوامع حاصل از تاثیر یک عامل می‌پردازیم. البته که خود این عامل دارای سطوح مختلفی می‌باشد. آنالیزواریانس یک‌راهه ثابت و تصادفی در محاسبات یکسان هستند و فقط در فرضیات تفاوت دارند.

آنالیزواریانس دوراهه: در آنالیز واریانس دوراهه به بررسی وجود یا عدم وجود اختلاف میانگین میان جوامع حاصل از تاثیر دو عامل به طور هم‌زمان می‌پردازیم. این نوع آنالیز خود به سه بخش دسته‌بندی می‌شود.

انواع آنالیزواریانس دوراهه : ۱. اثرثابت ۲. اثرتصادفی ۳. اثرآمیخته

اثرثابت: زمانی که ۲ عامل تیمار ما ثابت باشد به این معنا است که استنباط‌های گرفته شده از این تحلیل تنها قابل اعمال بر سطح منتخب است که توسط محقق برگزیده شده است. هنگامی که داده‌ها از تمام سطوح و حالت‌های ممکن یک عامل انتخاب شوند، ما آن فاکتور را تحت عنوان فاکتور دارای اثرات ثابت[۱] می‌شناسیم.

اثرتصادفی: زمانی که عامل مورد بررسی دارای سطوح و گروه‌های مختلف زیادی است، ما به همه این گروه‌ها به صورت یکسان علاقه‌مند هستیم و بین آن‌ها تفاوتی قائل نمی‌شویم، اما فقط می‌توانیم یک نمونه‌تصادفی از این سطوح را در مطالعه قرار دهیم. از آنجا که اثر این عامل بستگی به این خواهد داشت که کدام سطح انتخاب شود به آن اثرتصادفی[۲] گفته می‌شود .

اثرآمیخته: زمانی که یکی از عامل های مورد بررسی دارای اثرثابت باشد یعنی سطوح آن تنها سطوح موجود در جامعه باشند و عامل دیگر دارای اثر تصادفی باشد یعنی سطوح مورد ارزیابی به تصادف از میان تعداد زیادی از سطوح موجود در جامعه انتخاب شده باشند، آن‌گاه یک طرح آنالیزواریانس دوراهه با اثرات‌آمیخته[۳] می‌باشد.

---

۱ Fixed Effect Factor
۲ Random Factor Effect
۳ Mixed Effect

تحلیل واریانس

فرض کنید می‌خواهیم تعداد a تیمار با سطح مختلف یک عامل را با یکدیگر مقایسه کنیم. پاسخ مشاهده شده برای هر تیمار به عنوان یک متغیر تصادفی محسوب می‌شود. داده‌های حاصل از این آزمایش در جدول زیر نشان داده شده است. در این جدول، درایه y_{ij} بیانگر مشاهدهٔ j به ازای سطح عامل i مورد نظر است. به طور کلی، تعداد مشاهدات در سطح i برابر n در نظر گرفته می‌شود.

مدل‌هایی برای داده‌ها :

غالباً استفاده از یک مدل جهت توصیف مشاهدات حاصل از یک آزمایش می‌تواند مفید باشد. مدل زیر، یکی از روش‌های ارائه این مدل است:

$$y_{ij} = \mu_1 + \epsilon_{ij} \quad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

در این مدل، y_{ij} مشاهده ij ، μ میانگین سطح i عامل یا تیمار و ϵ مؤلفه خطای تصادفی[1] تعریف می‌شود. خطای تصادفی، کلیه منابع تغییرپذیری آزمایش از جمله اندازه‌گیری تغییرپذیری ناشی از عامل‌های غیرقابل کنترل، اختلاف بین واحدهای آزمایش(نظیر مواد آزمایش و غیره) و اغتشاشات پس زمینه موجود در فرآیند(نظیر تغییرپذیری در طول زمان، تأثیر عامل‌های محیطی و غیره) را شامل می‌شود. به منظور سهولت، میانگین خطاهای تصادفی برابر صفر در نظر گرفته می‌شود و یا به عبارت دیگر $E(y_{ij}) = \mu$ است. رابطه معرفی شده را مدل میانگین‌ها[2] می‌نامند. براساس رابطه زیر می‌توان مدل دیگری برای توصیف داده بدست آورد:

$$\mu_i = \mu + \tau_i \qquad i = 1, 2, \ldots, a$$

---

۱ Random error

۲ Means models

با جایگزین کردن μ_i در رابطه صفحه‌ی قبل مدل نهایی زیر حاصل می شود:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$

در مدل فوق، μ از پارامتر مشترک برای تمام تیمارها محسوب می‌شود و آن را تحت عنوان «میانگین کل»[۱] و τ_i که منحصر به تیمار i است را تحت عنوان« اثر تیمارi »[۲] می‌شناسیم. رابطه بالا را معمولاً« مدل‌اثرات»[۳] می‌نامند.

مدل میانگین‌ها و مدل اثرات، هر دو از مدل‌های آماری خطی[۴] محسوب می‌شوند. به عبارت دیگر، متغیر پاسخ y_{ij} یک تابع خطی از پارامترهای مدل است. گرچه هر دو شکل مدل مفید است ولی مدل اثرات را بیشتر از مدل میانگین‌ها در ادبیات طراحی آزمایش‌ها مشاهده می‌کنیم . این مدل به لحاظ تعریف (میانگین کل) به عنوان یک مقدار ثابت و τ_i (اثر تیمار i) به عنوان انحراف از یک مقدار ثابت (زمانی که سطح i تیمار استفاده می‌شود)، بنظر می‌رسد که مدل معقولی باشد.

رابطه‌های بالا را نیز به دلیل بررسی فقط یک عامل،مدل تحلیل واریانس«یک‌عاملی یا یک‌طرفه»[5]می‌نامند. از طرف دیگر، ترتیب انجام آزمایش باید تصادفی باشد تا محیط یکنواخت‌تری برای تیمارها(که غالباً واحدهای آزمایش نامیده میشوند) فراهم شود. در نتیجه، این طرح آزمایش یک طرح تصادفی شده کامل [۶]است. هدف اصلی در این‌جا، انجام آزمون‌های فرض مناسب برای میانگین‌ها و برآورد آن‌ها است. جهت انجام آزمون فرض، خطاهای مدل به عنوان متغیرهای تصادفی مستقل نرمال با میانگین صفر و واریانس σ^2 و در نظر گرفته است،

---

۱ Overall mean
۲ Treatment effect
۳ Effect model
٤ Linear Statistical Models
٥ one-way or single-factor analysis of variance
٦ completely randomized design

واریانس σ^2 برای تمام سطوح عامل مورد نظر ثابت فرض می شود. این بدین معنا است که مشاهدات از توزیع نرمال با میانگین ، $\mu+\tau_i$ و واریانس σ^2 پیروی می کنند و یا $y_{ij} \sim N(\mu + \tau_i\, \sigma^2)$ در این‌جا فرض می شود که مشاهدات از یکدیگر مستقل هستند.

عامل ثابت و تصادفی :

مدل آماری ارائه شده در رابطه بالا ، دو حالت مختلف را از لحاظ اثرات تیماری توصیف می کند، ابتدا اینکه، آزمایشگر می‌توانست a تیمار مختلف را به طور مشخص انتخاب کند. در چنین حالتی، او علاقه‌مند به انجام آزمون‌های فرضی در مورد میانگین‌های تیماری است و نتایج حاصل نیز فقط در مورد سطوحی از عامل که در تحلیل استفاده شده‌اند کاربرد دارد. به عبارت دیگر، نتایج حاصل را نمی‌توان به تیمارهای مشابهی که در آزمایش استفاده نشده‌اند تعمیم داد. همچنین ممکن است بخواهیم پارامتر های مدل (μ ،τi ، σ^2) را برآورد کنیم، چنین مدلی را «مدل اثرات ثابت»[۱] می‌نامند.

از طرف دیگر، ممکن است a تیمار مختلف به طور تصادفی از جامعه بزرگتری از تیمارها انتخاب شده‌باشد. در چنین حالتی، او علاقه‌مند به تعمیم نتایج حاصل از نمونه تیمارها به کلیه تیمارهای موجود در جامع (صرف نظر از این که تیمارها در تحلیل شامل گردیده‌اند یا خیر) است، تحت چنین شرایطی τ_i یک متغیر تصادفی محسوب می شود و دانش حاصل در مورد تعداد مشخصی از آن‌ها نسبتا بی‌فایده است. در عوض، فرض‌هایی در مورد تغییر پذیری τ_i مورد ارزیابی قرار گرفته و سعی می شود این تغییر پذیری تخمین زده شود، این مدل را «مدل اثرات تصادفی یامدل مؤلفه های واریانس»[۲] می‌نامند.

---

۱ fixed effects model
۲ random sample

تحلیل مدل اثرات ثابت

در این بخش روش تحلیل واریانس یک عاملی برای مدل اثرات ثابت ارائه می‌شود. به خاطر دارید که از $y_{i.}$ برای نشان دادن مجموع مشاهدات در سطح i عامل استفاده شد. حال اجازه دهید که میانگین سطح i عامل مورد مطالعه را با $\bar{y}_{i.}$ نشان دهیم، به روش مشابهی می‌توان از $y_{..}$ و $\bar{y}_{..}$ به ترتیب جهت نشان دادن مجموع کل و میانگین کل مشاهدات استفاده کرد. این کمیت‌ها به صورت زیر تعریف می‌شوند:

$$y_{i\cdot} = \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{i.} = y_{i\cdot}/n \qquad i = 1.2. \dots . a$$

$$y_{..} = \sum_{j=1}^{n} y_{ij} \quad \bar{y}_{..} = y_{..}/N$$

در رابطه فوق، N=an تعریف می‌شود. همانگونه که مشاهده می‌کنید، اندیسی که با نقطه جایگزین گردیده به معنای مجموع مشاهدات به ازای همان اندیس جایگزین شده است.

در تحلیل واریانس یک‌طرفه، علاقه‌مند به آزمون برابری میانگین‌های a تیمار و یا به عبارت دیگر $\mathrm{E}(y_{ij}) = \mu + \tau i = \mu i$ به ازای $i=1,2,\dots,a$ هستیم. فرض‌های مناسب جهت انجام این آزمون عبارتنداز:

$$\begin{cases} H_0 & \mu_1 = \mu_2 = \dots = \mu_a \\ H_1 & \mu_i \neq \mu_j \quad i \neq j \end{cases}$$

در مدل اثرات، میانگین سطح i عامل مورد مطالعه با هم به دو مؤلفه تقسیم می‌شود بطوریکه $\mu_{i=\mu+\tau i}$ باشد. معمولا μ به عنوان میانگین کل در نظر گرفته می‌شود و یا:

$$\frac{\sum_{i=1}^{a} \mu_i}{a} = \mu$$

بر اساس رابطه فوق نتیجه گیری می‌شود که:

$$\sum_{i=1}^{a} \tau_i = 0$$

به عبارت دیگر، اثرات عامل یا تیمار را می‌توان به عنوان انحرافات از میانگین کل در نظر گرفت، درنتیجه، یک روش دیگر برای بیان فرض‌های فوق آن است که از اثرات تیماری یا τ_i استفاده شود و یا:

$$\begin{cases} H_0 \quad \tau_1 = \tau_2 = \dots = \tau_a = 0 \\ H_1 \quad \tau_i \neq 0 \qquad \text{for at least one } i \end{cases}$$

بنابراین، آزمون برابری میانگین‌های تیماری به معنای آزمون بی‌اثر بودن اثرات عامل تیماری(τ_i) محسوب می‌شود، تحلیل‌واریانس را می‌توان به عنوان روش مناسبی برای آزمون برابری میانگین‌های a تیمار استفاده کرد.

تجزیه مجموع مربعات

واژه تحلیل‌واریانس از تقسیم تغییرپذیری کل به دو مؤلفه آن بر گرفته شده است، مجموع مربعات تصحیح شدهٔ کل و یا:

$$ss_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

به عنوان معیاری برای تغییر پذیری کل داده‌ها استفاده می‌شود. رابطه فوق معقول به نظر می‌رسد چرا که اگر SS_T را بر درجه آزادی مناسبی تقسیم کنیم (در این حالت an-1=N-1) آنگاه واریانس نمونه بدست می‌آید. بدیهی است که واریانس نمونه، معیار استانداردی برای تغییرپذیری است.

مجموع مربعات تصحیح شده کل SS_T را می‌توان به صورت زیر نوشت:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}\left[(\bar{y}_{i.} - \bar{y}) + (y_{ij} - \bar{y}_{i\cdot})^2\ \right]$$

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot\cdot})^2 = \mathrm{n}\sum_{i=1}^{a}\overline{(y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\cdot})^2$$
$$+2\sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})(y_{ij} - \bar{y}_{i\cdot})$$

در رابطه بالا، جمله آخر به دلیل زیر برابر صفر است:

$$\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\cdot}) = y_{i\cdot} - n\bar{y}_{i\cdot} = y_{i\cdot} - n\left(\frac{y_{i\cdot}}{n}\right) = 0$$

بنابراین :

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot\cdot})^2 = n\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\cdot})^2$$

این رابطه بیانگر این نکته است که تغییرپذیری کل داده‌ها که بر اساس مجموع مربعات کل تصحیح شده ارزیابی می‌شود را می‌توان به دو مجموع مربعات که متشکل از تفاوت بین میانگین‌های تیماری و میانگین گل، به اضافه مجموع مربعات ناشی از تفاوت بین مشاهدات درون تیماری و میانگین تیماری است تقسیم نمود. حال تفاوت بین میانگین‌های تیماری مشاهده شده و میانگین کل را می توان به عنوان معیاری برای تعیین تفاوت بین میانگین‌های تیماری در نظر گرفت. بدیهی است که اختلاف‌های موجود بین مشاهدات درون یک تیمار از میانگین آن تیمار می‌تواند فقط در اثر خطای تصادفی ایجاد شود. بنابراین، رابطه بالا را می‌توان به صورت زیر نوشت:

$$SS_T = SS_{treatments} + SS_E$$

در رابطه فوق، $SS_{Treatment}$ مجموع مربعات تیمارها(یا به عبارت دیگر بین تیمارها) و SS_E

مجموع مربعات خطا(یا به عبارت دیگر درون تیمارها) نامیده می‌شود. تعداد کل مشاهدات برابر n=N است. بنابراین، SS_T دارای N −۱ درجه آزادی است. از طرف دیگر، عامل مورد نظر دارای a سطح مختلف و a میانگین تیماری است. در نتیجه $SS_{Treatment}$ دارای a-1 درجه آزادی خواهدبود، نهایتاً، وجود n تکرار در درون هر تیمار باعث می‌شود تا n−۱ درجه آزادی برای برآورد خطای آزمایش وجود داشته‌باشد. درجه آزادی کل برای خطای آزمایش به دلیل وجود a تیمار برابر

a(n-1) =an-a=N-a است.

بررسی دو جمله سمت راست رابطه تحلیل‌واریانس می‌تواند مفید باشد. مجموع مربعات خطا را درنظر گیرید :

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2 = \sum_{j=1}^{a}\left[\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2\right]$$

همانگونه که مشاهده می‌کنید اگر جمله داخل کروشه بر n −۱ تقسیم شود آنگاه واریانس نمونه برای تیمار i بدست می‌آید و یا

$$S_i{}^2 = \frac{\sum_{j=1}^{n}(y_{ij}-y_{i\cdot})^2}{n-1} \qquad i = 1.2.\cdots.a$$

حال a واریانس نمونه را می‌توان با یکدیگر ترکیب نمود تا یک برآورد مشترک برای واریانس جامعه به صورت زیر بدست آورد:

$$\frac{(n-1)s_1{}^2 + (n-1)S_2{}^2 +\cdots +(n-1)S_a{}^2}{(n-1)+(n-1)+\cdots+(n-1)} = \frac{\sum_{i=1}^{a}\left[\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2\right]}{\sum_{i=1}^{a}(n-1)}$$

$$=\frac{SS_E}{(N-a)}$$

بنابراین، $SS_E/(N-a)$ یک برآورد ادغامی برای واریانس مشترک درون هر یک از a تیمار محسوب می شود. به روشی مشابه، اگر تفاوتی بین a تیمار وجود نداشته باشد آنگاه می توان به صورت زیر از انحرافات بین میانگین های تیماری و میانگین کل جهت برآورد و σ^2 استفاده کرد :

$$\frac{SS_T}{a-1} = \frac{n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2}{a-1}$$

به عبارت دیگر، رابطه فوق را می توان در صورت برابری میانگین های تیماری جهت برآورد و σ^2 استفاده نمود. از آنجاییکه کمیت $\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2/(a-1)$ واریانس میانگین های تیماری یا σ^2/n را برآورد می کند لذا در صورت عدم وجود تفاوت بین میانگین های تیماری، کمیت

$$n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 \Big/ (a-1)$$

باید σ^2 را برآورد کند.

همانگونه که مشاهده می‌کنید، رابطه تحلیل واریانس صفحه‌ی قبل دو برآورد، براساس تغییرات ذاتی درون تیمارها و تغییرات بین تیمارها، جهت برآورد σ^2 فراهم می‌سازد. اگر تفاوتی بین میانگین‌های تیماری وجود نداشته‌باشد آنگاه هردو برآورد باید یکسان باشند ولی اگر میانگین‌های تیماری با یکدیگر تفاوت داشته‌باشند آنگاه باید اختلاف بین میانگین‌های تیماری را علت وجود اختلاف مشاهده شده بین برآوردها دانست.

گرچه نتایج فوق به صورت توجیهی حاصل‌گردید ولی می‌توان همین نتایج را به صورت محاسباتی نیز به‌دست‌آورد .

کمیت‌های زیر را در نظر بگیرید :

$$MS_{treatment} = \frac{SS_{treatment}}{a-1}$$

$$MS_E = \frac{SS_E}{N-a}$$

میانگین مربعات [۱]نامیده می شوند. حال به بررسی امید ریاضی هر یک از این میانگین مربعات میپردازیم:

$$E(MS_E) = E\left(\frac{SS_E}{N-a}\right) = \frac{1}{N-a}\, E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{i.}\right)^2\right]$$

$$= \frac{1}{N-a}\, E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}\left({y_{ij}}^2 - 2y_{ij}\bar{y}_{i.} + {\bar{y}_{i.}}^2\right)\right]$$

$$= \frac{1}{N-a} E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}{y_{ij}}^2 - 2n\sum_{i=1}^{a}{\bar{y}_{i.}}^2 + n\sum_{i=1}^{a}{\bar{y}_{i.}}^2\right]$$

$$= \frac{1}{N-a} E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}{y_{ij}}^2 - \frac{1}{n}\sum_{i=1}^{a}{y_{i.}}^2\right]$$

با جایگزین کردن مدل مورد مطالعه (رابطه بالا) در رابطه فوق، رابطه زیر حاصل می شود :

$$E(MS_E) = \frac{1}{N-a}\, E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}(\mu + \tau_i + \epsilon_{ij})^2 - \frac{1}{n}\sum_{i=1}^{a}\left(\sum_{j=1}^{n}\mu + \tau_i + \epsilon_{ij}\right)^2\right]$$

پس از به توان رساندن و امیدریاضی گرفتن از کمیت‌های درون هر پرانتز، مشاهده می‌کنیم که جملات ϵ_{ij}^2 و $\epsilon_{i.}^2$ را به علت این که $E(\epsilon_{ij})=0$ است می‌توان به‌ترتیب با σ^2 و $n\sigma^2$ جایگزین کرد. از طرف دیگر، امیدریاضی تمام جملاتی که شامل ϵ_{ij} هستند برابر صفر است. بنابراین، رابطه آخر را بعد از به توان رساندن و تعیین امیدریاضی میتوان به صورت زیر نوشت؛

---

۱ mean squares

$$E(MS_E)=\frac{1}{N-a}\left[N\mu^2+n\sum_{i=1}^{a}{\tau_i}^2+N\sigma^2+N\mu^2-n\sum_{i=1}^{a}{\tau_i}^2-a\sigma^2\right]$$

$$E(MS_E)=\sigma^2$$

به روشی مشابه می‌توان نشان داد که :

$$E(MS_{Treatments})=\ \sigma^2+\ \frac{n\sum_{i=1}^{a}{\tau_i}^2}{a-1}$$

بنابراین، همان‌گونه که از قبل توضـــیح داده‌شـــد $MS_E = SS_E / (N-a)$ را برآورد می‌کند و اگر تفاوتی بین میانگین‌های تیماری وجود نداشته‌باشد (به این معنا که $\tau i = 0$ است) آنگاه :

$$MS_{treatment} = SS_{Treatment}/a-1$$

نیز σ^2 را برآورد خواهـد نمود. بـا این‌وجود، اگر میـانگین‌هـای تیمـاری بـا یکـدیگر تفـاوت داشـته‌باشـند آنگاه امیدریاضـی $MS_{Treatments}$ بزرگ‌تر از σ^2 خواهدبود. بدیهی اسـت که آزمون فرض عدم‌وجود تفاوت بین میانگین‌های تیماری را می‌توان با مقایسـه $MS_{Treatments}$ و MS_E انجام داد.

تحلیل آماری

حال درخصوص نحوه انجام یک آزمون آماری جهت بررسی عدم وجود اختلاف بین میانگین‌های تیماری $H_\circ:\ \mu_1=\mu_2=\cdots=\mu_a$ و یا معادل آن $H_1:\tau_1=\tau_2=\cdots=\tau_a=0$ بحث خواهد-شد. از آن‌جایی که فرض گردید خطاهای ϵij تشکیل متغیرهای تصادفی مستقل و هم‌توزیع نرمال با میانگین صفر و واریانس σ^2 را می‌دهند لذا می‌توان نتیجه‌گیری کرد که مشاهدات نیز دارای توزیع مستقل نرمال با میانگین $\mu+\tau i$ و واریانس σ^2 هستند. بنابراین مجموع مربعات متغیرهای تصادفی

نرمال است درنتیجه می‌توان نشان‌داد که SS_T/σ^2 دارای توزیع مربع‌کای با N=1 درجه‌آزادی است.

از طرف دیگر می‌توان نشان‌داد که SS_E/σ^2 دارای توزیع مربع‌کای با N-a درجه‌آزادی $SS_{Treatment}/\sigma^2$ و درصورت برقرار بودن فرض صفر $H_0 : \tau i = 0$ دارای توزیع مربع‌کای با a-1 درجه‌آزادی است.

با این‌حال، مجموع مربعات لزوما از یکدیگر مستقل نیستند چرا که جمع $SS_{Treatment}$, SS_E برابر SS_T است. قضیه زیر که یک حالت خاصی از قضیه ارائه شده توسط ویلیام کوکران[1] است را می‌توان جهت اثبات مستقل بودن SS_E از $SS_{Treatment}$ استفاده نمود.

فرض کنید Z_i به ازای i=1,2, … ,v دارای توزیع $NID(0,\sigma^2)$ است و

$$\sum_{i=1}^{v} z_i^2 = Q_1 + Q_2 + \cdots + Q_s$$

بطوریکه $s \leq v$ و Q_i دارای v_i درجه‌آزادی ($i=1,2,\ldots,s$) باشد.در این صورت می‌توان نشان داد که $Q_1 + Q_2 + \cdots + Q_s$ متغیرهای تصادفی مستقل مربع‌کای با درجه‌های‌آزادی به ترتیب $v_1 + v_2 + \cdots + v_s$ خواهند بود اگر و فقط اگر

$$v = v_1 + v_2 + \cdots + v_s$$

از آن‌جایی که مجموع درجه‌های‌آزادی $SS_{treatment}$ و SS_E برابر درجه‌آزادی کل یا ۱-N است لذا براساس قضیه کوکران می‌توان نتیجه‌گیری کرد که $SS_{treatment}/\sigma^2$ و SS_E/σ^2 متغیرهای تصادفی مستقل با توزیع مربع‌کای هستند .

بنابراین، اگر فرض صفر یا فرض عدم وجود تفاوت بین میانگین‌های تیماری برقرار باشد آنگاه نسبت.

$$F_0 = \frac{SS_{Treatment}/a-1}{SS_E/(N-a)} = \frac{MS_{Treatments}}{MS_E}$$

---
۱ William G. Cochran

دارای توزیع F با درجه های‌آزادی a-1 و N-a خواهد بود. رابطه بالا را می‌توان به عنوان آماره آزمون برای ارزیابی فرض عدم وجود تفاوت بین میانگین‌های تیماری استفاده نمود.

به‌طور کلی، با درنظرگرفتن امیدریاضـــی میانگین مربعات مشـــاهده‌می‌کنیم که MS_E یک برآوردکننده نااریب برای σ^2 است. همچنین تحت شرایط فرض صفر ، $MS_{Treatments}$ نیز یک برآوردکننده نااریب،برای σ^2 محسـوب می‌شـود. اگر فرض صـفر نادرسـت باشـد آنگاه امیدریاضـی $MS_{Treatments}$ بزرگ‌تر از σ^2خواهدبود.

بنابراین، تحت شـرایط فرض مقابل، امیدریاضـی صـورت آماره‌آزمون F بزرگ‌تر از امیدریاضـی مخرج آن خواهدبود و درنتیجه به ازای مقادیر خیلی بزرگ آماره‌آزمون F، فرض H_0 رد می‌شـــود. این بدین معنا است که ناحیه بحرانی به صورت یک‌طرفه و در سمت بالا قرار دارد.

بنابراین، اگر $F_0 > F_{\alpha,a-1,N-a}$ باشـــد آنگاه باید فرض H_0 را رد و نتیجه‌گیری کرد که بین میانگین‌های تیماری تفاوت وجود دارد. کمیت F_0 از رابطه بالا محاســـبه می‌شـــود. از مقدار P نیز می‌توان جهت تصمیم‌گیری در مورد فرض H_0 استفاده‌کرد.

با ساده‌کردن کمیت‌های $MS_{Treatments}$ و SS_T موجود در رابطه‌های بالا ، رابطه‌های محاسباتی زیر برای مجموع مربعات بدست می‌آید :

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$SS_{treatments} = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 = \frac{y_{..}^2}{N}$$

$$SS_E = SS_T - SS_{treatments}$$

در عمل از نرم‌افزارهای آماری جهت انجام این محاسـبات اسـتفاده می‌شـود. روش آزمون در جدول نشان داده‌شده‌است.

این جدول را جدول تحلیل‌واریانس[1] می‌نامند.

منبع	SSها	(درجه آزادی)	MSها	F
تیمار	$SS_{treatments} = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$	a-1	$MS_{Treatments}$	$F_{\circ} = \frac{MS_{Treatments}}{MS_E}$
خطا	$SS_E = SS_T - SS_{treatments}$	N-a	MS_E	
کل	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$	N-1		

تحلیل مدل اثرات تصادفی :

زمانی که ما بخواهیم اثر عامل را بر متغیر پاسخ بررسی کنیم ولی سطوح آن عامل از قبل مشخص نباشد یا سطوح آن خیلی زیاد باشد باید مدل با اثرتصادفی را برای آن درنظر بگیریم .از لحاظ متوسط تغییرات ، زمانی که تعداد سطوح زیاد باشد آنگاه تغییرات متوسط آنها سوال نیست بلکه می‌خواهیم بررسی کنیم تغییرات در سطوح عامل وجود دارد یا خیر؟ ، بنابراین نوع مقایسه و نوع آزمون فرض تغییر می‌کند.

مدل مورد استفاده :

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$

ظاهرا شبیه اثرعامل یک‌راهه است اما در باطن تفاوت‌های معناداری بین آن‌ها وجود دارد .

طرح با اثرثابت یک‌طرفه : بدنبال این موضوع هستیم که بررسی کنیم ، آیا متوسط تغیرات سطوح مختلف یکسان است یا خیر؟

---

[1] analysis of variance (or ANOVA) table

طرح با اثر تصادفی یک طرفه : **بررسی تغییرات این سطوح مختلف بر متغیر پاسخ به چه شکلی است.و همچنین می‌خواهیم بررسی کنیم چه مقدار از تغییرات حاصل از تیمار ما است و ه مقدار تصادفی است.**

آزمون فرض مورد بررسی :

$$\begin{cases} H_0: \sigma_t^2 = 0 \\ \mathrm{H}_1: \sigma_t^2 > 0 \end{cases}$$

تفاوت دیگر بین مدل اثرثابت با تصادفی در τ_i ها است :

مدل با اثر ثابت : τ_i ها عدد تصادفی نبوده‌اند و $\sum \tau_i = 0$

در مدل اثرتصادفی : τ_i یک متغیر تصادفی خواهدبود , چون از بین سطوح مختلف، a تا را انتخاب کرده‌ایم و چون تصادفی است دارای توزیع خواهدبود و به صورت پیش فرض توزیع آن را نرمال درنظر می‌گیریم $\tau i \sim N(0, \sigma_t^2)$ و $\epsilon ij \sim N(0, \sigma^2)$ و فرض می‌کنیم τ_i و ϵ_{ij} از هم مستقل خواهندبود .

در شرایطی که از طرفین مدل بالا واریانس گرفته شود :

$$\mathrm{Var}\,(y_{ij}) = V(\mu) + V(\tau_i) + V(\epsilon_{ij})$$
$$\mathrm{Var}\,(y_{ij}) = \sigma_t^2 + \sigma^2$$

نکته : **جدول تحلیل واریانس برای طرح با اثرتصـادفی مانند جدول تحلیل واریانس برای اثرات‌ثابت است.**

آنالیز واریانس دوطرفه:

آنالیز واریانس دوطرفه [۱] وقتی به کار می‌رود که هدف مطالعه، بررسی هم‌زمان در تأثیر دوعامل، مثل A و B باشـد. این کار تحلیل را برای آماردان جالب‌تر می‌کند زیرا ممکن اسـت یک اثر متقابل که با A و B نشـان داده می شـودبین دو عامل وجود داشـته باشـد؛ به این معنی که اثر عامل A

[۱] Two-Factor Factorial

ممکن است در همه سطوح عامل B یکسان نباشد. اگر این امر رخ دهد، آنگاه مقایسه میانگین‌های سطوح تیمار یک‌عامل مثل A ممکن است به‌دلیل اثرمتقابل AB پوشیده شود .

از این مطلب (به‌طور نادرست) چنین برداشت می‌شود که عامل A هیچ اثری بر پاسخ ندارد بنابراین، نخست باید تغییرپذیری وابسته به اثرمتقابل AB را امتحان کرد.

در آنالیزواریانس دوطرفه عامل A دارای a سطح و عامل B داری b سطح است و دارای اثرمتقابل نسبت به‌هم هستند که به آن اثرمتقابل می‌گویند.و به‌طور کلی، تعداد تکرار در یک آزمایش برابر n است.

به‌منظور تشریح حالت کلی، فرض کنید y_{ijk} پاسخ مشاهده‌شده به ازای سطح i عامل A ‎(i=1,2,……,a)‎ و سطح j عامل B ،‎(j=1,2,…,b)‎ در تکرار k ، ‎(k=1,2,…,n)‎ باشد . به‌طور کلی، مشاهدات مربوط به یک آزمایش دوعاملی به صورت جدول زیر ظاهر می‌شود. ترتیب انجام (abn) مشاهده به‌طور تصادفی است .

از آن‌جایی که یک آزمایش n مرتبه تکرار می‌شود لذا تعداد مشاهدات برابر abn خواهدبود .

		Factor B			
		1	2	…	n
Factor A	1	$y_{111}, y_{112}, \dots, y_{11n}$	$y_{121}, y_{122}, \dots, y_{12n}$		$y_{1b1}, y_{1b2}, \dots, y_{1bn}$
	2	$y_{211}, y_{212}, \dots, y_{21n}$	$y_{221}, y_{222}, \dots, y_{22n}$		$y_{2b1}, y_{2b2}, \dots, y_{2bn}$
	a	$y_{a11}, y_{a12}, \dots, y_{a1n}$	$y_{a21}, y_{a22}, \dots, y_{a2n}$		$y_{ab1}, y_{ab2}, \dots, y_{abn}$

مدل مناسب برای آنالیز واریانس دوطرفه:

مدل اثرات[1] به صورت زیر است :

$$Y_{ijk} = \mu + \tau_{\mathrm{i}} + \beta_{\mathrm{j}} + (\tau\beta)_{\mathrm{ij}} + \varepsilon_{\mathrm{ijk}} \begin{cases} i = 1\,,2\,,\ldots,a \\ j = 1\,,2\,,\ldots,b \\ k = 1\,,2\,,\ldots,n \end{cases}$$

در رابطه فوق، μ اثر میانگین‌کل، τ_i اثر سطح i ام در عامل ردیف (A) ، β_j اثر سطح j ام در عامل ستون (B)، $(\tau\beta)_{ij}$ اثرمتقابل بین τ_i و β_j مؤلفه ϵ_{ijk} خطای‌تصادفی را نشان می‌دهد.

مدل‌میانگین[2] ، مدل دیگری است که می‌توان ارائه داد :

$$Y_{ijk} = \mu_{\mathrm{ij}} + \varepsilon_{\mathrm{ijk}} \begin{cases} i = 1\,,2\,,\ldots,a \\ j = 1\,,2\,,\ldots,b \\ k = 1\,,2\,,\ldots,n \end{cases}$$

توجه داشته‌باشید که : $\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$

مفروضات مدل :

باتوجه به این موضوع که اثرات عامل یا تیمار را می‌توان به‌عنوان انحرافات از میانگین‌کل درنظر گرفت .

$$\sum_{i=1}^{a} \tau i = 0$$

$$\sum_{j=1}^{b} \beta j = 0$$

$$\sum_{i=1}^{a} (\tau\beta) ij = \sum_{j=1}^{b} (\tau\beta) ij = 0$$

[1] Effect model
[2] Means model

آنالیز واریانس دوطرفه با اثرثابت (هر دو عامل ثابت)[۱]:

زمانی که ۲ عامل تیمار ما ثابت باشد به این معنا است که استنباط‌های گرفته‌شده از این تحلیل تنها قابل اعمال بر سطح منتخب است که توسط محقق برگزیده شده‌است.در قسمت‌های قبل به‌طور کامل به تعریف اثرثابت پرداخته‌ایم .

آزمون فرض :

در آنالیز واریانس دوطرفه با اثرثابت ما علاقه‌مند هستیم موارد زیر را بررسی کنیم :

۱) بررسی برابری اثر عامل A با a سطح

$$\begin{cases} H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0 \\ H_1: at\ least\ one\ \tau_i \neq 0 \end{cases}$$

همان‌طور که قبلا در بخش آنالیزواریانس یک‌طرفه با اثرثابت گفته‌شد ، آزمون تاثیر اثر عامل (τi) با آزمون برابری میانگین های تیمار a یکسان است و یک نتیجه را به ما می‌دهد.

۳) بررسی برابری اثرعامل B با b سطح

$$\begin{cases} H_0: \beta_1 = \beta_2 = \cdots = \beta_a = 0 \\ H_1: at\ least\ one\ \beta_i \neq 0 \end{cases}$$

۴) بررسی وجود اثرمتقابل بین عامل A و B هستیم .

---

[۱] Two-Factor Factoria the Fixed Effects l

$$\begin{cases} H_0 \ : (\tau\beta)_{\text{ij}} = 0 \quad for\ all\ i,j \\ H_1 \ : at\ least\ one : (\tau\beta)_{\text{ij}} \neq 0 \end{cases}$$

تحلیل آماری مدل اثرات ثابت :

اجازه دهید مجموع مشاهدات سطح i عامل A را با $\text{y}_{\text{i}}..$، مجموع مشاهدات سطح j عامل B را با $\text{y}.j.$، مجموع مشاهدات سلول ij را با $\text{y}_{\text{ij}}.$ و مجموع کل مشاهدات را با $\text{y}...$ نشان دهیم. همچنین از $\bar{y}i..$، $\bar{\text{y}}.j.$ ، $\bar{\text{y}}ij.$ ، $\bar{\text{y}}...$ به ترتیب جهت تعریف میانگین‌های ردیف، ستون، سلول و کل استفاده می‌شود.

بصورت ریاضی:

$$y_{\text{.i.}} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} \quad \bar{\text{y}}_{\text{i..}} = \frac{y_{i..}}{bn} \qquad i = 1,2, \dots, b$$

$$y_{.j.} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk} \quad \bar{\text{y}}_{\text{.i.}} = \frac{y_{i.}}{an} \qquad j = 1,2, \dots, b$$

$$y_{ij\cdot} = \sum_{k=1}^{n} y_{ijk} \quad \bar{\text{y}}_{\text{ij.}} = \frac{y_{ij\cdot}}{n} \qquad i = 1,2, \dots, a$$

$$j = 1,2, \dots, b$$

$$y_{ij\cdot} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} \qquad \bar{\text{y}}_{\dots} = \frac{y_{\dots}}{abn}$$

تجزیه مجموع مربعات:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} (y_{ijk} - \bar{y}_{\dots})^2 =$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} \left[(\bar{y}_{i_{..}} - \bar{y}_{\dots}) + (\bar{y}_{i.j} - \bar{y}_{\dots}) + (\bar{\text{y}}_{\text{ij.}} - \bar{\text{y}}_{\text{i..}} - \bar{\text{y}}_{\text{.i.}} + \bar{y}_{\dots}) \ \right]^2$$

$$= bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 + an\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})$$

$$+n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.i.} + \bar{y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y^2 - \bar{y}_{ij.})^2$$

همان‌گونه که مشاهده می‌کنید مجموع مربعات کل، به مجموع مربعات ردیف‌ها یا عامل A یعنی (SS_A) و مجموع مربعات ستون‌ها یا عامل B یعنی (SS_B) ، مجموع مربعات اثرمتقابل بین A و B یعنی (SS_{AB})و مجموع مربعات خطا (SS_E) افراز شده‌است.

بنابراین می توان نتیجه گرفت : $SS_T = SS_A + SS_B + SS_{AB} + SS_E$

تعداد درجه های آزادی هر یک از مجموع مربعات برابر است با:

اثر	درجه آزادی
A	a-1
B	b-1
اثرمتقابل AB	(b-1) (a-1)
خطا	ab(n-1)
کل	abn-1

تقسیم هر یک از مجموع مربعات بر درجه آزادی متناظر آن ، تشکیل یک میانگین مربعات را می‌دهد امید ریاضی هر یک از میانگین مربعات برابر است با :

$$E(MS_A) = E\left(\frac{SS_A}{a-1}\right) = \sigma^2 + \frac{bn\sum_{i=1}^{a}{\tau_i}^2}{a-1}$$

$$E(MS_B) = E\left(\frac{SS_B}{b-1}\right) = \sigma^2 + \frac{an\sum_{i=1}^{b}\tau_i^{\,2}}{b-1}$$

$$E(MS_{AB}) = E\left(\frac{SS_{AB}}{(a-1)(b-1)}\right) = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^{\,2}}{(a-1)(b-1)}$$

$$E(MS_E) = E\left(\frac{SS_E}{ab(n-1)}\right) = \sigma^2$$

اگر فرضهای صفر مربوط به عدم وجود اثر تیماری ردیف، عدم وجود اثر تیماری ستون و عدم وجود اثر متقابل درست باشد آنگاه ، MS_A ، MS_B ، MS_{AB} و MS_E همگی σ^2 را برآورد می کنند. با این حال، اگر به عنوان مثال، تفاوتی بین اثرات تیماری ردیف وجود داشته باشد آنگاه , MS_A بزرگتر از MS_E خواهد بود. به همین صورت، اگر اثرات تیماری ستون و متقابل نیز وجود داشته باشد آنگاه میانگین مربعات آنها بزرگتر از MS_E خواهد بود. بنابراین، به منظور ارزیابی معنادار بودن هر یک از اثرات اصلی و اثر متقابل نیاز است که میانگین مربعات هر یک از آنها بر میانگین مربعات خطا تقسیم شود. مقادیر بزرگ این نسبتها حاکی از عدم تائید فرضهای صفر توسط داده های موجود است.

جدول آنالیز واریانس [1] دوطرفه با اثر ثابت (هر دو عامل ها ثابت)

منبع	Dfدرجه (آزادی)	SSها	MSها	F
Aتیمار	i-1	SSA	MSA	$FA = \frac{\text{MSA}}{\text{MSE}}$
Bتیمار	j-1	SSB	MSB	$FB = \frac{\text{MSB}}{\text{MSE}}$
ABاثرمتقابل	(i-1)(j-1)	SSAB	MSAB	$FAB = \frac{\text{MSAB}}{\text{MSE}}$
خطا	ij(k-1)	SSE	MSE	
کل	n-1			

[1] analysis of variance table

روش دیگر برای محاسبه ی مجموع مربعات :

مجوع مربعات کل را می توان :

$$SS_{\mathrm{T}} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} {y_{ijk}}^2 - \frac{y_{...}^2}{abn}$$

مجموع مربعات اثرات اصلی :

$$SS_{\mathrm{A}} = \frac{1}{bn}\sum_{i=1}^{a} {y_{i..}}^2 - \frac{y_{...}^2}{abn}$$

$$SS_{\mathrm{B}} = \frac{1}{an}\sum_{j=1}^{b} {y_{.j.}}^2 - \frac{y_{...}^2}{abn}$$

مجوع مربعات اثر متقابل :

محاسبه و SSAB در دو مرحله می تواند از سهولت بیشتری برخوردار باشد. ابتدا، مجموع مربعات مربوط به مجموع ab سلول که مجموع مربعات جزئی[۱] نامیده میشود را محاسبه می کنیم:

$$SS_{\mathrm{Subtotals}} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} {y_{ij.}}^2 - \frac{y_{...}^2}{abn}$$

این مجموع مربعات شامل SSA و SSB است. بنابراین، در گام دوم SSAB بصورت زیر محاسبه می شود:

$$SS_{AB} = SS_{\mathrm{Subtotals}} - SS_{\mathrm{A}} - SS_{\mathrm{B}}$$

SSE را می توان از طریق تفریق بصورت زیر محاسبه کرد:

$$SS_{E} = SS_{\mathrm{T}} - SS_{AB} - SS_{\mathrm{A}} - SS_{\mathrm{B}}$$

---

[۱] subtotals

آنالیزواریانس دوطرفه با اثر تصادفی (هر دو عامل تصادفی)[1]

همان گونه که در قسمت آنالیزواریانس یک طرفه با اثرتصادفی گفته شد در اینجا نیز صدق می‌کند. در برخی موقعیت های تجربی، سطوح عامل به طور تصادفی از جامعه بزرگتر انتخاب می‌شوند . در چنین حالتی، محقق علاقمند به تعمیم نتایج حاصل از نمونه تیمارها به کلیه تیمارهای موجود در جامعه (صرف نظر از این که تیمارها در تحلیل شامل گردیده اند یا خیر) است .

مدل مناسب :

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \quad \begin{cases} i = 1,2,\dots,a \\ j = 1,2,\dots,b \\ K = 1,2,\dots,n \end{cases}$$

در رابطه فوق، μ ما اثر میانگین کل، τ_i اثر سطح i عامل ردیف (A)، β_j اثر سطح j عامل ستون (B)، $(\tau\beta)_{ij}$ اثر متقابل بین τ_i و β_j مؤلفه ϵ_{ijk}, خطای تصادفی را نشان میدهد.
توجه داشته باشید که ($\tau_i, \beta_j, (\tau\beta)_{ij}, \epsilon_{ijk}$ به طور معمول دارای توزیع نرمال با میانگین صفر و واریانس های

$$v(\tau_i) = \sigma^2_{\tau_i}$$

$$v(\beta_j) = \sigma^2_{\beta_j}$$

$$v(\tau\beta)_{ij} = \sigma^2_{\tau\beta_{ij}}$$

$$v(\epsilon_{ijk}) = \sigma^2$$

بنابراین واریانس هرمشاهده ای است:

$$v(y_{ijk}) = \sigma^2_\tau + \sigma^2_\beta + \sigma^2_{\tau\beta} + \sigma^2$$

$(\sigma^2_\tau, \sigma^2_\beta, \sigma^2_{\tau\beta}, \sigma^2)$ مولفه‌های واریانس[2] هستند .

مدل در آنالیزواریانس دوطرفه با اثرتصادفی با آنالیزواریانس دوطرفه با اثرثابت یکی است ، اما تفاوت آن‌ها در آرمون فرض و جدول آنالیزواریانس می‌باشد .

---

[1] The Two-Factor Factorial with Random Factors
[2] variance components

آزمون فرض :

$$\begin{cases} H_0: \sigma_\tau^2 = 0 \\ H_1: \sigma_\tau^2 \neq 0 \end{cases}$$

$$\begin{cases} H_0: \sigma_\beta^2 = 0 \\ H_1: \sigma_\beta^2 \neq 0 \end{cases}$$

$$\begin{cases} H_0: \sigma_{\tau\beta}^2 = 0 \\ H_1: \sigma_{\tau\beta}^2 \neq 0 \end{cases}$$

توجه داشته باشید که محاسبات عددی در تجزیه مجوع مربعات و تحلیل واریانس بدون تغییر باقی می ماند. یعنی SSA، SSB، SSAB، SST و SSE همگی مانند حالت اثرات ثابت محاسبه می شوند . اما میانگین مربعات آن هامتفاوت می باشد .

میانگین مربعات :

$$E(MS_A) = \sigma^2 + n\sigma_{\tau\beta}^2 + bn\sigma_\tau^2$$

$$E(MS_B) = \sigma^2 + n\sigma_{\tau\beta}^2 + bn\sigma_\beta^2$$

$$E(MS_{BA}) = \sigma^2 + n\sigma_{\tau\beta}^2$$

$$E(MS_E) = \sigma^2$$

- **برای بررسی آزمون، $H_0: \sigma_{\tau\beta}^2 = 0$ می‌توان از آماره ازمون F_0 استفاده کرد.**

$$F_0 = \frac{MS_{AB}}{MS_E}$$

فرض H_0 ردمی‌شود اگر صورت‌کسر (MS_{AB}) بزرگ تر از مخرج (MS_E) باشد و توزیع F_0 به صورت $F_{(a-1)(b-1),ab(n-1)}$ می باشد .

- **برای بررسی آزمون ، $H_0: \sigma_\tau^2 = 0$ می توان از آماره ازمون F_0 استفاده کرد.**

$$F_0 = \frac{MS_A}{MS_{AB}}$$

توزیع F_0به صورت $F_{(a-1),(a-1)(b-1)}$می‌باشد .

- **برای بررسی آزمون $H_0: \sigma_\beta^2 = 0$ می‌توان از آماره ازمون F_0استفاده کرد.**

$$F_0 = \frac{MS_B}{MS_{AB}}$$

توزیع F_0به صورت $F_{(b-1),(b-1)(a-1)}$می‌باشد .

روش دیگر برای محاسبه میانگین مربعات :

$$\hat{\sigma}_{\tau\beta}^2 = \frac{MS_{AB} - MS_E}{n}$$

$$\hat{\sigma}_\beta^2 = \frac{MS_B - MS_{AB}}{an}$$

$$\hat{\sigma}_\tau^2 = \frac{MS_A - MS_{AB}}{bn}$$

$$\hat{\sigma}^2 = MS_E$$

جدول آنالیز واریانس دوطرفه با اثر تصادفی (هر دو عامل تصادفی)

منبع	Df (درجه آزادی)	SSها	MSها	F
تیمار A	i-1	SSA	MSA	$FA = \frac{\text{MSA}}{\text{MSAB}}$
تیمار B	j-1	SSB	MSB	$FB = \frac{\text{MSB}}{\text{MSAB}}$
اثر متقابل AB	(i-1) (j-1)	SSAB	MSAB	$FAB = \frac{\text{MSAB}}{\text{MSE}}$
خطا	Ij(k-1)	SSE	MSE	
کل	n-1	SST		

آنالیزواریانس دوطرفه با اثر آمیخته(یک عامل با اثرثابت ،عامل دیگربا اثرتصادفی)

اکنون ما شرایطی را در نظر می گیریم که یکی از عوامل A ثابت و دیگری عامل B

تصادفی است این تحلیل واریانس مدل آمیخته (مختلط)[۱] نامیده می‌شود

مدل مناسب :

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \quad \begin{cases} i = 1,2, \dots , a \\ j = 1,2, \dots , b \\ K = 1,2, \dots , n \end{cases}$$

در رابطه فوق، μ ما اثر میانگین کل، τ_i اثر سطح i عامل ردیف (A)، β_j اثر سطح j عامل ستون (B)، $(\tau\beta)_{ij}$ اثر متقابل بین τ_i و β_j مؤلفه ϵ_{ijk}, خطای تصادفی را نشان میدهد.

مفروضات مدل :

ما فرض می‌کنیم که τ_i ها اثرهای‌ثابت هستند به طوریکه :

$$\sum_{i=1}^{a} \tau_i = 0$$

و همچنین β_j ها اثر های تصادفی هستند به طوریکه : $[(a\text{-}1)a]/\sigma_{tB}^2$

$$NID(0, \sigma_\beta^2)$$

و اثر متقابل ، (τβ)ij دارای اثر تصادفی با آماره عادی با میانگین ۰ و واریانس $[(a-1)/\sigma_{tB}^2]$ با این حال :

$$\sum_{i=1}^{a} (\tau\beta)_{ij} = (\tau\beta)_{.j} = 0 \qquad j = 1,2, \dots , b$$

این بیانگر آن است که عناصر برهم‌کنش خاصی در سطوح مختلف عامل ثابت وجود دارد

مستقل نیستند در واقع، ممکن است ان را نشان دهیم :

---

[۱] The Two-Factor Mixed Model

$$Cov\left[(\tau\beta)_{ij}, (\tau\beta)_{i'j}\right] = -\frac{1}{a}\sigma^2_{\tau\beta} \qquad i \neq i'$$

کوواریانس بین $(\tau\beta)_{ij}, (\tau\beta)_{i'j}$ در حالی که $i \neq i'$ برابر ۰ است.وخطای تصادفی ϵ_{ijk} ، $NID(0,\sigma^2)$.
زیرا مجموع اثرات متقابل بر سطوح عامل ثابت برابر است صفر، این نسخه از مدل ترکیبی اغلب مدل محدود [۱] نامیده می شود.

آزمون فرض‌ها:

$$\begin{cases} H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0 \\ H_1: \tau_i \neq 0 \end{cases}$$

$$\begin{cases} H_0: \sigma^2_B = 0 \\ H_1: \sigma^2_B \neq 0 \end{cases}$$

$$\begin{cases} H_0: \sigma^2_{\tau\beta} = 0 \\ H_1: \sigma^2_{\tau\beta} \neq 0 \end{cases}$$

در مدل معرفی شده ، واریانس $(\tau\beta)_{ij}$ هست ، $[(a-1)a]/\sigma^2_{\tau\beta}$ که به جای‌آن $\sigma^2_{\tau\beta}$ استفاده می‌شود ،برای ساده کردن میانگین مربعات مورد انتظار ، ما فرض می‌کنیم $(\tau\beta)_{.j} = 0$ ، پس همچنین بر میانگین مربعات مورد انتظار،تأثیر دارد.

میانگین مربعات :

$$E(MS_A) = \sigma^2 + n\sigma^2_{\tau\beta} + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$$
$$E(MS_B) = \sigma^2 + an\sigma^2_{\tau}$$
$$E(MS_{BA}) = \sigma^2 + n\sigma^2_{\tau\beta}$$
$$E(MS_E) = \sigma^2$$

---

[۱] restricted model

- برای بررسی آزمون، $H_1: \tau_i \neq 0$ می‌توان از آماره آزمون F_0 استفاده کرد.

$$F_0 = \frac{MS_A}{MS_{AB}}$$

توزیع F_0 به صورت $F_{(a-1)),(a-1)(b-1)}$ می باشد .

- برای بررسی آزمون، $H_0: \sigma_B^2 = 0$ می‌توان از آماره آزمون F_0 استفاده کرد:

$$F_0 = \frac{MS_B}{MS_E}$$

توزیع F_0 به صورت $F_{(b-1),ab(n-1)}$ می‌باشد .

- برای بررسی آزمون ، $H_0: \sigma_{\tau\beta}^2 = 0$ می‌توان از آماره آزمون F_0 استفاده کرد.

$$F_0 = \frac{MS_{AB}}{MS_E}$$

توزیع F_0 به صورت $F_{(a-1)(b-1),ab(n-1)}$ می باشد .

روش های محاسبات :

در مدل مختلط می توان اثرات عامل ثابت را به صورت زیر تخمین زد

$$\hat{\mu} = \bar{y}_{...}$$

$$\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...} \qquad i = 1,2,\dots,a$$

مولفه های واریانس (σ_B^2 ، σ_{tB}^2 ، σ^2) می توان با استفاده از تحلیل واریانس تخمین زد

به روش هایی زیر :

$$\hat{\sigma}^2_{\tau\beta} = \frac{MS_{AB} - MS_E}{n}$$

$$\hat{\sigma}^2_{\beta} = \frac{MS_B - MS_E}{an}$$

$$\hat{\sigma}^2 = MS_E$$

جدول آنالیز واریانس دوطرفه با اثرآمیخته (یک عامل ثابت و دیگری تصادفی)

منبع	Df(درجه آزادی)	SSها	MSها	F
تیمار A	i-1	SSA	MSA	$FA = \frac{\mathrm{MSA}}{\mathrm{MSAB}}$
تیمار B	j-1	SSB	MSB	$FB = \frac{\mathrm{MSB}}{\mathrm{MSAB}}$
اثر متقابل AB	(i-1) (j-1)	SSAB	MSAB	$FAB = \frac{\mathrm{MSAB}}{\mathrm{MSE}}$
خطا	Ij(k-1)	SSE	MSE	
کل	n-1	SST		

ارزیابی کفایت مدل

تحلیل باقی‌مانده‌ها به‌عنوان ابزار اصلی برای بررسی کفایت مدل به حساب می‌آید .

روش به دست آوردن باقی‌مانده‌ها :

آنالیزواریانس یک‌طرفه : $e_{ij}=y_{ij}-\tilde{y}_{ij}$ $\longleftarrow$ $y_{ij} = \bar{y}_{i.}$

آنالیزواریانس دوطرفه : $e_{ijk}=y_{ijk}-\hat{y}_{ijk}$ $\longleftarrow$ $y_{ijk}= y_{ij.}$

تجزیه تغییرپذیری موجود در مشاهدات از طریق رابطه تحلیل واریانس یک روش کاملا ریاضی است، با این حال، استفاده از روش تجزیه یا افراز کردن جهت انجام آزمون عدم وجود تفاوت بین میانگین-های تیماری، نیاز به برقراری مفروضات خاصی دارد.

و همچنین برقراری توزیع مستقل نرمال برای خطاها با میانگین صفر و واریانس نامعلوم ولی ثابت σ^2 اگر این مفروضات مورد تأیید قرارگیرند آنگاه روش تحلیل‌واریانس، آزمون دقیقی برای ارزیابی فرض عدم وجود تفاوت بین میانگین های تیماری خواهد بود با این‌حال، در عمل این مفروضات معمولا به طور دقیق برقرار نخواهد بود. در این صورت اعتماد به نتایج حاصل از تحلیل واریانس، قبل از تأیید این مفروضات، تصمیم عاقلانه‌ای نخواهد بود. نقض مفروضات اولیه و کفایت مدل را می‌توان به راحتی از طریق بررسی باقی‌مانده‌ها [1] (مانده‌ها یا پس‌مانده‌ها) انجام داد.

بررسی باقی‌مانده‌ها، بخش متداولی از تحلیل‌واریانس است. اگر مدل استفاده شده مناسب باشد آنگاه باقیمانده‌ها باید فاقد ساختار [2] باشند و یا به عبارت دیگر، باقیمانده‌ها نباید روند قابل رؤیتی را منعکس کنند. با مطالعه باقیمانده‌ها می‌توان به عدم کفایت مدل و نقض مفروضات در نظر گرفته شده پی برد. در این بخش نشان می‌دهیم که چگونه می‌توان از طریق روش های نموداری، کفایت مدل را به سهولت بررسی و با موارد غیر عادی متداول مقابله نمود.

فرض نرمال:

فرض نرمال بودن داده ها را می توان بوسیله رسم هیستوگرام باقیمانده‌ها بررسی کرد. اگر فرض $\mathrm{NID}(\cdot, \sigma^2)$ برای باقیمانده‌ها برقرار باشد آنگاه هیستوگرام باقیمانده‌ها باید شبیه توزیع نرمال و در حول مقدار صفر متمرکز باشد. متأسفانه، نمونه های کوچک غالبا تغییرات قابل ملاحظه‌ای را از خود نشان می‌دهند. بنابراین، فاصله گرفتن از توزیع نرمال به میزان متعارف لزوما به معنای نقض جدی مفروضات محسوب نمی شود. فاصله گرفتن بیش از اندازه از توزیع نرمال را باید جدی تلقی نمود و بیشتر مورد بررسی قرار داد .

رسم نمودار احتمال نرمال برای باقیمانده‌ها، روش مفیدی برای بررسی فرض نرمال بودن مشاهدات محسوب می شود. در تحلیل واریانس، انجام این کار از طریق باقیمانده‌ها معمولا می تواند

---

[1] Structureless
[2] Residuals

اثر بخش تر (و راحتر) باشد. اگر توزیع باقیمانده‌ها نرمال باشد آنگاه نمودار حاصل بصورت یک‌خط راست ظاهر می‌شود. در زمان ارزیابی این نمودار باید بر مقادیر مرکزی تأکید داشت تا نقاط دور افتاده.

بطور کلی، در تحلیل واریانس مدل اثرات ثابت، انحراف از توزیع نرمال به میزان متعارف یک مشکل جدی محسوب نمی شود (مباحث مربوط به آزمونهای تصادفی سازی در بخش قبلی را بخاطر بیاورید). یک توزیع خطا با دنباله های خیلی‌نازکتر باضخیم تر از توزیع نرمال، به مراتب از اهمیت بیشتری در مقایسه با توزیعی که دارای چولگی است برخوردار است. از آنجائیکه، آزمون F فقط به میزان نسبتا کمی تحت تأثیر قرار میگیرید، اصطلاحا گفته می‌شود که تحلیل‌واریانس (و روش‌های مرتبط با آن نظیر مقایسات چندگانه) نسبت به فرض نرمال، مقاوم یا استوار [1] است. فاصله گرفتن از توزیع نرمال معمولا باعث میشود تا سطح معنادار واقعی و توان آزمون به میزان نسبتا کمی با مقادیر اعلام شده تفاوت داشته باشد. بطور کلی، توان آزمون از مقدار واقعی آن کمتر است. نرمال نبودن داده ها به میزان نسبتا زیادی مدل اثرات تصادفی را تحت تاثیر قرار میدهد.

یک مسأله بسیار متداول در نمودارهای احتمال نرمال، مشاهده یک باقیمانده بزرگ در بین باقیمانده ها است. چنین باقیمانده ای را غالبأ یک نقطه دور افتاده می‌نامند. وجود یک یا چند نقطهٔ دور افتاده می‌تواند به طور جدی تحلیل‌واریانس را با مشکل مواجه سازد. بنابراین، نقاط دور افتاده نیاز به یک بررسی جدی دارند. در برخی مواقع، دلیل وجود مشاهدات دور افتاده می تواند اشتباه در محاسبات، کد کردن داده‌ها یا ثبت داده‌ها باشد.

اگر موارد مذکور، علل رسم یک نقطه دور افتاده باشد آنگاه باید شرایط انجام آزمایش برای این مورد خاص بررسی شود. اگر نقطهٔ دور افتاده مورد نظر دارای مقدار مناسبی باشد ،آنگاه اطلاعات موجود در این نقطه می‌تواند به مراتب بیشتر از سایر داده‌ها باشد، بنابراین، در چنین شرایطی باید با احتیاط عمل نمود و چنین نقاطی را به راحتی حذف نکرد، مگر اینکه شواهد غیر آماری معقولی برای انجام این کار وجود داشته باشد. در بدترین حالت، ممکن است مجبور به انجام دو تحلیل مختلف باشیم. یک تحلیل با در نظر گرفتن نقاط دور افتاده و یک تحلیل بدون در نظر گرفتن نقاط دور افتاده .

---

[1] Robust

چند روش آماری مختلف برای شـناسـایی نقاط دور افتاده وجود دارد ، یک ارزیابی تقریبی برای نقاط دور افتاده، استفاده از باقیمانده‌های استاندارد شده و یا:

$$d_{ij} = \frac{e_{ij}}{\sqrt{MS_E}}$$

است. اگر خطاها ϵ_{ij} را از توزیع ($N(0,\sigma^2)$) پیروی کند آنگاه باقیمانده های استاندارد شده باید تقریبا دارای توزیع نرمال با میانگین صفر و واریانس واحد باشد. بنابراین، حدود ۶۸ درصد باقیمانده های استاندارد شده باید بین حدود + -۱، حدود ۹۵ درصد آنها بین ۲+ -و تقریبا همه آنها بین ۳ + -واقع شود، باقیمانده‌ای که دورتر از ۳ یا ۴ انحراف معیار از میانگین صفر واقع شود بعنوان یک نقطهٔ دور افتاده بالقوه محسوب می‌شود.

نمودار باقیمانده ها بر حسب زمان:

رسم داده‌ها بر حسب ترتیب زمانی آنها می‌تواند در شناسایی همبستگی بین باقیمانده‌ها مفید واقع شود، وجود تسلسل های مثبت و منفی بیانگر همبستگی مثبت بین باقیمانده ها و به معنای نقض فرض استقلال خطاها است. این مسأله یک مشکل جدی بالقوه محسوب می شود و برطرف کردن آن معمولا کار ساده‌ای نیست. بنابراین، باید سعی شود تا در زمان جمع آوری . داده از بروز چنین مشکلی پیشگیری بعمل آید. تصادفی سازی مناسب آزمایش می تواند گام مهمی جهت دستیابی به استقلال مشاهدات باشد.

نمودار باقیمانده‌ها برحسب مقادیر برازش شده :

اگر مدل صحیح و مفروضات برقرار باشد آنگاه باقیمانده‌ها نباید ساختار خاصی از خود نشان دهند و یا ارتباط خاصی با متغیرهای دیگر از جمله پاسخ پیش بینی شده داشته باشند. با رسم نمودار باقیمانده ها بر حسب مقادیر پیش بینی شده $\hat{y}_{ij}$ می‌توان به سهولت این نکته را بررسی کرد. به خاطر دارید که برای مدل آزمایش یک‌عاملی، مقدار پیش‌بینی‌شده $\hat{y}_{ij}$ برابر میانگین تیمار i و یا $\bar{y}_{i.}$ است. در این نمودار نباید روند مشهودی مشاهده شود.

گاهی با افزایش مقادیر مشاهدات، واریانس آنها نیز افزایش می یابد. این حالت زمانی رخ می‌دهد که خطا یا اغتشاشات پس زمینه آزمایش، درصد ثابتی از مقدار مشاهده باشد. (معمولا در

اغلب ابزارهای اندازه گیری، خطا درصدی از مقیاس اندازه گیری است. در چنین شرایطی، با افزایش مقدار y_{ij} مقدار باقیمانده افزایش یافته و نمودار باقی‌مانده‌ها برحسب $\hat{y}_{ij}$ شبیه به انتهای بزرگ یک قیف با بلندگو به نظر می رسد. همچنین واریانس غیرثابت را می‌توان در داده های غیرنرمال با توزیع های چوله که معمولا در این توزیع‌ها، واریانس تابعی از میانگین است انتظار داشت.

اگر فرض همگن بودن واریانس نقض شود آنگاه آزمون F در مدل اثرات ثابت متعادل (اندازه نمونه مساوی برای تمام تیمارها) به میزان نسبتا کمی تحت‌تأثیر قرارمی‌گیرد. اما در طرح‌های غیرمتعادل یا حالاتی که یک واریانس خیلی بزرگ‌تر از سایر واریانس‌ها باشد، این مشکل خیلی حادتر خواهد بود، به ویژه، اگر تیمارهایی که از واریانس بزرگتری برخوردار هستند دارای اندازه نمونه کوچکتری نیز باشند آنگاه خطای نوع I واقعی بیشتر از مقدار مورد انتظار خواهد بود (یا سطوح اطمینان مربوط به فواصل، کمتر از مقادیر تعیین شده است). برعکس، اگر تیمارهایی که از واریانس بزرگ‌تری برخودار هستند دارای اندازه نمونه‌های بزرگ‌تری نیز باشند آنگاه سطوح معنادار حاصل، کوچک‌تر از مقادیر مورد انتظار خواهد بود (سطوح اطمینان بیشتر است، این موضوع را می توان توجیح خوبی برای انتخاب اندازه نمونه های یکسان درنظرگرفت، در مدل‌های اثرات تصادفی، ثابت نبودن واریانس می‌تواند به میزان قابل ملاحظه ای استنباط های حاصل در مورد مؤلفه های واریانس را، حتی برای طرح های متعادل، تحت تأثیر قرار دهد.

استفاده از تبدیل تثبیت واریانس و انجام تحلیل واریانس با داده‌های تبدیل شده، روشی برای برطرف سازی مشکل واریانس غیر ثابت که به دلایل فوق ایجاد شده باشد محسوب می شود. در صورت معلوم بودن توزیع مشاهدات می‌توان از این اطلاعات جهت انتخاب نوع تبدیل استفاده‌کرد.

نمودار باقیمانده‌ها بر حسب متغیرهای دیگر :

اگر مشاهدات بر حسب متغیرهای دیگری که احیانا پر پاسخ تأثیر گذارند تهیه شده باشد آنگاه باید باقیمانده ها را بر حسب این متغیرها نیز رسم کرد. مشاهده هرگونه روند غور تصادفی پر روی نمودارهای باقیمانده به معنای تأثیرگذار بودن متغیر مورد نظر بر روی پاسخ است. این بدین معنا است که متغیر مورد مطالعه را باید با دقت بیشتری کنترل و یا باید آن را در آزمایش لحاظ کرد.

مقایسه های زوجی میانگین های تیماری:

در اغلب مواقع علاقه‌مند به مقایسه فقط یک زوج میانگین هستیم.

به عبارت دیگر با آزمون اختلاف بین دو میانگین تیماری مشخص می‌شود که کدام‌یک از میانگین‌ها باعث ایجاد اختلاف شده است.

در آنالیزواریانس یک‌عامله درصورت رد فرض صفر (یعنی تفاوت معناداراست)ما می‌توانیم برای تشخیص تفاوت درون‌گروه‌ها از آزمون‌های تعقیبی استفاده کنیم.در واقع ببینیم این تفاوت در بین کدام‌یک از گروه‌ها وجود دارد.

تحلیل واریانس نشان می‌دهد که آیا نمونه‌ها متعلق به جامعه هستند یا خیر. در صورتی که فرض صفر رد شود، معلوم نیست که کدام‌یک از نمونه‌ها در کدام جامعه قرار دارند. به عبارت دیگر، معنی‌دار شدن نسبت F به ما نمی‌گوید که اختلاف بین کدام جفت از میانگین‌ها معنی‌دار است. بلکه با آماره F تنهاه می‌توان پی‌ببریم که اختلاف بین میانگین گروه‌ها معنی‌دار است .

در هنگام تحلیل‌واریانس و به عبارتی آزمون تفاوت میانگین‌ها در بین سه گروه و بیشتر، علاوه بر آزمون معنی‌داری این تفاوت میانگین‌ها، لازم است که به کیفیت این تفاوت نیز پی‌ببریم. چرا که آشنایی با کیفیت این تفاوت، نقش مهمی در آزمون فرضیه و تفسیر نتایج آن و همچنین در جمع‌بندی و ارائه راه‌کارها برای گزارش دارد. در نرم‌افزار SPSS ، نوع آزمون مقایسه چندگانه تعریف‌شده که به کمک آن‌ها می‌توانیم چگونگی تفاوت میانگین نمره گروه‌ها از هم‌دیگر را تشخیص دهیم. هدف این آزمون‌ها، بررسی تفاوت دو به دو میانگین‌ها یا ترکیب خطی از آن‌هاست.در خصوص برتری استفاده از این آزمون‌ها، باید گفت که آزمون‌هایی بیش‌تر از همه مورد استفاده قرار می‌گیرند که نرخ خطای نوع اول را تعدیل کرده و به عبارتی از میزان تورم آن بکاهد. همچنین، انتخاب نوع آزمون مقایسه چندگانه بر اساس دو معیار، برابری حجم نمونه و برابری واریانس انجام می‌گیرد.

انواع آزمون های تعقیبی

حداقل تفاوت معنی‌دار فیشر (LSD)

این آزمون یکی از قدیمی‌ترین و قوی‌ترین آزمون‌ها برای مقایســـه پس از تجربه اســـت. در صـورتی که تعداد میانگین‌ها از سـه تا بیشـتر نباشـد، بهتر اسـت از این آزمون اسـتفاده شـود. اما اگر میانگین‌های مورد مقایسه بیش از سه مورد باشد، بهتر است سایر آزمون‌ها مورد استفاده قرار گیرد.

شفه (Scheffe)

اگر قصد مقایسه میانگین گروه‌های با حجم نابرابر را داریم، روش شفه مناسب‌ترین آزمون است. اما اشکال عمده این روش، محتاطانه یا محافظه‌کارانه بودن آن است. بدین معنی که چون آزمون شفه تمامی ترکیب‌های خطی احتمالی میانگین گروه‌ها را آزمون می‌کند، بنابراین، در این آزمون، صرفاً ترکیب‌های جفتی آزمون نمی‌شوند. در نتیجه آزمون شفه نسبت به سایر آزمون‌ها محافظه‌کارتر است. به همین خاطر، برای این که تفاوت بین میانگین‌ها معنی‌دار باشد، نیازمند میزان بالایی از این تفاوت هستیم. همچنین، آزمون شفه، در مقایسه با آزمون توکی، برای آزمون یک دسته اطلاعات یکسان، فرض صفر را کمتر رد می‌کند. مهم‌ترین مزیت‌های آزمون شفه نسبت به آزمون توکی، امکان کاربرد آن در مورد گروه‌های با حجم‌های نابرابر و عدم حساسیت آن نسبت به انحراف از پیش فرض‌های نرمال بودن توزیع داده‌ها و همگونی واریانس ها می‌باشد. این روش به α بزرگ تری نیاز دارد. به همین دلیل، برخی از پژوهش گران هنگام استفاده از این آزمون، از سطح ۰/۱ به جای ۰/۰۵ استفاده می‌کنند.

نیومن-کلز استودنت شده (S-N-K)

آزمون ساده‌ای است که به جای آزمون دانکن بکار می‌رود. در آزمون نیومن-کلز، ابتدا میانگین‌ها از بالاترین تا پایین‌ترین مقدار مرتب می‌شوند و سپس تفاوت بین هر جفت میانگین محاسبه می‌شود. در نهایت نیز، ارزش مقایســه‌ای که برای هر جفت از میانگین‌ها به طور جداگانه محاسـبه می‌شـود، با هم مقایسه می‌گردند.

توکی (Tukey)

این آزمون که به آزمون‌های کمترین تفاوت‌های راستین توکی و ای. توکی نیز معروف است، از آماره طیف استودنت شده برای تمامی مقایسه‌های جفتی بین گروه‌ها استفاده کرده و نرخ خطای تجربی را با نرخ خطای حاصل از جمع‌آوری برای تمامی مقایسه‌های جفتی هماهنگ می‌کند. پیشنهاد می‌شود زمانی که قصد دارید تعداد زیادی جفت میانگین را آزمون کنید، از آزمون توکی استفاده کنید زیرا این آزمون از آزمون بونفرونی قوی تر است.استفاده از آزمون توکی مستلزم تعیین یک اندازه بحرانی HSD برای داده‌های مورد مطالعه است. هرگاه تفاوت میانگین هر جفت از گروه‌های مورد مطالعه برابر یا بیشتر از اندازه بحرانی HSD باشد، فرض صفر درباره معنی‌دار بودن تفاوت بین آن‌ها رد می‌شود. به عبارت دیگر تفاوت بین آن‌ها معنادار است. در شرایطی که حجم نمونه‌ها مساوی باشد استفاده از روش توکی مناسب است.

آزمون توکی

فرض کنید بعد از انجام تحلیل واریانس و رد فرض برابری میانگین های تیماری، می خواهیم تمام زوجی بین میانگینها را در قالب فرض زیر و به ازای $i \neq j$ مورد ارزیابی قرار دهیم :

$$H_0 : \mu_i = \mu_j$$

$$H_1 : \mu_i \neq \mu_j$$

توکی روشی را برای آزمون فرض هائی که سطح معنادار آنها به ازای اندازه نمونه های یکسان، دقیقا برابر α و به ازای اندازه نمونه های متفاوت حداکثر α است ارائه نمود. روش او را می توان همچنین جهت تعیین فواصل اطمینان برای اختلاف های زوجی بین میانگین ها استفاده کرد. برای این فواصل اطمینان، سطح معنادار توأم به ازای اندازه نمونه های یکسان برابر $(1-\alpha)$۱۰۰ درصد و به ازای اندازه نمونه های متفاوت، حداقل برابر $(1-\alpha)$۱۰۰ درصد است. بعبارت دیگر، روش توکی نرخ خطا در سطح آزمایش یا نرخ خطای جمعی را به ازای سطح α انتخاب شده کنترل می کند. اگر بررسی

میانگین های زوجی مورد نظر باشـــد آنگاه این روش می تواند روش ایده آلی برای داده کاوی محسوب شود.

دانکن (Duncan)

در این آزمون، که به آزمون چند دامنه دانکن نیز معروف اســـت، چنانچه قدر مطلق اختلاف میانگین‌های مورد مقایســه بزرگ تر یا مســاوی $r_\alpha (s_{\bar{x}})$ باشـد، اختلاف بین میانگین‌های مورد مقایسه معنی‌دار است. در این آزمون برای مقایسه هر جفت میانگین، مقدار $r_\alpha (s_{\bar{x}})$ خاص آن مقایسه محاسبه می شود.

والر-دانکن (Waller-Duncan)

این آزمون که به آزمون تی. والر-دانکن نیز معروف اســـت، از رویکرد بیزی (Bayesian) اســـتفاده می‌کند. زمانی که اندازه‌های نمونه با هم برابر نباشـــند، این آزمون از میانگین هارمونیک اندازه نمونه استفاده می‌کند.

کدام روش مقایسه های زوجی را انتخاب کنم؟

مطمئنا یک سوال منطقی که میتوان در این مرحله بیان کرد این است که کدام یک از روشهای ارائه شده را باید استفاده نمود؟ متأسفانه، پاسخ واضحی برای این سؤال وجود ندارد و غالبا در بین آماردانان خبره در مورد عملکرد روشهای فوق اتفاق نظر مشاهده نمی شود. در مطالعه ای خصوص مقایسه چندین روش متفاوت مقایسه های چند گانه به کمک شبیه سازی انجام شده است. در نتایج گزارش شــده، از روش کمترین اختلاف معنادار بعنوان یک روش بســیار اثربخش جهت شــناســایی اختلاف های واقعی بین میانگین هانام برده شـــده اســـت، البته با این شـــرط که از این روش بعد از معنادار بودن تحلیل واریانس به ازای ســـطح معنادار ۵ درصـــد اســـتفاده شـــود. با این وجود، روش کمترین اختلاف معنادار از نرخ خطای جمعی آزمایش برخوردار نیست. از آنجائیکه روش توکی نرخ خطای جمعی را کنترل می کند، اغلب آماردانان ترجیح میدهند از این روش استفاده کنند.

فصل سوم (پروژه)

۳. مقدمه ای بر موضوع داده‌های پروژه:

گندم به عنوان یکی از اصلی‌ترین مواد غذایی بشر و مهمترین محصول زراعی و ماده غذایی در اکثر کشورها، از جایگاه ویژه‌ای برخوردار است و بیشترین سطح کشت را در دنیا دارد.

یکی از مهم ترین عوامل مؤثر در کیفیت گندم یعنی درصد پروتئین موجود در دانه گندم شرایط آب و هوایی است. وقوع یک سرمای نامناسب و یا تغییر ارتفاع از سطح دریا، ریزش بیش از اندازه باران و خشکی آخر دوره رشد دانه می تواند باعث کاهش معنی دار کیفیت گندم تولیدی شود.

انواع کشت گندم :

گندم به دوشیوه دیم و آبی کشت می شود. دیم‌کاری یا کشت دیم نوعی کشاورزی مخصوص مناطق خشک است که در آن تنها بارش‌های آسمانی تأمین‌کنندهٔ آب مورد نیاز کشتزار و زمین کشاورزی هستند.

کود آلی (ارگانیک) : به کودهایی اطلاق میشوند که منشا طبیعی دارند.

در کود آلی می توان هر ماده آلی که به وسیله میکروب‌ها قابل تجزیه باشد به کار برد. اما انواع کودهای آلی مختلف از نظر کیفیت و دوام در خاک و قیمت بسیار متفاوت اند. ارزش اصلی کودهای آلی به علت

تغییرات فیزیکی است که در خاک ایجاد می‌کنند. کودهای آلی عبارتست از : کود حیوانی ، کود گیاهی و یا کود سبز و کود کمپوست .

کودهای بیولوژیک یا زیستی:

کودهای بیولوژیک به مواد حاصل خیزکننده ای گفته میشود که حاوی تعداد کافی از یک یا چند گونه از میکروارگانیسم های مفید خاکزی هستند که ابتداکشت وتکثیر داده می شوند و سپس به همراه نگهدارنده های خاصی به صورت مایع یا خشک و بسته بندی شده، عرضه می شوند..

تفاوت کود آلی با بیولوژیک و معدنی:

بهترین معادل فارسی واژه ارگانیک همان طبیعی است، بنا به همین معنی ساده مشخص می شود هر نوع کودی که به شکل طبیعی تهیه شود یا از معادن استخراج گردد (مثل گچ یا نمکهای سولفاته) به عنوان کود ارگانیک حساب می شود.

در کنار آن محصولاتی مثل کود دامی کمپوست و هیومیکها هم کود ارگانیک حساب می شوند و نوع دیگر کودهای ارگانیک کودهای زنده یا بیولوژیک هستند. یعنی کودهایی که حاوی موجودات زنده می باشند که به اصطلاح بیولوژیک می گویند. پس ما سه نوع کود ارگانیک داریم:

۱) معدنی

۲) آلی

۳) بیولوژیک

البته اگر بخواهیم دقیق تر رده بندی کنیم باید بگوییم کودهای بیولوژیک یک پله از کودهای معدنی و آلی بالاتر هستند چون این محصولات در واقع کود به حساب نمی آیند بلکه کود ساز هستند، محصولات بیولوژیک مانند کودهای حاوی باکتری های مفید در خاک شما شروع به ساخت کود و مواد معدنی و.... می کنند.

۳.۱ طرح مسئله :

در این فصل قصد داریم داده های مربوط به طول خوشه‌های گندم را که از ترکیب ۴ نوع کود مختلف و ۳ نوع بذر مختلف رشد کرده اند با استفاده از آنالیز واریانس بررسی کنیم تا ببینیم که ترکیب بذرها و کودهای مختلف آیا تاثیری بر رشد خوشه های گندم دارد؟ و اگر این ترکیب ها تاثیری بر طول خوشه ها دارد دقیقا کدام یک از ترکیب ها تاثیر بیشتری دارد و باعث ایجاد اختلاف در طول خوشه ها شده است؟برای آنالیز واریانس میتوان از نرم افزار های مختلفی استفاده نمود در این پروژه از نرم افزار SPSS استفاده شده است.

۳.۲. معرفی متغیر ها:

در این پروژه نمونه های ما مربوط به سـه متغیر طول خوشـه، کودهای آلی و شـیمیایی ، و بذر های گندم میباشند.

طول خوشه: طول خوشه متغیر پاسخ ما میباشد که یک متغیر کمی با مقادیر پیوسته است.

انواع کود: این متغیر یکی از عامل های ما در آنالیز واریانس است که دارای ۴ سطح میباشد که یک متغیر کیفی است و قرار است تاثیر سطوح مختلف آن بر متغیر پاسخ(طول خوشه) سنجیده شود تا ببینیم اسـتفاده از کود ها مختلف میتواند روی طول خوشـه ها اثر بگذارد یا خیر.نام سـطوح آن به ترتیب از ۱ تا ۴ به شرح رو به رو میباشد: هیومیکا، هیومیسل ، نیتروکسین ، سوپرنیتروپلاس

انواع بذر گندم دیم: این متغیر نیز همچون نقش متغیر کود را دارد.متغیری اسـت کیفی که دارای ۳ سطح مختلف میباشد.نام هر سطح متغیر بذر گندم به ترتیب از ۱ تا ۳ به شرح رو به رو میباشد: هما ، کوهدشت ، UN

۳.۳. اهداف فصل ۳:

۱. آنالیز واریانس یک راهه با عامل کود
۲. آنالیز واریانس دو راهه با اثر ثابت
۳. آنالیز واریانس دوراهه با اثر تصادفی
۴. آنالیز واریانس دوراهه با اثرات آمیخته
۵. بررس مناسبت مدل

نکته: در هر کدام از مراحل بالا به بررسی آمار توصیفی و آزمون های جفتی برای آن مرحله نیز خواهیم پرداخت.

۳.۴. نکات مهم در قدم اول:

۱. جمع آوری نمونه های متغیر پاسخ از هر گروه حتما باید به صورت تصادفی صورت بگیرد تا نماینده ی مبینی از جمعیت خود باشند.یعنی نمونه گیری بدون دخالت سلیقه ی نمونه گیر و یا عوامل دیگر انجام شده باشد.
۲. نمونه های درون یک گروه نیز باید مستقل از یکدیگر باشند.
۳. نمونه های جمع آوری شده باید از توزیع نرمال پیروی کنند.در غیر این صورت باید انها را به توزیع نرمال تبدیل کرد و یا اینکه برای تحلیل آنها از روش های ناپارامتری استفاده کرد.

۳.۵. آنالیز واریانس یک راهه با عامل کود

در این بخش به بررسی تاثیر عامل کود بر روی طول خوشه های گندم میپردازیم تا بدانیم آیا استفاده از کود های مختلف میتواند بر روی رشد خوشه های گندم اثر بگذارد یا خیر.به این صورت که فرض میکنیم داده ها مربوط به طول خوشه از بین خوشه های رشد یافته از ۴ سطح مختلف کود جمع آوری شده اند. از هر سطح به تعداد ۹ نمونه جمع اوری شده است بنابراین به طور کلی متغیر پاسخ مادارای ۳۶ داده میباشد.

۳.۵.۱ آمار توصیفی

	تعداد	میانگین	انحراف معیار	فاصله اطمینان ٪۹۵		مینیمم	ماکسیمم
				Lower Bound	Upper Bound		
هیومیکا	9	8.5867	.27798	8.3730	8.8003	8.27	8.97
هیومیسل	9	8.9422	.46054	8.5882	9.2962	8.47	9.58
نیتروکسین	9	9.3144	.23212	9.1360	9.4929	9.06	9.66
سوپرنیتروپلاس	9	8.9144	.60040	8.4529	9.3760	8.37	9.75
کل	36	8.9394	.47872	8.7775	9.1014	8.27	9.75

تفسیر: در جدول فوق برای هر سطح میانگین ، انحراف معیار ، مینیمم داده، ماکسیمم داده و یک فاصل اطمینان برای میانگین هر سطح محاسبه شده است.همانطور که در ستون میانگین ها مشاهده میکنیم میانگین طول خوشه های حاصل از کود سوپرنیتروپلاس ازمحصول بقیه کود ها تفاوت زیادی دارد.و در کل با اطمینان ٪۹۵ میتوان نتیجه گرفت طول خوشه های حاصل از این ۴ نوع کود در بازه (8.77,9.101) قرار میگیرد.

۳.۵.۲ جدول آنالیز واریانس

	مجموع مربعات خطا	درجه ازادی	میانگین مربعات خطا	اماره F	Sig.
خطای بین گروهی	2.391	3	.797	4.531	.009
خطای درون گروهی	5.630	32	.176		
کل	8.021	35			

تفسیر: همانطور که در جدول فوق مشاهده میکنیم مقدار۰.۰۵ > Sig=0.009 است در نتیجه اماره F در سطح۰.۰۵ معنا دار است یعنی میانگین سطوح مختلف اختلاف معنا داری با یکدیگر دارند.میتوان

اینطور نتیجه گرفت که خوشه های حاصل از کود های مختلف دارای میانگین طوا های برابری نیستند و عامل کود یک عامل تاثیر گذار بر روی میانگین طول خوشه های رشد یافته از آنهاست.

۳.۵.۳آزمون تعقیبی Tukey

برای اینکه بدانیم دقیقا کدام یک از سطوح تفاوت معنا داری بین میانگین ها ایجاد میکند از ازمون توکی استفاده میکنیم.

(I) ترکیب کودی بیولوژیک و آلی	(J) ترکیب کودی بیولوژیک و آلی	(I-J) تفاوت میانگین	انحراف معیار	Sig.
هیومیکا	هیومیسل	-.35556	.19773	.293
	نیتروکسین	-.72778*	.19773	.004
	سوپرنیتروپلاس	-.32778	.19773	.362
هیومیسل	هیومیکا	.35556	.19773	.293
	نیتروکسین	-.37222	.19773	.256
	سوپرنیتروپلاس	.02778	.19773	.999
نیتروکسین	هیومیکا	.72778*	.19773	.004
	هیومیسل	.37222	.19773	.256
	سوپرنیتروپلاس	.40000	.19773	.201
سوپرنیتروپلاس	هیومیکا	.32778	.19773	.362
	هیومیسل	-.02778	.19773	.999
	نیتروکسین	-.40000	.19773	.201

تفسیر: همانطور که مقدار sig را برای سطوح مختلف میبینیم تنها سطح نیتروکسین و سطح هیومیکا هستند که مقدار sig آنها معنا دار است یعنی میانگین طول خوشه های حاصل از این دو سطح با یکدیگر اختلاف معنا داری دارد و بقیه سطوح تقریبا خوشه هایی با طول یکسان را رشد میدهند و در صورت نیاز میتوان آنها را به جای یکدیگرو به کار برد.

نتیجه: در این بخش به آنالیز واریانس یک راهه پرداختیم و تاثیر عامل کود را بر طول خوشه ها بررسی نمودیم و متوجه شدیم که استفاده از کود های مختلف میتواند بر طول خوشه ها تاثیر بگذارد سپس برای اینکه بدانیم کدام کود ها باعث ایجاد این اختلاف شده اند ازمون توکی را انجام دادیم و متوجه شدیم مود هیومیکا و کود نیتروکسین تفاوت ها معنا داری در میانگین طول خوشه ها ایجاد میکنند.

نکته: در مدل انالیز واریانس یک راهه طرح اثرات ثابت و اثرات تصادفی پاسخ های یکسانی را به ما ارائه میدهند بنا براین برای انالیز واریانس یک راهه اثر ثابت و تصادفی در نظر نگرفته ایم.

۳/۶. آنالیز واریانس دوراهه با اثرات ثابت

در این بخش میخواهیم تاثیر عامل کود و عامل بذر گندم را به طور همزمان بر طول خوشه ها بررسی کنیم قابل ذکر است که در این بخش فرض میکنیم هر دو عامل ما دارای اثرات ثابت هستند یعنی سطوح بررسی شده در پروژه تنها سطوح موجود در جامعه ی هر عامل بوده است.گروه های ما هر کدام دارای ۳ مشاهده هستند که هر گروه شامل ترکیب یک نوع کود و یک نوع بذر است.و در کل ۳۶ مشاهده داریم بنابراین ۱۲ گروه در اختیار داریم که باید با یکدیگر مقایسه شوند.

۳.۶.۱ آمار توصیفی

گندم دیم	ترکیب کودی بیولوژیک و آلی	میانگین	انحراف معیار	تعداد
هما	هیومیکا	8.5300	.03606	3
	هیوميسل	9.5300	.04583	3
	نیتروکسین	9.1000	.04583	3
	سوپرنیتروپلاس	8.6433	.07506	3
کوهدشت	هیومیکا	8.9300	.04000	3
	هیوميسل	8.5000	.03606	3
	نیتروکسین	9.2333	.03786	3
	سوپرنیتروپلاس	9.7000	.04583	3
UN	هیومیکا	8.3000	.03606	3
	هیوميسل	8.7967	.04041	3
	نیتروکسین	9.6100	.04583	3
	سوپرنیتروپلاس	8.4000	.03606	3

تفسیر: جدول بالا میانگین و انحراف معیار را برای تمام گروه ها محاسبه کرده است همانطور که در جدول میبینیم بیشترین میانگین طول خوشه متعلق به گوهیست که از ترکیب کود سوپرنیتروپلاس و گندم کوهدشت به عمل به امده است.

۳.۶.۲ جدول آنالیز واریانس

	مجموع مربعات خطا	درجه آزادی	میانگین مربعات خطا	آماره F	Sig.
اثر اصلی گندم	.595	2	.297	149.257	.000
اثر اصلی کود	2.391	3	.797	400.232	.000
اثر متقابل	4.987	6	.831	417.361	.000
خطا	.048	24	.002		
کل	2884.913	36			

تفسیر: جدول فوق نشان میدهد میزان Sig برای اثر متقابل گندم و کود برابر ۰ است که کوچکتر از ۰.۰۵ است برنابراین نتیجه میگیریم اثر متقابل آنها معنا دار است.یعنی استفاده از کود و گندم های مختلف باعث ایجاد اختلاف در میانگین طول خوشه ها میشود.همچنین اثرات اصلی هر دو عامل نیز معنا دار است حال میتوانیم بررسی کنیم تا ببینیم دقیقا کدام سطوح در هر دو عامل باعث ایجاد تفاوت معنا دار در میانگین ها شده است.

۳.۶.۳. آزمون تعقیبی Tukey برای هر دو عامل

گندم دیم (I)	گندم دیم (J)	تفاوت میانگین	انحراف معیار	Sig.
هما	کوهدشت	-.1400*	.01822	.000
	UN	.1742*	.01822	.000
کوهدشت	هما	.1400*	.01822	.000
	UN	.3142*	.01822	.000
UN	هما	-.1742*	.01822	.000
	کوهدشت	-.3142*	.01822	.000

تفسیر: جدول فوق آزمون توکی برای عامل گندم است همانطور که در جدول میبینیم مقدار sig برای تمام سطوح بذر گندم برابر 0 است و معنا دار هستند یعنی تمام سطوح گندم میانگین های یکسانی ندارند و اختلاف معناداری بایکدیگر دارند.

ترکیب کودی بیولوژیک و آلی (I)	ترکیب کودی بیولوژیک و آلی (J)	اختلاف میانگین	انحراف معیار	Sig.
هیومیکا	هیومیسل	-.3556*	.02104	.000
	نیتروکسین	-.7278*	.02104	.000
	سوپرنیتروپلاس	-.3278*	.02104	.000
هیومیسل	هیومیکا	.3556*	.02104	.000
	نیتروکسین	-.3722*	.02104	.000
	سوپرنیتروپلاس	.0278	.02104	.559
نیتروکسین	هیومیکا	.7278*	.02104	.000
	هیومیسل	.3722*	.02104	.000
	سوپرنیتروپلاس	.4000*	.02104	.000
سوپرنیتروپلاس	هیومیکا	.3278*	.02104	.000
	هیومیسل	-.0278	.02104	.559
	نیتروکسین	-.4000*	.02104	.000

تفسیر: جدول فوق آزمون توکی برای عامل کود با در نظر گرفتن ثابت بودن تاثیر عامل گندم است.همانطور که میبینیم همه ی سطوح با یکدیگر تفاوت معنا دار دارند بجز سطوح هیومیسل و سوپرنیترو پلاس چرا که sig برای این دو سطح بیشتر از ۰.۰۵ است و این یعنی طول خوشه های حاصل از این دو کود اختلاف معناداری با یکدیگر ندارند.

نتیجه:

در این بخش متوجه شدیم که اگر عامل های ما با اثرات ثابت باشند هر دو عامل گندم و کود عامل های تاثیر گذاری بر طول خوشه هستند و استفاده از سطوح مختلف آنها میتواند در میانگین طول خوشه ها تفاوت معنا داری ایجاد کند با استفاده از آزمون توکی دست یافیتم که استفاده از گندم کوهدشت یا هما یا UN میتواند تفاوت معنا داری در میانیگن طول خوشه ها ایجاد کند و در کود ها نیز استفاده از کود هیومیسل به جای سوپرنیترو پلاس میتواند تفاوت معنا داری ایجاد کند.

۳.۷. آنالیز واریانس دوراهه با اثرات تصادفی

در این بخش قصد داریم تاثیر عامل گندم و عامل کود را به طور همزمان بر روی طول خوشه بررسی کنیم اما در این بخش فرض میکنیم هر دوعامل ما دارای اثرات تصادفی میباشند.یعنی سطوح مورد بررسی به تصادف از بین تمام سطوح موجود در جامعه انتخاب شده است.مانند بخش قبل ۱۲ گروه داریم که در هر گروه ۳ مشاهده قرا دارد و مجموع تمام مشاهدات ۳۶ میباشد.

نکته ۱: قسمت امار توصیفی این بخش شبیه به بخش اثرات ثابت است بنابراین از تکرار آن صرف نظر میکنی.

نکته۲: در طرح اثرات تصادفی امکان انجام ازمون ها تعقیبی وجود ندارد.بنابراین برای این بخش آزمون توکی را انجام نمیدهیم.

۳.۷.۱ جدول آنالیز واریانس

		مجموع مربعات خطا	درجه ازادی	میانگین مربعات خطا	آماره F	Sig.
اثر اصلی گندم	Hypothesis	.595	2	.297	.358	.713
	خطا	4.987	6	.831b		
اثر اصلی کود	Hypothesis	2.391	3	.797	.959	.470
	خطا	4.987	6	.831b		
اثر متقابل	Hypothesis	4.987	6	.831	417.361	.000
	خطا	.048	24	.002c		

تفسیر: جدول فوق جدول آنالیز واریانس برای اثرات تصادفی است همانطو که میبینیم اثر متقابل معنا دار شده است یعین دو عامل گندم و کود عامل های موثر در ایجاد اختلاف بین میانگین گروه ها هستند.اما اثرات اصلی آنها معنا دار نشده است.

۳.۸.آنالیز واریانس دوراهه با اثرات آمیخته

در این بخش نیز میخواهیم تاثیر عامل کود و عامل گندم را به طور همزمان بر طول خوشه ها بررسی کنیم اما در این بخش عامل گندم را ثابت و عامل کود را تصادفی در نظر گرفته ایم.بناراین مدل ما انالیز واریانس دوراهه با اثرات آمیخته میباشد.مانند دو بخش قبل دارای ۱۲ گروه است که بایکدیگر مقایسه میکنیم و در هر گروه ۳ مشاهده وجود دارد.

نکته ۱: آمار توصیفی این بخش نیز مانند اثرات ثابت است بنابراین از پرداختن به آآمر توصیفی در این بخش صرف نظر میکنیم.

نکته ۲: در اینجا یکی از عامل ها یعنی عامل کود اثر تصادفی است بنابراین برای عامل کود نمیتوانیم ازمون تعقیبی انجام دهیم.

۳.۸.۱ جدول آنالیز واریانس

		مجموع مربعات خطا	درجه آزادی	میانگین مربعات خطا	آماره F	Sig.
اثر اصلی گندم	Hypothesis	.595	2	.297	.358	.713
	خطا	4.987	6	.831[b]		
اثر اصلی کود	Hypothesis	2.391	3	.797	.959	.470
	خطا	4.987	6	.831[b]		
اثر متقابل	Hypothesis	4.987	6	.831	417.361	.000
	خطا	.048	24	.002[c]		

تفسیر: جدول فوق جدول آنالیز واریانس برای اثرات امیخته است همانطور که میبینیم مقادیر محاسبه شده شبیه به مقادیر جدول آنالیز واریانس اثر تصادفی میباشد.این دو مدل در محاسبات یکان هستند و هر دو تصادفی در نظر گرفته میشوند اما در فرضیات بایکدیگر تفاوت دارند.

همانطور که میبینیم در این مدل نیز اثر متقابل معنا دار است و تفاوت بین میانگین گروه ها را نشان میدهد اما اثرات اصلی معنا دار نیستند.

۳.۸.۲. آزمون Tukey برای عامل گندم

گندم دیم (I)	گندم دیم (J)	تفاوت میانگین	انحراف معیار	Sig.
هما	کوهدشت	-.1400*	.01822	.000
	UN	.1742*	.01822	.000
کوهدشت	هما	.1400*	.01822	.000
	UN	.3142*	.01822	.000
UN	هما	-.1742*	.01822	.000
	کوهدشت	-.3142*	.01822	.000

تفسیر: جدول فوق آزمون توکی برای سطوح عامل گندم که عامل ثابت ما بود میباشد. همانطور که میبینیم تمام سطوح گندم با یکدیگر اختلاف معنادار دارند.

۳.۹ بررسی نرمال بودن توزیع نمونه ها

	گروه کود	کولموگروف اسمیرنو		
		Statistic	df	Sig.
طول خوشه cm	هیومیکا	.196	9	.200*
	هیومیسل	.254	9	.096
	نیتروکسین	.259	9	.082
	سوپرنیترو پلاس	.294	9	.024

تفسیر : با توجه به آزمون کولموگروف میتوانیم نتیجه بگیریم که توزیع نمونه های مربوط به کود هیومیکا و هیومیسل و نیتروکسین دارای توزیع نرمال هستند و فقط نمونه های کود سوپرنیتروپلاس فرض نرمال بودن را رد میکند.

	گروهگندم	کولموگروف اسمیرنوو		
		Statistic	df	Sig.
طول خوشه cm	هما	.210	12	.149
	کوهدشت	.143	12	.200*
	UN	.234	12	.069

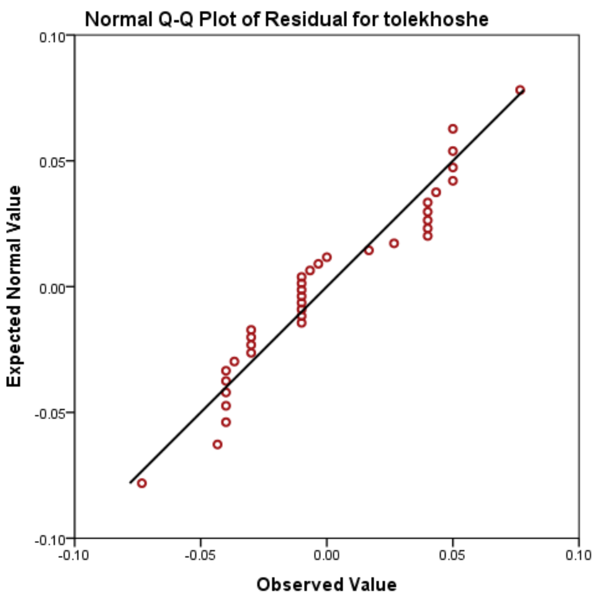
تفسیر : با توجه به آزمون کولموگروف برای داده های مربوط به سطوح گندم میتوان نتیجه گرفت که تمام مشاهدات در این سه سطح دارای توزیع نرمال میباشند چرا که Sig آنها بزرگتر از ۰.۰۵ است

۳.۱۰. بررسی مناسبت مدل

۳.۱۰.۱ محاسبه مانده ها

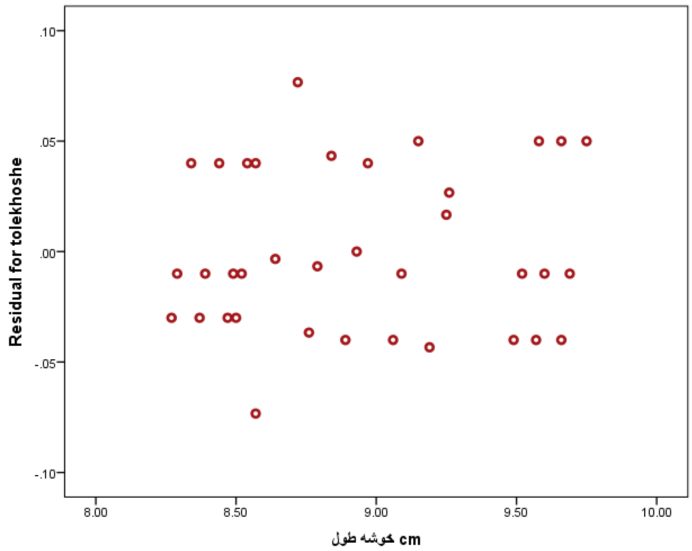
برای این بخش نیاز داریم تا مانده ها را محاسبه کنیم که روش محاسبه ی آن را در نرم افزار spss در ابتدای فصل سوم صفحه ۶۰ توضیح داده ایم.از انجا که مقدار مانده ها برای هر سه مدل انالیز واریانس دوراهه ثابت ،تصادفی و آمیخته برابر است تنها یک بار مناسبت مدل را بررسی میکنیم.

۳.۱۰.۲. بررسی نرمال بودن مانده ها



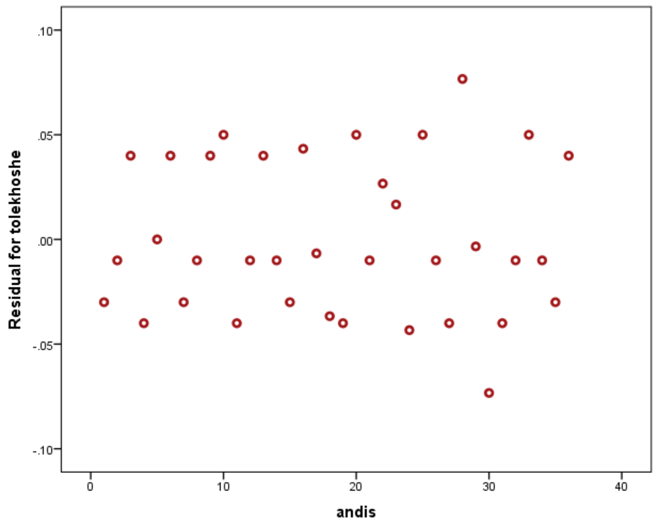
تفسیر: باتوجه به نمودار فوق میتوانیم نتیجه بگیریم که توزیع مانده های ما نرمال میباشد چرا که تقریبا نقاط روی خط نیم ساز که چندک توزیع نرمال است قرار دارند.

۳.۱۰.۳. بررسی همگونی واریانس



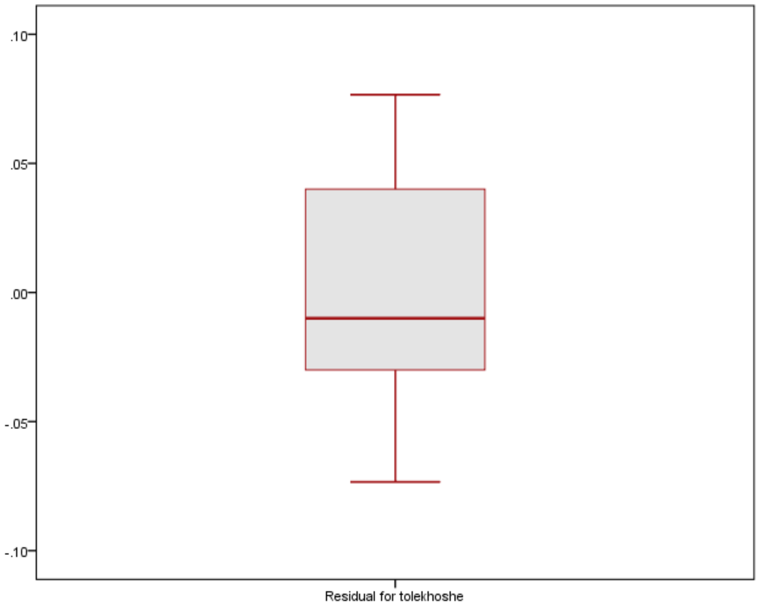
تفسیر: برای بررسی همگونی واریانس مانده ها از نمودار پراکنش استفاده نمودیم بدین گونه که مانده ها را در مقابل مقادیر برازش داده شده رسم کردیدم و انتظار داریم که هول صفر پراکنده باشند و حالت قیفی شکل نداشته باشند بلکه به صورت ثابت مثل یک مستطیل پراکنده شده باشند. همانگونه که در تصویر بالا میبینیم نقاط به طور ثابتی حول صفر پراکنده شده اند بنابراین فرض همگونی واریانس را برای مانده ها میپذیریم.

۳.۱۰.۴. بررسی فرض استقلال مانده ها



تفسیر: برای بررسی فرض استقلال، مانده ها را در مقابل اندیس رسم نمودیم و انتظار داریم حول صفر به طور ثابت پرا کنده شده باشند که همینطور نیز هست.بنا براین فرض استقلال مانده هارا میپذیریم.

۳.۱۰.۵.بررسی عدم حضور مشاهده پرت در مانده ها



تفسیر : برای بررسی این فرض ازنمودار جعبه ای مانده ها استفاده نمودیم همانطور که مشاهده میکنیم نمودار جعبه ای هیچ داده پرتی را به ما نشان نمیدهد پس فرض عدم حضور داده پرت را میپذیریم.

۳.۱۱. جدول داده ها

	G	K	tolekhoshe
1	1	2	8.50
2	1	2	8.52
3	1	2	8.57
4	2	2	8.89
5	2	2	8.93
6	2	2	8.97
7	3	2	8.27
8	3	2	8.29
9	3	2	8.34
10	1	3	9.58
11	1	3	9.49
12	1	3	9.52
13	2	3	8.54
14	2	3	8.49
15	2	3	8.47
16	3	3	8.84
17	3	3	8.79
18	3	3	8.76
19	1	4	9.06
20	1	4	9.15
21	1	4	9.09
22	2	4	9.26
23	2	4	9.25
24	2	4	9.19
25	3	4	9.66
26	3	4	9.60
27	3	4	9.57
28	1	5	8.72
29	1	5	8.64
30	1	5	8.57
31	2	5	9.66
32	2	5	9.69
33	2	5	9.75
34	3	5	8.39
35	3	5	8.37
36	3	5	8.44

متغیر G : عامل گندم است که دارای سه سطح به ترتیب هما ، کوهدشت و UN است.

متغیر K : عامل کود است که دارای ۴ سطح به ترتیب هیومیکا ، هیومیسل ، نیتروکسین و سوپر نیترو پلاس.

متغیر tolekhoshe : متغیر پاسخ است که مقادیر کمی را در بر دارد.

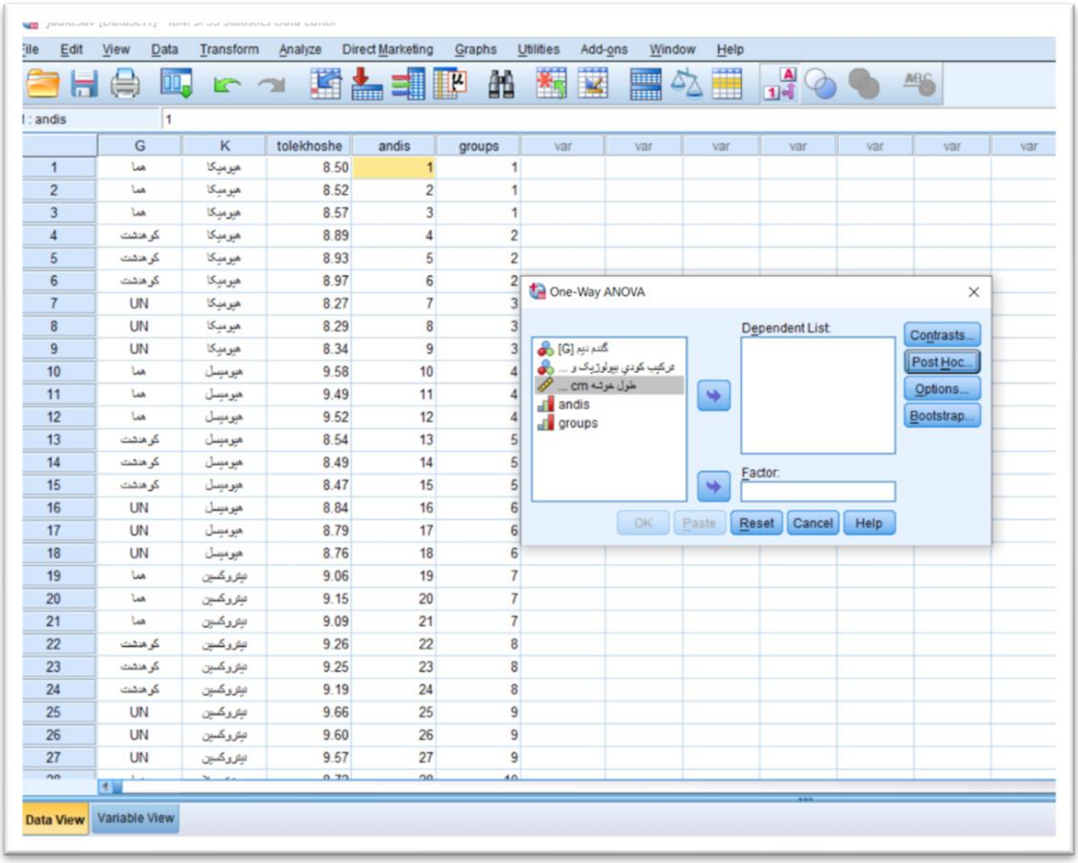
"پیوست"

. معرفی مسیر های مورد استفاده در نرم افزار SPSS

مسیر آنالیز واریانس یک راهه

از مسیر زیر میتوان محاسبات مربوط به آنالیز واریانس را انجام داد:

Analyze >   Comper Means   >   One-Way Anova



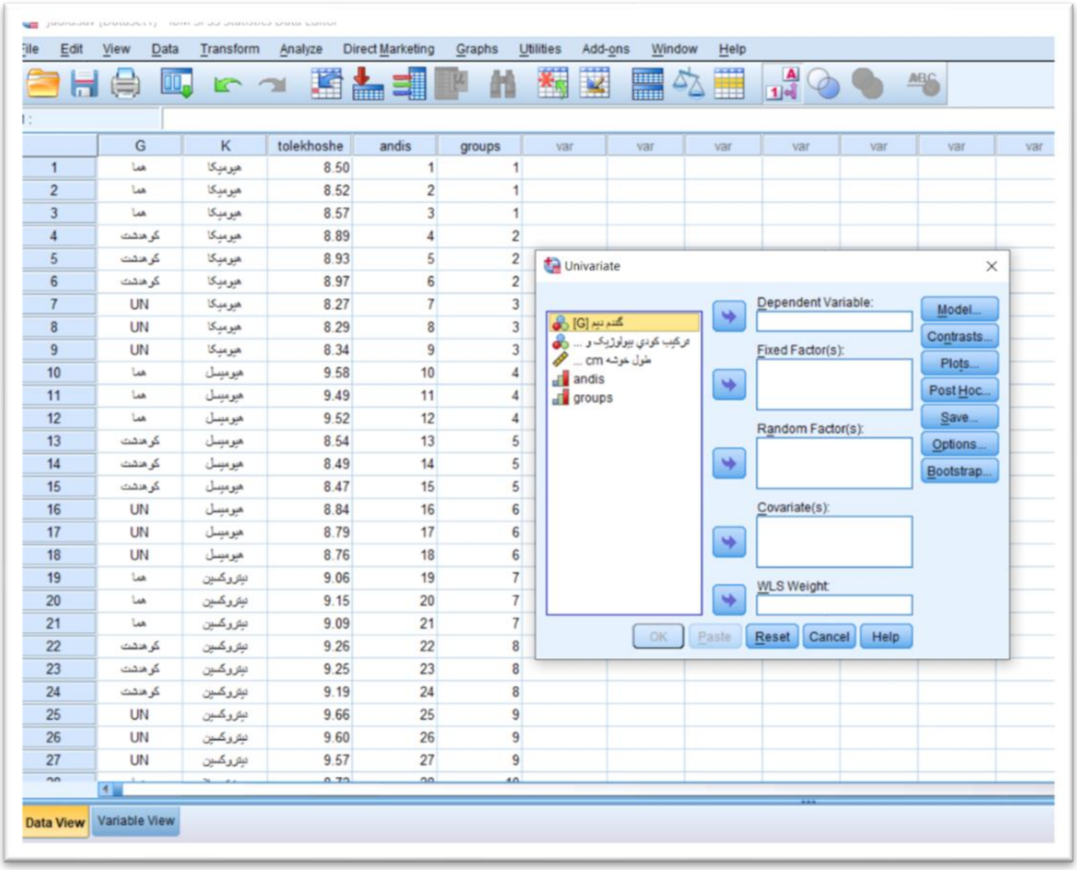
نکته ۱: متغیر پاسخ را در قسمت Dependent List وارد میکنیم و متغیر عامل را درقسمت Factor قرار میدهیم.

نکته ۲: برای از انجام آزمون های تعقیبی از دکمه Post Hoc کمک میگیرم و آزمون مورد نظر را علامت میزنیم.

مسیر آنالیز واریانس دوراهه

برای انجام آنالیز واریانس دوراهه از مسیر زیر کمک می‌گیریم:

Analyze > General Linear model > Univariate



با توجه به شکل بالا موارد زیر را اجرا می‌کنیم:

– متغیر پاسخ را در قسمت Dependent Variable **وارد می‌کنیم.**

– اگر مدل ما ، مدل آنالیزواریانس دوراهه با اثرات‌ثابت **باشد لازم است هردو عامل را در قسمت** Fixed factor **وارد کنیم.**

– اگر مدل ما ، مدل آنالیزواریانس دوراهه با اثرات‌تصادفی **باشد لازم است هردو عامل را در قسمت** Random Factor **وارد کنیم.**

- اگر مدل ما ، مدل آنالیز واریانس دوراهه با اثرات‌آمیخته باشد عاملی را که سطوح آن به تصادف انتخاب شده‌اند در قسمت Random factor و عاملی را که سطوح آن ثابت است در قسمت Fixed factor وارد می‌کنیم.

- با استفاده از دکمه Post Hoc آزمون‌تعقیبی موردنظر را انتخاب می‌کنیم. مانند آزمون LSD، Duncan، Tukey و... . توجه داشته‌باشید آزمون‌های‌تعقیبی تنها مناسب برای عامل‌هایی با اثرثابت انجام می‌شوند.

*- با استفاده از دکمه Save و انتخاب گزینه‌ی Unstandardized از قسمت Residual می‌توانیم مقادیر مانده‌ها را محاسبه کنیم.

- با استفاده از فشردن دکمه‌ی option سپس انتقال عامل‌ها به باکس سمت راست و سپس انتخاب گزینه‌ی Descriptive می‌توانیم میانگین و واریانس هرگروه را محاسبه کنیم و به‌عنوان آمارتوصیفی از آن‌ها استفاده کنیم.

- در آخر برای گرفتن خروجی دکمه ok را می‌زنیم.

مسیر بررسی مناسبت مدل

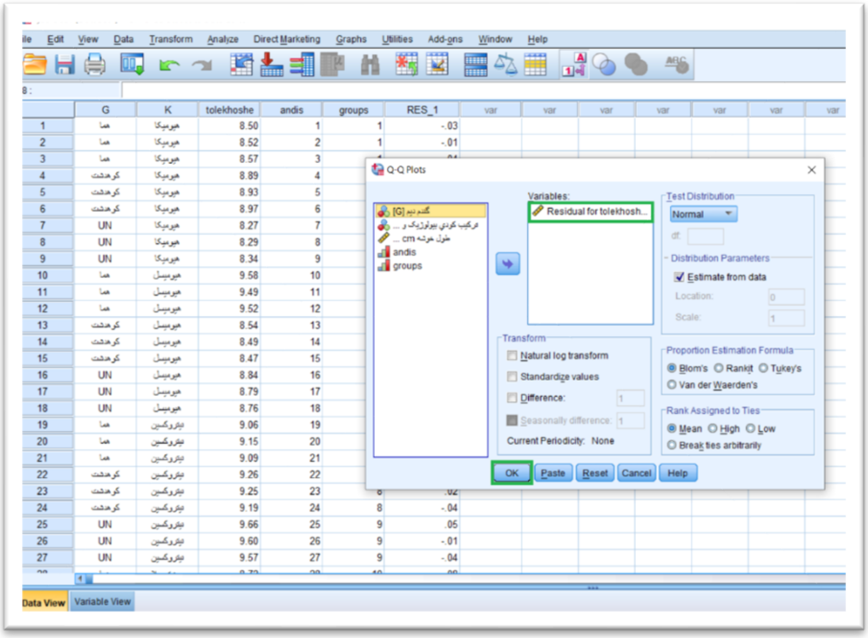
با استفاده از نرم‌افزار spss لازم است فرضیات زیر بررسی شوند:

۱. نرمال بودن توزیع مانده‌ها که ما از طریق نمودار QQ-plot بررسی خواهیم‌کرد.
۲. همگونی واریانس مانده‌ها از طریق رسم مانده‌ها در مقابل مقادیر برازش یافته بررسی میشود.
۳. مستقل بودن مانده از طریق رسم نمودار مانده ها در مقابل اندیس تعداد (N)بررسی خواهد شد.
۴. عدم حضور داده‌ی پرت در مانده‌ها که با استفاده از نمودارجعبه‌ای بررسی خواهیم‌کرد.

رسم نمودار QQ-plots

از طریق مسیر زیر نمودار QQ-plots را رسم می‌کنیم :

Analyze > Descriptive Statistics > QQ-plots



– متغیرمانده‌ها را مانند تصویر بالا در کادر موردنظر وارد می‌کنیم و سپس دکمه ok را می‌زنیم.

رسم نمودار پراکنش برای همگونی واریانس

از مسیر زیر نمودار پراکنش را رسم میکنیم:

Graphs > Legacy Dialogs > Scatter

متغیرمانده‌ها را در کادرY و متغیرمقادیربرازش‌یافته را در کادر X وارد می‌کنیم و oK را می‌زنیم تا نمودار پراکنش رسم شود.

رسم نمودار پراکنش برای استقلال مانده ها

مانند مرحله قبل از مسیر زیر نمودارپراکنش را رسم می‌کنیم:

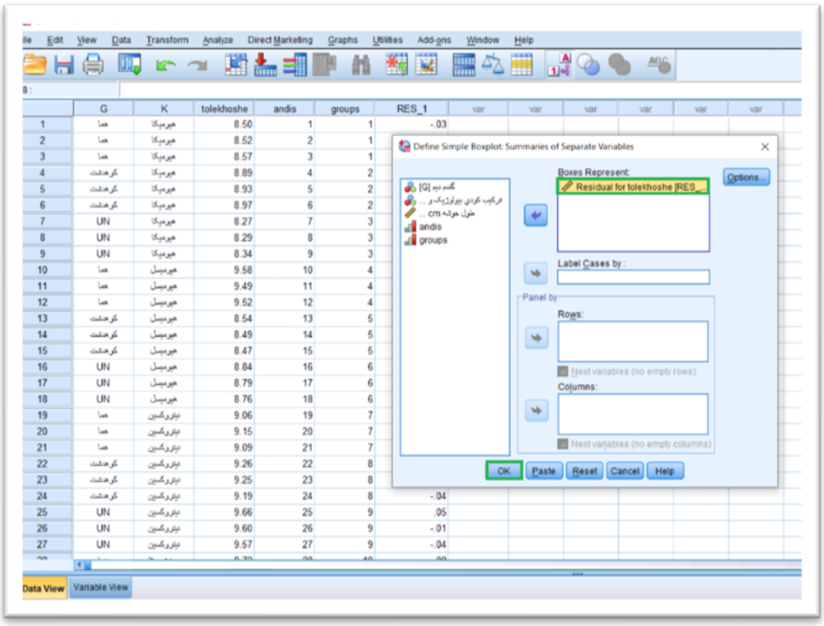
Graphs > Legacy Dialogs > Scatter

با این تفاوت که در ستون Xها مقادیر برازش‌یافته را قرار نمی‌دهیم.لازم است در ستون داده های خود یک ستون به نام andis ایجاد کنیم و از ۱ تا N شماره‌گذاری کنیم. پس از این کار در قسمت Y مانده‌ها را وارد می‌کنیم و در قسمت X متغیر Andis را وارد می‌کنیم و ok را میزنیم تا نمودار پراکنش رسم شود.

رسم نمودار جعبه ای

از مسیر زیر میتوانیم نمودار جعبه ای را به منظور تشخیص مشاهده پرت رسم کنیم:

Graphs > Legacy Dialogs > Box plots > Summaries of separate variables



طبق تصویر بالا متغیر مانده ها را وارد میکنیم و سپس دکمه Ok را میزنیم تا نمودار جعبه ای رسم شود.

منابع :

[1]کابررا ، خ. ، مک‌دوگال ، ا. ،"مشاورهٔ آماری "، چاپ اول ، انتشارات مرکز نشر دانشگاهی سال انتشار ۱۳۹۴

[2] تیم علمی دایان (۱۳۹۹) . راهنمای جامع مصرف کود گندم و جو با هدف افزایش عملکرد
. https://www.agrodayan.com/fertilizer-application-to-increase-yield-of-wheat-and-barley/

[3] آمار پیشرو (۱۳۹۹) . آنالیز واریانس دوطرفه چیست؟-اجرای آن در SPSS با مثالی کاربردی .
https://amarpishro.com/parametric-inferential-statistics/two-way-anova/

[4] Wikipedia (Last edited time (2021/11/21)). Analysis of variance.
https://en.m.wikipedia.org/wiki/Analysis_of_variance

[5] Montgomery, Douglas C. ,"Design and analysis of experiments", Eighth edition ,1943

[6] Rancher, "Methods of Multivariate Analysis", Second Edition, Wiley & Sons, J, Inc., Alvin C., 1934

[7] Alvin C. Rancher and G. Bruce Schaller, "LINEAR MODELS INSTATISTICS", Second Edition, Wiley & Sons, Inc., Alvin C., 1934