Subject : (   )   Date : ....

ضریب همبستگی:

محاسبه ضریب همبستگی   (قیمت، هزینه، فروش) F = سود فعلی است

رابطه علی یک سویه $y = F(x_1, x_2, \ldots x_n)$

وابسته   مستقل

متغیر مستقل تعدادی در برون داد است

اگر بخواهیم رابطه دو متغیر را بسنجیم که رابطه علت و معلولی نداشته باشند

ضریب همبستگی، شدت و جهت همبستگی دو متغیر را با هم می‌سنجد.

وابسته و مستقل بررسی اثر از کلمه اثر استفاده می‌کنیم،

وقتی از correlation صحبت می‌کنیم از کلمه رابطه استفاده می‌شود.

هر موقع جهت اثرگذاری مهم نباشد از ضریب همبستگی استفاده می‌کنیم

هر موقع جهت اثرگذاری مهم باشد از تحلیل‌های دیگری مانند رگرسیون استفاده می‌کنیم. ضریب همبستگی همیشه بین ۱- و ۱+ نوسان دارد، صفر تا ۱-

اثر معکوس در متغیره، بین ۱- تا ۰ یعنی ۱۰۰٪ ارتباط عکس وجود دارد.

اگر بین ۱ تا ۰ باشند یعنی ارتباط مستقیم و هم جهت دارند. عدد ضریب همبستگی شدت و علامت جهت ارتباط را نشان می‌دهد

به صرف محاسبه ضریب همبستگی تفسیری اتفاق نمی‌افتد باید بررسی کرد معنی داری ضریب همبستگی محاسبه (مورد آزمون) و تایید آماری قرار گیرد اگر ضریب محاسبه شده معنی دار نشود، همبستگی کاذب وجود دارد.

ضریب همبستگی تا چه اندازه قابل قبول است و یا چه اندازه قابل قبول است

انواع ضریب همبستگی

۱- ضریب همبستگی داده‌های مقیاس فاصله‌ای یا کمی ← آمار پارامتریک

ضریب همبستگی گشتاوری پیرسون

۲- دیگر داده‌ها با مقیاس کیفی یا دارای مقیاس رتبه‌ای باشند در این صورت ← آمار ناپارامتریک ← ضریب همبستگی رتبه‌ای اسپیرمن

ضریب همبستگی عددی است بین ۱+ و ۱- که شدت و جهت رابطه خطی بین دو متغیر عددی را نشان می‌دهد. نزدیک ۱+ یعنی رابطه مستقیم و قوی، نزدیک ۱- یعنی رابطه معکوس قوی و نزدیک صفر یعنی تقریباً هیچ رابطه خطی دیده نمی‌شود. وقتی دو متغیر با هم بالا و پایین می‌روند مثل قد و وزن می‌گوییم همبستگی مثبت دارند و ضریب همبستگی بین ۰ و ۱ است.

وقتی یکی زیاد می‌شود و دیگری کم می‌شود (مثلاً افسردگی و شادکامی) همبستگی منفی داریم و ضریب بین ۰ و ۱- قرار می‌گیرد

اگر ضریب همبستگی نزدیک صفر باشد یعنی الگوی خطی مشخصی بین دو متغیر دیده نمی‌شود.

فرمول و مثال محاسبه ضریب همبستگی

$$r = \frac{\Sigma xy - n\bar{x}\bar{y}}{\sqrt{\Sigma x^2 - n\bar{x}^2}\sqrt{\Sigma y^2 - n\bar{y}^2}}$$

$\bar{x}$	$\bar{y}$	x	y	xy	x^2	y^2
		۲	۵	۱۰	۴	۲۵
		۳	۶	۱۸	۹	۳۶
		۴	۷	۲۸	۱۶	۴۹
		۵	۸	۴۰	۲۵	۶۴
		۹	۱۱	۹۹	۸۱	۱۲۱
		۲۴	۳۷	۱۹۵	۱۳۵	۲۹۵

$$\bar{y} = \frac{۳۷}{۵} = ۷,۴$$

$$\bar{x} = \frac{۲۴}{۵} = ۴,۴$$

$$r = \frac{۱۹۵ - ۵(۴,۴)(۷,۴)}{\sqrt{۱۳۵ - (۵)(۴,۴)^2}\sqrt{۲۹۵ - ۵(۷,۴)^2}} = \frac{۱۹۵ - ۱۷۰,۲}{\sqrt{\underset{۲۹,۲}{۱۳۵ - ۱۰۵,۸}}\sqrt{۲۹۵ - ۲۷۳,۸}}$$

$$= \frac{۱۹۵ - ۱۷۰,۲}{۵,۴(۴,۶)} = \frac{۲۴,۸}{۲۴,۸۴} = ۰$$

همبستگی ندارند

فرض کنیم جواب بشه ۰٫۱۹

حدود ۲٫ که می‌شود ۲٪ همبستگی وجود دارد

معنی‌داری:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{۰٫۱۹}{\sqrt{\frac{۱-(۰٫۱۹)^2}{۵-۲}}} = ۰٫۳۷$$

توزیع t با n-۲ درجه آزادی

— *Arman*

فرض صفر تایید می‌شود همبستگی کاذب است پس فرض ما رد می‌شود.

Subject : (   )   Date : ....

مثال کتاب

x [illegible]	y [illegible]	xy	$x^۲$	$y^۲$
۴	۵	۲۰	۱۶	۲۵
۲	۲	۴	۴	۴
۴	۶	۲۴	۱۶	۳۶
۲	۱	۲	۴	۱
۶	۴	۲۴	۳۶	۱۶
		۷۴	۷۶	۸۲

$\bar{x} = \frac{۱۸}{۵} = ۳,۶$

$\bar{y} = \frac{۱۸}{۵} = ۳,۶$

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^۲ - n\bar{x}^2}\sqrt{\sum y^۲ - n\bar{y}^2}} = \frac{۷۴ - (۵)(۳,۶)(۳,۶)}{\sqrt{۷۶ - ۵(۱۲,۹۶)}\sqrt{۸۲ - ۵(۳,۶)}}$$

۱۲,۹۶

$$= \frac{۷۴ - ۶۴,۸}{\sqrt{۷۶ - ۶۴,۸}\sqrt{۸۲ - ۶۴,۸}} = \frac{۹,۲}{\sqrt{۱۱,۲}\,\sqrt{۱۷,۲}} = \frac{۹,۲}{۳,۳۵(۴,۱۵)} = \frac{۹,۲}{۱۳,۹۰}$$

$= ۰,۶۶$ این دو تا پس از هم مستقلند و ضریب همبستگی نسبتاً بالایی دارند یعنی ۶۶ درصد همپوشانی دارند.

آیا دو متغیر تصادفی x و y همبسته‌اند یا خیر؟ به عبارت دیگر آیا ضریب همبستگی جامعه برابر صفر است یا نه؟

$H_0: \rho = ۰$

$H_1: \rho \neq ۰$

$$t = \frac{r}{\sqrt{\frac{۱ - r^۲}{n - ۲}}} = \frac{۰,۶۶}{\sqrt{\frac{۱ - (۰,۶۶^۲)}{۵ - ۲}}} = \frac{۰,۶۶}{\sqrt{\frac{۰,۵۶۴۴}{۳}}} = \frac{۰,۶۶}{\sqrt{۰,۱۸۸}}$$

$$= \frac{۰,۶۶}{\sqrt{\ }} = \frac{۰,۶۶}{۰,۴۳} = ۱,۵۳ \checkmark$$

مقادیر بحرانی $n = ۵$

t

$\alpha = ۰,۰۵$

Arman

Subject : (   )   Date : ....

$t = ۱/۵۲۴$ $n = ۵$ $\alpha = ۰/۰۵$

$t_{\frac{\alpha}{۲}, ۵-۱} = t_{۰/۰۲۵, ۴} = \pm ۳/۱۸۲$

سطح معنی‌دار

درجه آزادی

چون $t = ۱/۵۲۴$ بین ۳/۱۸۲ و ۳/۱۸۲ - قرار می‌گیرد

نمی‌توان فرض صفر را رد کرد بنابراین شماره‌های ظاهر شده در دو تاس مستقل هستند

و هیچ همبستگی معنی‌دار بینشان نیست.

مثال بعدی از کتاب صفحه ۱۸۵ ۴-۱۳ را حل کنیم.

Arman

آنالیز واریانس (تحلیل واریانس)

مقایسه میانگین های سه گروه یا بیشتر

① اگر چند جامعه را بخواهیم مقایسه کنیم باید چه روشی به کار ببریم؟

② در آنالیز واریانس جهت برای ما مطرح نیست؟

آیا این چند جامعه از نظر متغیر مدنظر با هم یکسان هستند یا نیستند؟

مثلاً آیا بسته بندی در فروش اثرگذار است؟

فروشگاهی با ۵ نوع بسته بندی می خواهیم بررسی کنیم با آنالیز واریانس استفاده می کنیم اگر فرض صفر رد شد، پس نتیجه می گیریم بسته بندی های مختلف بر فروش تاثیر دارند.

تحلیل واریانس ANOVA یک روش آماری پارامتریک می باشد که از آن جهت تعیین معنی دار بودن تفاوت های بین میانگین های داده های سه گروه یا بیشتر استفاده می شود.

در آزمون t اختلاف بین میانگین های دو گروه با اختلاف تصادفی قابل انتظار مقایسه می شود، اما در تحلیل واریانس تغییرات بین میانگین های گروه ها (تغییرات بین گروهی) با تغییرات نمرات داخل گروه ها (تغییرات درون گروه ها) مقایسه می شود.

در اینجا نسبتی به نام F حاصل می شود که: F برابر است با متوسط واریانس بین گروهی تقسیم بر متوسط واریانس درون گروه ها

آنالیز واریانس خیلی کمک کننده است:

H_0 ۱- میانگین همه گروه ها برابر است

H_1 ۲- یک گروه متفاوت از بقیه است

Arman

به این علت آنالیز واریانس می‌گوییم چون پراکندگی‌ها را آنالیز می‌کنیم

چنانچه تغییرات بین گروهی، از آنچه که انتظار می‌رود به صورت صرفاً تصادفی حادث شود، بیشتر از تغییرات درون گروهی باشد، می‌توان نتیجه‌گیری نمود که حداقل یکی از میانگین‌های گروه‌ها از میانگین دیگر متفاوت است.

فرضیه صفر H را آزمون F به صورت ذیل نشان می‌دهند.

$\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$

در فرضیه صفر فرض بر این است که میانگین‌های سه گروه یا بیشتر که به صورت تصادفی از جامعه آماری شبیه انتخاب شده‌اند با یکدیگر مساوی هستند هنگامی که حداقل میانگین یکی از نمونه‌ها به طور معنی‌داری از میانگین‌های دیگر متفاوت باشد، F معنی‌دار تلقی شد، و فرضیه صفر را رد می‌کنیم

منابع تغییرات

محاسبه F به طور نظری بسیار آسان بوده و مستلزم عملیات پیچیده نیست

هنگامی که چندین گروه مختلف از داده‌ها مورد مقایسه قرار می‌گیرند، هر گروه دارای میانگین است. (میانگین گروه‌ها) و کل داده، یا تمام میانگین‌های گروه‌ها نیز دارای میانگین است

میانگین کل : میانگین کل از طریق مجموع نمرات تمام گروه‌ها تقسیم بر کل تعداد آزمودنی‌ها (N) حساب می‌شود

$\bar{X}_T$

میانگین هر یک از گروه‌ها ممکن است با میانگین کل تفاوت داشته باشند.

تغییرات یا انحراف میانگین گروه از میانگین کل، تغییرات بین گروهی خوانده می‌شود.

Between groups

علاوه بر تفاضل یا تغییرات بین میانگین‌ها، هر نمره واحد نیز یک مقدار معین از میانگین گروه خود انحراف دارد. این منبع تغییرات، تغییرات درون گروهی within groups خوانده می‌شوند.

بنابراین منابع تغییرات بین گروهی (B) و درون گروهی (W) در محاسبه تحلیل واریانس از اجزا اصلی به شمار می‌روند.

منبع تغییرات سوم یا تغییرپذیری کل را می‌توان از طریق تعیین میزان انحراف هر یک از نمرات گروه‌ها از میانگین کل محاسبه نمود. تغییرات کل برابر است با مجموع تغییرات بین گروهی و درون گروهی

تغییرات کل = تغییرات بین گروهی + تغییرات درون گروهی

مجموع مربعات درون گروهی را می‌توان با استفاده از نمرات فردی گروه و میانگین گروه محاسبه نمود

درون گروهی $SS_w = \sum (x - \bar{x}_{\text{گروه}})^2$

CS CamScanner

مجموع مربعات بین گروه‌ها را می‌توان از طریق تفریق میانگین کل از میانگین هر گروه محاسبه نمود.

$SS_B = \Sigma(\bar{x}_{\text{گروه}} - \bar{x}_{\text{کل}})^2$ بین گروه‌ها

مجموع مربعات کل را می‌توان از طریق تفریق میانگین کل از هر یک واحدها جدا کرد

$$SS_T = \Sigma(x - \bar{x}_{\text{کل}})^2$$

بنابراین: مجموع مربعات کل همیشه برابر است با مجموع مربعات بین گروه‌ها به علاوه مجموع مربعات بین گروه

$$SS_T = SS_B + SS_W$$

چون اشکال مجموع مربعات به تنهایی فایده ندارند بلکه ما به میانگین مربعات نیاز داریم بنابراین:

جهت تخمین یا برآورد مقدار اختلاف بین تعداد افرادی که مجموع مربعات بین گروه‌ها (تعداد گروه‌ها) و درون گروه‌ها (تعداد نمرات افراد) را تشکیل می‌دهند، از میانگین مربعات استفاده می‌کنیم.

میانگین مربعات (MS) از طریق تقسیم هر یک از مجموع مربعات (SS) بر درجه آزادی مربوط به آن محاسبه می‌شود.

$MS = SS/df$

در نهایت با استفاده از این روش امکان مقایسه تغییرات بین گروه‌ها SS_B و درون گروه‌ها SS_W فراهم می‌شود.

CS CamScanner

درجه آزادی ها   درون گروهی $Df_w = N - K$

N = تعداد کل نمونه های گروه ها

K = تعداد گروه ها   بین گروهی $Df_B = K - 1$

جهت تعیین مربعات درون گروهی ، مجموع مربعات درون گروهی را بر درجات آزادی درون گروهی تقسیم می کنیم.

$$MS_w = SS_w / df_w$$

جهت تعیین میانگین مربعات بین گروهی، مجموع مربعات بین گروهی را بر درجات آزادی بین گروهی تقسیم می کنیم.

$$MS_B = SS_B / df_B$$

مقادیر مربوط به میانگین مربعات پس از محاسبه به طور مستقیم قابل مقایسه هستند و جهت محاسبه F می توان از آنها استفاده کرد.

$$F = \frac{MS_B}{MS_w}$$

روش نمرات خام

داده هایی که شامل اعداد اعشار است و مقدار N نیز بزرگ است روش ترجیحی جهت محاسبه F خسته کننده است.

فرمول های بیشتر جهت محاسبه که به جای نمرات انحراف از نمرات خام استفاده می کن

در این روش ابتدا SS_B و SS_T را محاسبه می‌کنیم و سپس از فرمول زیر SS_w بدست می‌آید و سایر مراحل طبق روش توضیحی خواهد بود:

$SS_w = SS_T - SS_B$

فرمول محاسبه SSB

$$SS_B = \frac{(\sum x_1^2)}{n_1} + \frac{(\sum x_2)^2}{n_2} + \frac{(\sum x_3)^2}{n_3} + \dots + \frac{(\sum x_k)^2}{n_k} - \frac{(\sum x_T)^2}{N}$$

$$SST = \sum x^2 - \frac{(\sum x_T)^2}{N}$$

درجه آزادی $df_B = K - 1$

$$SS_w = SS_T - SS_B$$

$df_w = N - K$

مثال:

x_1 A	x_2 B	x_3 C
۲	۳	۴
۵	۴	۳
۳	۲	۱
۱	۱	۱
۲	۳	۲

$$SS_B = \frac{(۲+۵+۳+۱+۲)^۲}{۵} + \frac{(۳+۴+۲+۱+۳)^۲}{۵} + \frac{(۴+۳+۱+۱+)^۲}{۵} - \frac{(۳۷)^۲}{۱۵} = \frac{۱۶۹}{۵} + \frac{۱۶۹}{۵} + \frac{۱۲۱}{۵} - \frac{۱۳۶۹}{۱۵}$$

$$= (۳۳,۸ + ۳۳,۸ + ۲۴,۲) - ۹۱,۲ = ۰,۶$$

$\sum X_T = (۲+۵+۳+۱+۲+۳+۴+۲+۱+۳+۴+۳+۱+۱+۲) = ۳۷$

$N = ۱۵$ $\sum x^۲ = ۴+۲۵+۹+۱+۴+۹+۱۶+۴+۱+۹+۱۶+۹+۱+۱+۴ = ۱۱۳$

$K = ۳$ تعداد گروه‌ها $SST = \sum x^۲ - \frac{(\sum x_T)^۲}{N} = ۱۱۳ - ۹۱,۲ = ۲۱,۷۴$

$SS_w = SS_T - SS_B = ۲۱,۷۴ - ۰,۶ = ۲۱,۱۴$ $df_B = k-1 = ۳-۱ = ۲$

$df_w = N - K = ۱۵ - ۳ = ۱۲$

Arman

$MS_w = \frac{۲۱,۱۴}{۱۲} = ۱,۷۶$ $F = \frac{۰,۳}{۱,۷۶} \simeq ۰,۱۷$

$MS_B = \frac{۰,۶}{۲} = ۰,۳$

CamScanner

Subject : (   )   مثال کتاب   Date :

						ΣX	$X^{۲}$
X_1	۸۶	۷۹	۸۱	۷۰	۸۴	۴۰۰	۱۴۰,—
X_2	۸۹	۸۲	۸۷	۷۶	۹۰	۴۲۴	۱۷۹,۷۷۶
X_3	۸۲	۶۸	۷۲	۷۱	۸۱	۳۷۵	۱۴۰,۶۲۵

$$\frac{(\Sigma x_1)^{۲}}{۵} = \frac{(۴۰۰)^{۲}}{۵} = \frac{۱۶۰,۰۰۰}{۵} = ۳۲,۰۰۰$$

$$\frac{(\Sigma x_2)^{۲}}{۵} = \frac{۱۷۹۷۷۶}{۵} = ۳۵,۹۵۵$$

$$\frac{(\Sigma x_3)^{۲}}{۵} = \frac{۱۴۰۶۲۵}{۵} = ۲۸۱۲۵$$

$$(\Sigma x_T)^{۲} = (۴۰۰ + ۴۲۴ + ۳۷۵)^{۲} = (۱,۱۹۹)^{۲} = ۱,۴۳۷۶۰۱$$

$$\Sigma x^{۲} = (۸۶)^{۲}+(۷۹)^{۲}+(۸۱)^{۲}+(۷۰)^{۲}+(۸۴)^{۲}+(۸۹)^{۲}+(۸۲)^{۲}+(۸۷)^{۲}$$
$$+(۷۶)^{۲}+(۹۰)^{۲}+(۸۲)^{۲}+(۶۸)^{۲}+(۷۲)^{۲}+(۷۱)^{۲}+(۸۱)^{۲} = ۹۶۵۲۳$$

$$SS_B = (۳۲۰۰۰ + ۳۵۹۵۵ + ۲۸۱۲۵) - \frac{۱۴۳۷۶۰۱}{۱۵} = ۹۶۰۸۰ - ۹۵۸۴۰ = ۲۴۰$$

$$SS_T = ۹۶۵۲۳ - ۹۵۸۴۰ = ۶۸۳ \qquad SS_W = ۶۸۳ - ۲۴۰ = ۴۴۳$$

$$N - k = ۱۲ \Rightarrow df_W \qquad MS_B = \frac{۲۵۰}{۲} = ۱۲۵ \qquad F = \frac{MS_B}{MS_W} = ۳,۳۷$$

$$k - ۱ = ۳ - ۱ = ۲ \quad df_B \qquad MS_W = \frac{۴۴۳}{۱۲} = ۳۶,۹$$

در سطح $\alpha = ۰.۰۵$

$$F_{\alpha, n-1, k(n-1)} = F_{۰.۰۵, ۲, ۱۲} = ۳,۸۸$$

— ***Arman***

F محاسبه شده چون ۳,۳۷ از ۳,۸۸ کمتر است فرض صفر مبنی بر تساوی میانگین‌ها تأیید

می‌شود ۵٪ است این نتوان را رد کرد و اختلاف در میانگین‌ها ناشی از تصادفی است.

آنالیز واریانس :

این امکان را بوجود می آورد که جوامع را با یک متغیری که تغییر ایجاد می‌کند تا نه بررسی می‌کند و جهت اثرگذاری مطرح نیست. آنالیز واریانس بر اساس نوع مشکل یک راهه و دو راهه دارد. برای آنالیز واریانس یک راهه ما فقط اثر یک عامل را چک می‌کنیم. مثلاً نوع بسته‌بندی، نوع مدیریت از نظر مصرف کننده کالا تفاوتی داشته باشند یا نه ؟

اما در آنالیز واریانس دو عامله هم زمان دو عامل را بر روی متغیر مورد نظر سنجش می‌کنیم و اثرات متقابل را می‌سنجیم. آیا میزان فروش ما با تغییر مکان تغییر می‌کند، مغازه ما بیش از جاهای دیگر را بررسی می‌کنیم. از روش آنالیز واریانس دوراهه [illegible] و این نوع آنالیز واریانس دوراهه از اسلاید ۱۴۷ تا ۱۵۹ حذف است.

CamScanner

تعیین معنی داری بودن F

آزمونهای تعقیبی Pos Hoc که نرم افزار spss میتواند آنها را انجام دهد جزء آزمونهای برای تعیین معنی دار هستند.

فرآیند تعیین معنی داری

جهت تعیین معنی دار بودن مقدار F محاسبه شده، باید به جدول F رجوع شود که درجات آزادی ویژه مربوط به تغییرات درون گروهی و بین گروهی نشان داده شده است. در این جدولها درجه آزادی بین گروهی dfB که از ردیف افقی بالا و درجه آزادی خطا (dfE که) از ردیف عمودی سمت چپ خوانده میشود. یعنی گاهی $SS_E = SS_w$ ، تغییرات ~~بدون~~ درون گروهی SS_w یا SS_{error} را SS_E نیز میگویند.

اما معنی دار بودن F به تنهایی مشخص نمی‌کند که کدام یک از گروهها با یکدیگر تفاوت دارند. F معنی دار نشان میدهد که صرفاً تفاوتی بین گروهها وجود دارد، حال برای تعیین گروههایی که به طور معنی داری از هم متفاوت هستند، باید یکی از آزمونهای تعقیبی یا Post Hoc را اجرا نمود.

دو تا از متداولترین آزمونهای مورد استفاده، آزمون فاصله اطمینان شفه (I) و آزمون تفاوت صحیح معنی دار HSD یا توکی هستند.

در آزمون شفه امکان انجام تمام مقایسه‌ها وجود دارد. اما دقتش به اندازه توکی نیست. درحالیکه آزمون توکی آسانتر بوده و تنها مقایسه‌ها به صورت جفت مرتب به یکدیگر انجام میشود. آزمون توکی برای بیشتر طرحهای تحقیقاتی آزمون مناسبی است.

CS CamScanner

مثال: در دانشکده‌ای دروس تحقیق در عملیات ۱، ۲، ۳ و پیشرفته با یک استاد تدریس می‌شود. میانگین نمره‌های کلاسها در چهار ترم اخیر طبق جدول، آیا در سطح معنی داری ۵٪ می‌توان ادعا کرد که میانگین نمره‌های این چهار درس با هم متفاوت است؟

$$\begin{cases} H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_1 \text{ حداقل یکی از میانگین‌ها با بقیه مساوی نیست} \end{cases}$$

$N = ۱۲$

$K = ۴$

x^2	$\sum X$					
۲۶۳۲	۵۱٫۳	۱۳٫۵	۱۲	۱۲٫۸	۱۳	تحقیق در عملیات ۱
۱۸۹۲	۴۳٫۵		۱۴٫۵	۱۵	۱۴	〃 〃 ۲
۱۸۸۴	۴۳٫۴		۱۵٫۴	۱۳	۱۵	〃 〃 ۳
۶۴۰	۲۵٫۳			۱۲٫۲	۱۳٫۱	〃 پیشرفته
=۲۲۲۸						

$$\frac{(\sum x_T)^2}{N} = \frac{(۵۱٫۳+۴۳٫۵+۴۳٫۴+۲۵٫۳)^2}{N(۱۲)} = \frac{(۱۶۳٫۵)^2}{۱۲} = \frac{۲۶۷۳۲}{۱۲}$$

$$\frac{(\sum x_1)^2}{n} = \frac{۲۶۳۲}{۴} = ۶۵۸$$

$$SS_B = (۶۵۸ + ۶۳۰٫۷ + ۶۲۸ + ۳۲۰) - ۲۲۲۸$$

$$\frac{(\sum (x_2))^2}{n} = \frac{۱۸۹۲}{۳} = ۶۳۰٫۷$$

$$= ۲۲۳۶٫۷ - ۲۲۲۸ \simeq ۸٫۷$$

$$\frac{(\sum x_3)^2}{n} = \frac{۱۸۸۴}{۳} = ۶۲۸$$

$$\frac{(\sum x_4)^2}{n} = \frac{۶۴۰}{۲} = ۳۲۰$$

$$SS_T = \sum x^2 - \frac{(\sum x_T)^2}{N} = ۲۲۴۱٫۹۵ - ۲۲۲۸$$

$$= ۱۳٫۹۵$$

$$\sum x_i^2 = (۱۳)^2 + (۱۲٫۸)^2 + (۱۲)^2 + (۱۳٫۵)^2 + (۱۴)^2 + (۱۵)^2 + (۱۴٫۵)^2 + (۱۵)^2 + (۱۳)^2 + (۱۵٫۴)^2 + (۱۳٫۱)^2 + (۱۲٫۲)^2 = ۱۶۹ + ۱۶۹ + ۱۷۱٫۴۱ + ۱۸۲٫۲۵ + ۱۶۳٫۸۴ + ۱۴۴ + ۱۴۸٫۸۴ + ۱۹۶ + ۲۱۰٫۲۵ + ۲۲۵ + ۲۲۵ + ۲۳۷٫۱۶ = ۲۲۴۱٫۹۵$$

$$SS_W = ۱۳٫۹۵ - ۸٫۷ = ۵٫۲۵$$

CamScanner

$$F = \frac{MS_B}{MS_W} = \frac{۲٫۹}{۰٫۶۵۶} = ۴٫۴۲$$

$$MS_B = \frac{SS_B}{K-1} = \frac{۸٫۷}{۴-۱} = \frac{۸٫۷}{۳} = ۲٫۹$$

$$MS_W = \frac{SS_W}{N-K} = \frac{۵٫۲۵}{۱۲-۴} = \frac{۵٫۲۵}{۸} = ۰٫۶۵۶$$

بحرانی جدول $F_{۰٫۰۵,۳,۸} = ۴٫۰۷$

بدست آمده $F = ۴٫۴۲$

چون آماره محاسبه شده از مقدار بحرانی جدول بیشتر است، بنابراین فرض صفر که برابری میانگین های جامعه ها باشد در سطح معنی دار ۵٪ رد شده، یعنی محقق توانسته ادعا کند که تفاوت در میانگین های جامعه در سطح ۵٪ معنی دار است.

چون تفاوت میانگین های جامعه معنادار است پس نیاز به آزمونهای تعقیبی داریم. آزمون شفه را بررسی میکنیم که محافظه کار ترین نوع آزمون تعقیبی است.

در آزمون شفه (I) هیچ گونه محدودیتی در انجام تعداد مقایسه های ممکن وجود ندارد.

آزمون فاصله اطمینان شفه به ما این امکان را میدهد تا بتوانیم میانگین های تمام دو گروه یا ترکیبی از گروه ها را با هم مقایسه کنیم.

$$MS_W = MS_E$$

$$I = \sqrt{(K-1)(F_\alpha)\left(\frac{۲MS_E}{n}\right)}$$

فرمول آزمون شفه
تعداد گروه ها مساوی

Arman

از جدول $\rightarrow F_{۰٫۰۵,۳,۸}$ مثلاً

CamScanner

اگر تعداد گروه‌ها مساوی نباشند فرمول زیر استفاده می‌شود:

$$I = \sqrt{(k-1)(F_\alpha)(MS_e)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

چنانچه تعداد گروه‌ها مساوی نباشند، معادله فوق برای اینکه دو گروه با تعداد n نامساوی قابل مقایسه باشند به شرح ~~زیر~~ بالا می‌باشد.

رسم جدول اختلاف میانگین‌ها برای مثال قبل.

	$\bar{X}_۱$	$\bar{X}_۲$	$\bar{X}_۳$	$\bar{X}_۴$
$\bar{X}_۱$	۰	۱٫۶۷۵	۱٫۶۴۷	۰٫۱۷۵
$\bar{X}_۲$	۱٫۶۷۵	۰	۰٫۰۳	۱٫۸۵
$\bar{X}_۳$	۱٫۶۴۵	۰٫۰۳	۰	۱٫۸۲
$\bar{X}_۴$	۰٫۱۷۵	۱٫۸۵	۱٫۸۲	۰

$$\bar{X}_۱ = \frac{۵۱٫۳}{۴} = ۱۲٫۸۲۵$$

$$\bar{X}_۲ = \frac{۴۳٫۵}{۳} = ۱۴٫۵$$

$$\bar{X}_۳ = \frac{۴۳٫۴}{۳} = ۱۴٫۴۷$$

$$\bar{X}_۴ = \frac{۲۵٫۴}{۲} = ۱۲٫۶۵$$

آماره تاکی یا شفه را محاسبه می‌کنیم و با اختلاف میانگین‌های بدست آمده در جدول مقایسه می‌کنیم هر گروهی که متفاوت بود

که شفه برای این مسئله استفاده می‌شود، که تعداد گروه‌ها مساوی نیستند

$$I = \sqrt{(۴-۱)(۴٫۰۷)(۰٫۶۵۶)\left(\frac{۱}{۴}+\frac{۱}{۳}+\frac{۱}{۳}+\frac{۱}{۲}\right)}$$

$$= \sqrt{(۳)(۴٫۰۷)(۰٫۶۵۶)(۱٫۴۳)}$$

$$F_{۹۵,۳,۸} = ۴٫۰۷ \quad MS_e = MS_w$$

$$= \sqrt{۱۱٫۴۵۳۹} = \boxed{۳٫۳۸۴}$$

CamScanner