بسم الله الرحمن الرحیم

هفته اول ۹۳/۰٦/۲۳

تشکیل نشد

هفته دوم ۹۳/۰٦/۳۰ (جلسه اول)

بیان کلیات و نقشه راه کلاس

جلسه دوم ۹۳/۰۷/۰٦

سرفصل ها

۱-مقدمه و مرور بر مباحث گذشته.

۲-آزمون های آماری:

الف:آزمون های اسمی و وابسته

ب: آزمون های ترتیبی

ج: آزمون های فاصله ای

۳-توزیع نرمال

۴-همبستگی و رگرسیون

۵-تحلیل عاملی، تحلیل خوشه ای، تحلیل مسیر

<table dir="rtl">
<tr><td>مستقل ها</td><td>اسمی</td><td>ترتیبی</td><td colspan="2">فاصله ی نسبتی</td><td></td></tr>
<tr><td rowspan="2">اسمی</td><td rowspan="2">کای دو و وابسته های آن X^2</td><td rowspan="2">X^2</td><td>دو شقی</td><td>T و انواع آن</td><td rowspan="2">تحلیل واریانس و وابسته های آن</td></tr>
<tr><td>چند شقی</td><td>F</td></tr>
<tr><td rowspan="2">ترتیبی</td><td rowspan="2">X^2</td><td rowspan="2">کندال ها t_b و t_c و سامرزها D_s</td><td>دو شقی</td><td>T</td><td rowspan="2"></td></tr>
<tr><td>چند شقی</td><td>F</td></tr>
<tr><td>فاصله ای/ نسبتی</td><td>باید حتما به صفر و یک تبدیل کنیم و بود و نبود را بررسی کنیم در غیر این صورت آزمونی نداریم</td><td></td><td colspan="2"></td><td></td></tr>
</table>

مثال : هر دو اسمی می باشد

*مبنای محاسباتی آزمون:تضاد فراوانی مشاهده شده با فراوانی های مورد انتظار است

*فرضیه:به نظر می رسد میزان تماشای تلویزیون زنان و مردان تفاوت وجود دارد.

این فرضیه H_1 است چون لااقل یک تفاوت را بیان می کند.اما اگر بگوییم:

به نظر می رسد بین میزان تماشای تلویزیون زنان و مردان تفاوت وجود ندارد آن گاه فرضیه H_0 را ساخته ایم .

در آمار همیشه H_0 را بررسی می کنیم.

متغییر ها	طبقه بندی	جهت	صفر(هر نوع)	آماره ها
اسمی	+	–	–	mo (نما)
ترتیبی	+	+	–	mo md
فاصله ای	+	+	قراردادی	mo Md
نسبتی	+	+	مطلق	Mo Md و $\bar{x}$

مثال :مستقل اسمی و وابسته ترتیبی .

مثال:اگر هر دو ترتیبی باشند.

فرضیه :به نظر می رسد بین سن و تماشای تلویزیون رابطه وجود دارد.

سن تماشا [نوزاد ۰ ، کودکان ۱ ، نوجوان ۲، جوانان ۳، میان سال ۴، بالاتر ۵]

میزان تماشا [خیلی زیاد ۵، زیاد ۴ ، متوسط ۳ ، کم ۲ ، خیلی کم ۱ ، اصلا ۵]

جلسه سوم ۹۳/۷/۲۷ (دوهفته قبل ۱۳ و ۲۰ مهر تعطیل بود)

شرایط آزمون کای ۲

جدول۲*۲	اختلاف فراوانی مورد انتظار و مشاهده شده نباید کمتر از ۵ باشد. ۲۰ درصد فراوانی های مورد انتظار نباید کمتر از ...
غیر از۲*۲	اختلاف فراوانی مورد انتظار و مشاهده شده وجود دارد و نباید از ۱ کمتر باشد. ۲۰ درصد فراوانی های مورد انتظار وجود دارد نباید از ۵ کمتر باشد.

* آزمون متغیر های ترتیبی:

* مبنای آزمون متغییر های ترتیبی: تفاضل جفت های موافق و مخالف

۱- آزمون Dسامرز (نامتقارن)

$$D_{s_{xy}} = \frac{ns - nd}{2!\, ns + nd + ty}$$

Ns = جفت های موافق

Nd = جفت های مخالف

متغیر مستقل ←

متغیر وابسته ↓

تحصیلات	کمتر از دیپلم	دیپلم	عالیه	کل
زیاد	۲۰	۳۰	۴۰	۹۰
کم	۱۰	۱۵	۲۰	۴۰
اصلا	۵	۱۰	۱۵	۳۵
کل	۳۵	۵۵	۷۵	۱۶۵

* **خط دیاگونال** (مورب) خطی که اطلاعات را به ۲ قسمت تقسیم کند. برای این کار باید جهت هر متغیر (بالا و پایین)آن را پیدا کنیم.

	کمتر از دیپلم	دیپلم	عالیه	کل
اصلا	۲۰	۳۰	۴۰	۹۰
کم	۱۰	۱۵	۲۰	۴۰
زیاد	۵	۱۰	۱۵	۳۵
کل	۳۵	۵۵	۷۵	۱۶۵

* **جفت های موافق:**

برای خط دیاگونال جفت های موافق جایی است که هر دو متغیر در بالاترین سطوح است یعنی خطی که ۴۰را به ۵ وصل می کند.

* **جفت های مخالف:**

از لحاظ سطح یکی از متغیر ها در بیشترین حالت است و متغیر دیگر در کمترین حالت. که در جدول اول می شود خطی که ۲۰ را به ۱۵ وصل کرده است.

$$\begin{aligned} Ns &= 40(15 + 10 + 5 + 10) + 30(10 + 5) + 20(0) + 20(10 + 5) + 15(5) \\ &\quad + 10(0) + 15(0) + 10(0) + 5(0) \\ &= 40(40) + 30(15) + 20(15) + 15(15) = 2425 \end{aligned}$$

سر نقطه را به خط دیاگونال در نظر می گیریم (مثلا ۴۰ یا ۵) . سپس سطر و ستونی را که ۴۰ یا ۵ در آن است را حذف می کنیم و بقیه داده ها را با هم جمع می کنیم . سپس به جلو می رویم . عدد بعدی ۳۰ است . این جا نیز سطر و ستونی را که ۳۰ را در خود جای داده حذف می کنیم و آن چه می ماند فقط اعداد ۱۰ و ۵ است . نباید به عقب بر گردیم.

حالا محاسبه جفت های مخالف

$$Nd = 15(15+10+20+30) + 20(30+20) + 40(10) + 10(20+10) + 15(20) = 2725$$

$$Ty = 20(30+40) + 30(40) + 40(0) + 10(15+20) + 15(20) + 20(0) + 5(0+15) + 15(0) = 3525$$

برای Ty از مقوله y که به X وابسته است به صورت ستونی همان مراحل بالا را انجام می دهیم.

$$D_{s_{xy}} = \frac{-2425 - 2725}{22425 + 2725 + 3525} = -0/034$$

علامت منفی نشان می دهد که رابطه بین تماشای تلویزیون و تحصیلات معکوس است .یعنی هرچه تحصیلات زیادتر باشد میزان تماشا کم می شود.

***اگر $|D_s|$ از ۰/۳ کمتر باشد تفاوت معنا دار نیست .اگر از ۰/۳ بیشتر باشد تفاوت معنا دار است.**

-1<d<+1

H_0: به نظر می رسد بین میزان تماشای تلویزیون و میزان تحصیلات رابطه وجود **ندارد**.

H_1: به نظر می رسد بین میزان تماشای تلویزیون و میزان تحصیلات رابطه وجود **دارد**.

در این مثال H_0 تایید شد چون D_s نشان می دهد که تفاوت معنا دار نیست و H_1 نیز رد می شود پس هیچ رابطه ای بین میزان تماشای تلویزیون و میزان تحصیلات وجود ندارد.

مثال : اکنون مثال قبل را برای حالتی حساب می کنیم که $D_{s_{yx}}$است. یعنی در اینجا X وابسته است. در این مثال Ns و Nd تغییر نمی کند و فقط باید به جای Ty، Tx را حساب کنیم.

این محاسبه به شکل ستونی است:

$$Tx = 20(10+5) + 10(5) + 30(15+10) + 15(10) + 4(20+15) + 20(15) = 2950$$

$$D_{s_{yx}} = \frac{2425 - 2725}{8100} = -0/037$$

*زمانی که نمی دانیم کدام متغیر وابسته و کدام مستقل است ، باید از میانگین استفاده کنیم به این صورت که میانگین ۸۱۰۰ و ۸۶۷۵ را محاسبه می کنیم و آن گاه ۳۰۰ را بر مقدار حاصل تقسیم می کنیم تا رابطه ای به دست آید.

جلسه چهارم ۹۳/۸/٤

T_b کندال: برای جداول مربعی شکل است و **T_c کندال**: جداول مستطیلی است.

کندال۲تا نقد به سامرز وارد کرد :

۱- این که فقط برای مربعی ها جواب میدهد و برای مستطیلی ها جواب تفاوت دارد

۲- سامرز هر بار یکی از متغیرها را وابسته و دیگری را مستقل می گرفت اما ممکن است هر دو برهم موثر باشند.

بنابراین:

$$T_b = \frac{Ns - Nd}{\sqrt{(Ns + Nd + Ty)(Ns + Nd + Tx)}}$$

*** این فرمول در حقیقت $D_{s_{xy}} \times D_{x_{yx}}$ است.**

مثال : می خواهیم ارتباط بین مطالعه روزنامه و مشارکت اجتماعی را بررسی کنیم.

مطالعه/مشارکت	اصلا	کم	زیاد
زیاد	۵۰	۴۰	۲۰
کم	۴۰	۳۰	۱۰
اصلا	۳۰	۲۰	۱۰

دراینجا فراوانی مهم نیست بلکه برای تعیین سرهای دیاگونال تنها ترتیب مهم است یعنی در نقطه ۳۰، هم مشارکت اصلا است و هم مطالعه روزنامه.حتی اگر این فراوانی یعنی ۳۰ عدد دیگری مانند ۱۰ یا صفر باشد باز هم همین خانه یک سردیاگونال است:

Ns=20(30+40+30+20)+40(40+20)+10(30+20)+ 30+(20)=5900
Nd=10(30+40+50+40)+30(50+40)+10(40+50)+ 30(50)=6700

* تا اینجا معلوم می شود که علامت T_b منفی است اما محاسبه ادامه دارد :

Ty=50(40+20)+40(20)+40(30+10)+30(10)+20(30+10)+30(10)=6800 ty=6800
Tx=50(40+20)+40(20)+40(30+30)+30(30)+20(10+10)+10(10)=7600 tx=7600

$$T_b = \frac{5900 - 6700}{\sqrt{(5900 + 6700 + 6800)(5900 + 6700 + 7600)}} = \frac{-8}{19795} = -0.04$$

$$\Rightarrow T_b = -0.04$$

در حقیقت یعنی مشارکت اجتماعی با مطالعه روزنامه رابطه معکوس دارد و رابطه معنادار نیست چون از ۰/۳ کمتر است . یعنی مطبوعات این جامعه مشارکت اجتماعی را از بین برد .

*** فرمول T_c کندال:**

$$\boldsymbol{T_c} = \frac{Ns - Nd}{\frac{N^2(m-1)}{2m}}$$

m: مینیمم تعدا سطر یا ستون. مثلا اگر جدول ۲*۳ باشد آنگاه m برابر ۲ است.

مطالعه/مشارکت	اصلا	کم	زیاد	کل
زیاد	۵۰	۴۰	۲۰	
کم	۶۰	۶۰	۲۰	
اصلا	۱۱۰	۱۰۰	۴۰	N=۲۵۰

در این جدول تعداد سطرها ۲ وستون ها و ۳ است پس

M = 2
N=250
Ns=60(40+20)+60(20)=3600+1200=4800
Nd=20(40+50)+60(50)=1800+3000=4800

$$\boldsymbol{T_c} = \frac{Ns - Nd}{\frac{N^2(m-1)}{2m}} = \frac{4800 - 4800}{\frac{250^2(2-1)}{2*2}} = 0$$

* یعنی هیچ رابطه ای بین مشارکت اجتماعی و مطبوعات وجود ندارد.

*وقتی متغیر وابسته (y) فاصله ای است و نیز وقتی متغیر مستقل (x) اسمی دو شقی باشد از آزمون t استفاده می شود.

استفاده می شود.

n ها برابر

۱- مقایسه میانگین

واریانس برابر

n ها نا برابر

n ها برابر

واریانس نابرابر

n ها نا برابر

n ها برابر

واریانس برابر

n ها نا برابر

۲- مقایسه ۲ نسبت

n ها برابر

واریانس نابرابر

n ها نا برابر

۹۳/۸/۱۱(عاشورا و تاسوعا- تعطیل)

جلسه پنجم ۹۳/۸/۱۸

در آزمون t (تی استیودنت) یک مقدار از قبل داریم و می خواهیم اطلاعات جمع آوری شده را با آن مقایسه کنیم.

$$t_c = t_1 = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

t_c حالت اول: t

میانگین به دست آمده از نمونه ها: $\bar{x}$

میانگین از قبل تعیین شده: μ

انحراف معیار: s

کل نمونه ها: n

- در آزمون t حتما باید متغیر فاصله ای باشد. در غیر این صورت اصلا نمی توانیم از میانگین استفاده کنیم
درجه آزادی: $dp = N - 1$

- اگر $\left.\begin{matrix} t_t = t_c \\ t_t > t_c \end{matrix}\right|$ باشد آن گاه فرضیه H_0 تایید می شود. یعنی متغیرها باهم رابطه ای ندارند. اما اگر $t_t < t_c$ باشد آن گاه فرضیه H_1 تایید می شود. یعنی متغیرها باهم رابطه دارند.

در حالت دیگر می خواهیم با متغیرهای نسبتی از آزمون t استفاده کنیم.

بعنوان مثال فرض کنیم میزان بیکاری ۰/۰۹ اعلام شده است.

$$z = \frac{\pi - p}{\sqrt{\frac{p\theta}{N}}}$$

$$Z = \frac{\text{نسبت از قبل تعیین شده} - \text{نسبت به دست آمده از نمونه}}{\sqrt{\frac{\text{واریانس}}{\text{کل ها نمونه}}}}$$

$$S^2(V) = \frac{\sum n_i x_i - \frac{(\sum n_i x_i)^2}{n}}{N/_{N-1}} = \frac{\sum x_i (x_i - \bar{x})^2}{N/_{N-1}}$$

شغل	***xi***
بلی	فرد ۱
خیر	فرد ۲
بلی	فرد ۳
بلی	فرد ۴
بلی	فرد ۵
بلی	فرد ۶
خیر	فرد ۷
بلی	فرد ۸
خیر	فرد ۹
خیر	فرد ۱۰

از جدول نسبت بیکاری $\frac{4}{10}$= ۰/۴ حساب می شود. یعنی ۰/۴ = q=π

نسبت به دست آمده $\Rightarrow q = 0.4 \Rightarrow p=1-q = 1- 0.4 = 0.6$

P به دست آمده یعنی۰/۴ است.

q = ۰/۴

p = ۰/۶ احتمال شاغل بودن

n =۱۰

p = ۰/۰۹

p = ۱ − q احتمال غیر شاغل

$$\Rightarrow\ z = \frac{0.4 - 0.09}{\sqrt{\frac{0.6 \times 0.4}{10}}} = 2.06$$

در داده هایی که فراوانی ندارند واریانس را با ضرب pq محاسبه می کنیم.

در این جا اگر فرض کنیم Z جدول یعنی z_t برابر ۲/۰۰۱ باشد (z_t = ۲/۰۰۱) ⇦ ۲/۰۰۱ < ۲/۰۶

⇦ $z_c > z_t$ ⇦ یعنی H_1 تأئید می شود و مثلاً آماری که وزیر برای بیکاری داده است (۰/۰۹) رد می شود.

H_1: به نظر می رسد میزان بیکاری در جامعه با میزان بیکاری اعلام شده از سوی وزیر تفاوت دارد.

H_0:به نظر می رسد میزان بیکاری در جامعه برابر با میزان بیکاری اعلام شده از سوی وزیر است.

- در جدول محاسبه شده چون رقم مثبت است (۲/۰۶) در نتیجه، نتیجه می گیریم که میزان بیکاری از آن چه وزیر گفته بیشتر است. اما اگر Z_C منفی شود آنگاه نتیجه می گیریم که آمار وزیر اشتباه است اما میزان بیکاری جامعه کمتر از میزان اعلام شده از سوی وزیر است.
- مثال برای فاصله ای و استفاده از آزمون تی استیودنت: مثلاً گفته شده که مردم به طور متوسط در طول روز ۱۰ دقیقه اخبار گوش می کنند ⇦ ۱۰= μ

فرض کنیم از ۲۰ نمره نمونه گرفته ایم که در جدول زیر آمده است:

تماشای خبر (x_i)	F_i	$F_i x_i$	$\bar{x}$	$F_i {x_i}^2$
۵	۲	۲×۵ = ۱۰	۹/۸	۲×۲۵ = ۵۰
۷	۳	۷×۳ = ۲۱	۹/۸	۳×٤۹ = ۱٤۷
۸	۵	۸×۵ = ٤۰	۹/۸	۵×٦٤ = ۳۲۰
۱۰	۵	۱۵×۵ = ۷۵	۹/۸	۵×۱۰۰ = ۵۰۰
۱۵	۵	۱۰×۵ =۵۰	۹/۸	۵×۲۲۵ = ۱۱۲۵
کل	۲۰	۱۹٦		۲۱٤۲

$$\bar{x} = \frac{\sum F_i x_i}{N} = \frac{196}{20} = 9.8$$

$$S^2(V) = \frac{\sum n_i x_i - \frac{(\sum n_i x_i)^2}{n}}{N/_{N-1}} = \frac{\sum x_i (x_i - \bar{x})^2}{N/_{N-1}} = \frac{2.42 - \left(\frac{196}{20}\right)^2}{19} = 11.64$$

$$t_c = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{N}}} = \frac{9.8 - 10}{\sqrt{\frac{11.64}{20}}} = -0.26$$

براساس خطای ۰/۰۵ درصد و با درجه آزادی ۱۹ از جدول ص ۵۲۲ ← t_t = ۲/۰۹۳

$|-0/26| = 0/26 < 2/093$

⇦ تفاوت معنادار نیست و H_0 تأئید و H_1 رد می شود. یعنی به نظر می رسد میزان ملاحظه اخبار توسط مردم با آنچه آقای ضرغامی اعلام کرده برابر است.

H_1: به نظر می رسد میزان ملاحظه اخبار با آنچه آقای ضرغامی اعلام کرده تفاوت دارد. که این فرضیه رد شد.

اگر نتیجه به گونه ای باشد که t_c بیشتر از t_t بود و ۲ میزان تفاوت داشتند آن گاه علامت + یا – t_c محاسبه شده مهم بود به این صورت که اگر t_c مثبت بود، یعنی میزان تماشای اخبار توسط مردم از آنچه آقای ضرغامی گفته بیشتر است و اگر t_c منفی بود میزان تماشای اخبار توسط مردم از آنچه آقای ضرغامی گفته کمتر است.

جلسه ششم ۹۳/۸/۲۵

اگر بخواهیم ۲ میزان مستقل را با هم مقایسه کنیم، مثلاً میزان دریافتی معلمان با میزان دریافتی اعضای شرکت نفت باید از آزمون t استفاده کنیم.

- در این‌جا مهم است که آیا واریانس ها برابر است یا نه؟

بنابراین آزمون ۲ نمونه ای ۲ مستقل میانگین ها برای برابری واریانس‌ها به شرح زیر است:

$D.F = n_1 + n_2 - 2$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right)\left(\frac{n_1+n_2}{n_1 n_2}\right)}}$$

ردیف	xi_2^2	xi_1^2	نفت x_2	معلمان x_1
۱	۲۲۵۰۰۰۰	۱۰۰۰۰۰۰	۱۵۰۰	۱۰۰۰
۲	۴۰۰۰۰۰۰	۱۲۱۰۰۰۰	۲۰۰۰	۱۱۰۰
۳	۴۸۴۰۰۰۰	۸۱۰۰۰۰	۲۲۰۰	۹۰۰
۴	۱۶۰۰۰۰۰۰	۲۲۵۰۰۰۰	۴۰۰۰	۱۵۰۰
۵	۲۰۲۵۰۰۰۰	۴۰۰۰۰۰۰	۴۵۰۰	۲۰۰۰
۶	۶۲۵۰۰۰۰	۲۸۹۰۰۰۰	۲۵۰۰	۱۷۰۰
۷	۲۷۶۰۰۰۰	۳۲۴۰۰۰۰	۴۶۰۰	۱۸۰۰
۸	۴۴۱۰۰۰۰	۳۶۱۰۰۰۰	۲۱۰۰	۱۹۰۰

۹	۲۵۰۰۰۰۰۰	۹۰۰۰۰۰۰	۵۰۰۰	۳۰۰۰
۱۰	۳۰۲۵۰۰۰۰	۹۶۱۰۰۰۰	۵۵۰۰	۳۱۰۰
۱۱	۱۳۴۴۱۰۰۰	۳۷۶۲۰۰۰۰	۳۳۹۰۰	۱۸۰۰۰
میانگین	۱۳۴۴۱۰۰۰۰	۳۷۶۲۰۰۰۰	۳۳۹	۱۸۰۰

$$v = \frac{x_i^2 - \frac{(\sum x_i)^2}{n}}{N}$$

$$v_1 = \frac{37620000 - \frac{(1800)^2}{10}}{10} = 522000 \Rightarrow v_1 = 522000$$

$$v_2 = 1948900$$

اگر S_1 از ۳ برابر S_2 بیشتر باشد، واریانس ها نابرابراند. در نتیجه باید از فرمول واریانس های نابرابر استفاده کنیم. البته در اینجا ما از فرمول برابری استفاده می کنیم به جهت آموزش آن.

$$t = \frac{1800 - 3390}{\sqrt{(\frac{(10-1)522000 + (10-1-1)1948900}{10+10-2})(\frac{10+10}{10*10})}} = \frac{-1590}{157/19} = -10/11$$

$$t_c = -10/11$$

d.f=10+10-2=18

$$t_t|\alpha = 0.05\,, d.f = 18 \Rightarrow 2.1009$$

$\Rightarrow |t_c| > t_t \Rightarrow$

تفاوت معنادار است و ۱۵۹۰ تومان به طور متوسط حقوق شرکت نفت از معلمان بیشتر است.

* در اینجا می خواهیم از فرمول نابرابری واریانس ها استفاده کنیم و همین مثال را از فرمول نابرای حل کنیم.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}}$$

فرمول محاسبه درجه آزادی در واریانس های نابرابر $D.F = \frac{[\left(\frac{s_1^2}{n_1-1}\right)+\left(\frac{s_2^2}{n_2-1}\right)]^2}{[\left(\frac{s_1^2}{n_1-1}\right)^2(\frac{1}{n_1-1})]+[\left(\frac{s_2^2}{n_2-1}\right)^2(\frac{1}{n_2-1})]}$

$$t = \frac{1800 - 3390}{\sqrt{\frac{522000}{10 - 1} + \frac{1948900}{10 - 1}}} = -3.03$$

بنابراین از این روش نیز تفاوت زیاد حقوق شرکت نفت نسبت به معلمان حاصل شد.

- **انواع فرضیه از نظر جهت**

۱. فرضیه جهت دار

۲. فرضیه بدون جهت

فرضیه جهت دار: به نظر می رسد میزان متوسط درآمد کارکنان شرکت نفت بیشتر از متوسط درآمد معلمان است.

فرضیه بدون جهت: به نظر می رسد میزان متوسط درآمد کارکنان شرکت نفت و معلمان متفاوت است.

اکنون برای در نظر گرفتن t_t از جدول اگر فرضیه ما جهت دار است باید نصف t_t را در نظر بگیریم یعنی اگر $t_t = 2/179$ باشد و فرضیه جهت دار باشد باید $\frac{t_t}{2} = 1/08$ را حساب کنیم.

جلسه هفتم ۹۳/۹/۲

۳- حالت سوم (وابسته یا جفت شده)

انواع تحقیقات از لحاظ زمان:

۱. مقطعی

۲. طولی که شامل :

- روند پژوهشی (تغییرات ناخالص)

- پانلی (تغییرات خالص)

- نسبتی

فرض کنیم می خواهیم اثر یک فیلم را بر یک مجموعه بسنجیم. ابتدا از افراد یک پرسشنامه را پر می کنیم. سپس فیلم را پخش می کنیم و بعد از همان فردی که قبلاً سوال شده همان پرسش نامه قبلی را پر می کنیم. به این روش پانلی می گویند. اما اگر افراد تغییر کنند و قبل از پخش از یک سری افراد پرسش نامه پر شود و بعد از پخش از افراد دیگر آنگاه روش ما، روند پژوهشی است.

مثال: می خواهیم اثر دیدن فیلم رادر افراد بسنجیم. اما در هر فرد نسبت به قبل از دیدن فیلم و بعد از دیدن فیلم. در این جا بر خلاف قبلاً که میانگین حقوق معلمان را با میانگین حقوق شرکت نفت مقایسه کردیم، محاسبات هر فرد نیاز است.

$$t = \frac{\bar{d}}{\sqrt{\dfrac{\sum d^2 - \dfrac{\left(\sum d\right)^2}{n}}{n(n-1)}}}$$

$$D.F = N - 1$$

ردیف	di^2	di	اجرای مرحله اول	اجرای مرحله دوم
۱	۳۵	۵	۱۵	۱۰
۲	۴	۲	۱۰	۸
۳	۴	-۲	۱۰	۱۲
۴	۲۲۵	۱۵	۲۰	۵
۵	۱۹۶	۱۴	۱۵	۱
جمع	۴۵۴	۳۴		

$$\bar{d} = \frac{\sum d_i}{N} = \frac{34}{5} = 6.8$$

$$t = \frac{6.8}{\sqrt{\frac{454 - \frac{34^2}{5}}{5(5-1)}}} = \frac{6.8}{3.3} = 2.04, d.f = 5 - 1 = 4 \Rightarrow t_t = 2.77 \Longrightarrow t_c < t_t$$

یعنی این برنامه اثر گذاشت اما اثر آن و تفاوتش معنی دار نیست. به بیان دیگر H_1 رد می شود و H_0 تأیید می شود.

جلسه هشتم ۹۳/۹/۹

جلسه ۹۳/۹/۹

t
- تک نمونه ای
  - میانگین
  - نسبت
- دو نمونه
  - دو میانگین
  - دو نسبت

$$z = \frac{p_1 - p_2}{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$\text{نسبت} = \frac{\text{میزان مبدأ} - \text{میزان مقصد}}{\text{میزان مبدأ}}$$

قیمت گوشت زمان   قیمت گوشت زمان   سال

	خاتمی	هاشمی	
	۱۰۰۰	۲۰۰	۱
	۱۲۰۰	۲۲۰	۲
	۱۳۰۰	۲۵۰	۳

خاتمی   هاشمی

$$\frac{1200-1000}{1000}=0/2 \qquad \frac{220-200}{200}=0/1$$

$$\frac{1300-1200}{1200}=0/08 \qquad \frac{350-220}{220}=0/13$$

$$\frac{0/28}{2}=0/14 \qquad \frac{0/23}{2}=0/115$$

برای معناداری این تفاوت باید در ادامه . . . :

$$z=\frac{0/115-0/14}{\sqrt{\frac{0/115\times0/885}{200}+\frac{0/14\times0/86}{200}}}=\frac{-0/025}{\sqrt{\frac{0/1}{200}+\frac{0/12}{200}}}=\frac{-0/025}{0/03}=-0/83$$

فرض کنیم نمونه ۲۰۰ است $q=1-p$

$$df=n_1+n_2-2 \ \rightarrow df=398 \ \Rightarrow \underset{t}{\text{جدول}} \text{ یا } \underset{z}{\text{جدول}}=1/96$$

فرضیه H_0 تاییده شده و تفاوت معنا داری وجود ندارد.

متغیرهای اسمی

Exp: میزان استفاده از روزنامه در دو شهر متفاوت را با هم مقایسه کنید.

Exp: میزان بیکاری در دو شهر متفاوت را با هم مقایسه کنید.

بیکار

$$A:\ 600 \leftarrow 30$$
$$B:400 \leftarrow 25$$

$$P_1=\frac{\text{بیکاری}}{\text{کل نمونه}}=\frac{300}{600}=\frac{1}{20}=0/05$$

$$P_2=\frac{25}{400}=0/0625$$

$$q_1=0/95 \quad ,q_2=0/9375$$

$$z=\frac{0/05-0/0625}{\sqrt{\frac{0/05\times 0/95}{600}+\frac{0/0625\times 0/9375}{400}}}=\frac{-0/0125}{\sqrt{\frac{0/0475}{600}+\frac{0/0585938}{400}}}=\frac{-0/0125}{\sqrt{0/000792+0/0001465}}=-0/831$$

$$df=n_1+n_2-2=600+400-2=998 \Rightarrow Z_{\text{جدول}}=1/96$$

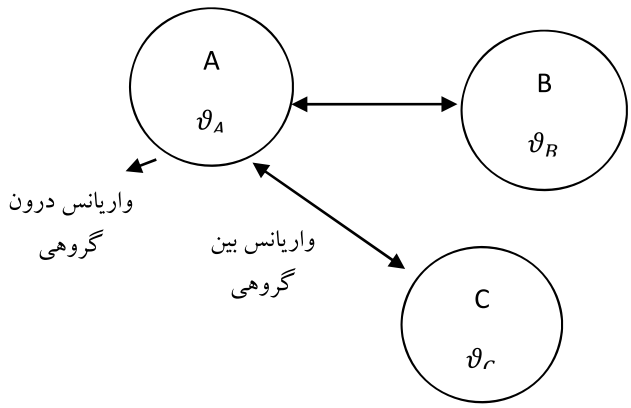
$$\propto=0/05$$

یعنی قبول H_0 و تفاوت معنا داری بین دو شهر وجود ندارد.

جلسه نهم ۹۳/۹/۱٦

اگر بیش از ۳ گروه داشته باشیم، می شود از آزمون t به صورت ۲به۲ استفاده کرد که اما از آن جایی که در هر بازمقایسه حدود ۵ مورد خطا وجود دارد و اگر مقایسه ها زیاد شود خطاها زیاد می شود. در نتیجه، از آزمون f استفاده می کنیم.

آزمون f در حاصل تقسیم واریانس بین گروهی بر درون گروهی به دست می آید. پس کل آزمون f محاسبه واریانس است.



شرط های آزمون f

۱-متغیر وابسته ما فاصله ای باشد.

۲-متغیر مستقل بیش از ۲ گروه باشد.

۳-متغیر وابسته تقریباً نرمال باشد.

۴-واریانس بین گروه ها تقریباً برابر باشد.

*شرط های اول و دوم با نگاه کردن تشخیص داده می شود.

*شرط سوم: باید کشیدگی و چولگی را باید محاسبه کنیم واگر کمتر از یک شده تقریباً نرمال شد.

*شرط چهارم: زمانی واریانس بین گروه ها تقریباً برابر است که واریانس هیچ کدام از گروه ها ۲ برابر گروه دیگر نباشد باید این ترتیب اگر واریانس یک گروه ۲ برابر یا بیشتر از گروه دیگر باشد. آن گاه واریانس های بین گروهی تقریباً برابر است

واریانس بین گروهی: Sb

واریانس درون گروهی:SW

$$F = \frac{Sb}{Sw} \text{ یا } Sb = \frac{SSb}{h-1}$$

$$V = \frac{\sum(x_i - \bar{x})^2}{n}$$

مثال:

A_i	B_i	C_i	A_I^2	B_I^2	C_I^2
۲	۷	۱	۴	۴۹	۱
۵	۸	۲	۲۵	۶۴	۴
۴	۹	۳	۱۶	۸۱	۹
۷	۵	۲	۴۹	۲۵	۴
۹	۶	۱	۸۱	۳۶	۱
جمع ۲۷	۳۵	۹	۱۷۵	۲۵۵	۱۹

اگر جدول بالا نمرات ۳ گروه دانشجویان و میانگین نمرات باشد مطلوب است مقایسه نمرات این دانشجویان و تعیین گروه برتر.

فرضیه سازی در آزمون F:

H_0= به نظر می رسد بین میانگین نمرات دانشجویان گروه های A, B, C تفاوت وجود ندارد.

H_1 = به نظر می رسد، حداقل بین میانگین نمرات دانشجویان ۲ گروه B, A و C تفاوت وجود دارد.

*** محاسبه حالت های مقایسه بین چند گروه** = $\frac{n(n-1)}{2}$

تعداد مجموعه ها ***n:***

مثلاً در این مثال $\frac{3(3-1)}{2} = 3$ یعنی ۳ حالت مقایسه داریم Aبا B، Bبا C و Aبا C

واریانس کل

$$S_t^2 = \frac{SS_t}{N-1}$$

مجموع مجذورات کل

$$SS_t = \sum_i^w x_i^2 - \frac{(\Sigma_i^w x_i)^2}{N}$$

واریانس درون گروهی

$$S_w^2 = \frac{SS_w}{N-1}$$

$$SS_b = \Sigma \frac{(\Sigma x_k)^2}{nk} - \frac{(\Sigma x_i)^2}{N(\text{کل})}$$

تعداد گروه ها: ***K***

$$\Longrightarrow SS_b = \left[\frac{(27)^2}{5} + \frac{(35)^2}{5} + \frac{(9)^2}{5}\right] - \frac{(35+27+9)^2}{15}$$

$$= \left[\frac{729}{5} + \frac{1225}{5} + \frac{18}{5}\right] - \frac{(71)^2}{15}$$

$$\left[\frac{1972}{5}\right] - \frac{5041}{15} = \frac{5916+1972}{15} = \frac{875}{15} = 58 / 33$$

$\Longrightarrow SS_b=58/33$

$S_b = \frac{SS_b}{\frac{n-1}{k}} = \frac{58/33}{4} = 29 / 16$

مجموع مجذورات درون گروهی

$$SS_w = \Sigma_{\mathrm{i}}^{\mathrm{w}} x_i^2 - \Sigma_{\mathrm{i}}^{\mathrm{w}} \frac{(x_k)^2}{n_k}$$

واریانس بین گروهی

$$S_b^2 = \frac{SS_b}{n-1}$$

مجموع مجذورات بین گروهی

$$SS_b = \Sigma_{\mathrm{i}}^{\mathrm{w}} \frac{(\Sigma x_k)^2}{n_k} - \frac{(\Sigma xi)^2}{N}$$

منبع تغییرات	مجموع مجذورات	درجه آزادی	واریانس ها
بین گروهی	مجموع مجذورات بین گروهی	*K-1*	S_{b}^2
درون گروهی	مجموع مجذورات درون گروهی	*N-K*	S_{w}^2
کل	مجموع مجذورات کل	*N-1*	S_{t}^2

$$SS_w = (175 + 255 + 19) - \frac{1972}{5}$$

449-394/4=54/6

$$\Longrightarrow SS_w = 54/6$$

$$S_w = \frac{SS_w}{N-K} - \frac{54/6}{15-3} = 4/55$$

منبع	مجموع مجذورات	درجه آزادی	واریانس	آزمون *F*
بین گروهی	۵۸/۳۳	*K*-۱=۳-۱=۲	۲۹/۱۶	

درون گروهی	۵۴/۱۶	N-K=۱۵-۳=۱۲	۴/۵۵	$\frac{29/16}{4/55} = 6/458$
کل	۱۱۲/۹۳			

$\Longrightarrow F = 6/408$

مجموع مجذورات کل

از جدول ۲ عدد می توان یافت | با ۱ درصد خطا

برای F_t | و با ۵ درصد خطا

برای مثال فوق :

برای : $\propto = 0/01$ $\qquad F_t\left(\frac{d.f_1}{2}, \frac{d.f_2}{12}\right) = 2\ /\ 81$

برای $\propto = 0/05$ $\qquad F_t\left(\frac{d.f_1}{2}, \frac{d.f_2}{12}\right) = 3\ /\ 89$

بنابراین در مثال چون f_c بزرگ تر از f_t (جدول) در ۵ درصد و یک درصد است فرض H_0 هم با احتمال ۹۹ درصد و هم ۹۵ درصد رد می شود و فرض H_1 تأئید می شود.

*اما تا این جا فقط می توانیم بگوئیم حداقل یک تفاوت وجود دارد بین گروه ها، و نمی دانم بین کدام گروه ها تفاوت داریم.

آزمون F توانایی این را ندارد که بگوید کجا تفاوت وجود دارد.

*حال اگر یکی از شرایط ۴گانه آزمون F فراهم نبود، مثلاً توزیع نرمال نبود چه می کنیم؟

از آزمون LSD استفاده می کنیم.

آزمون *LSD*

$$S_d = \sqrt{\frac{2SMw}{n}}$$

M_w=58/33+54/6=112/93

$$S_{\mathrm{d}} = \sqrt{\frac{2 \times 112/93}{15} = \sqrt{15 / 05 =}}\, 3 / 87$$

$\bar{x}_a = 5/4 \quad \bar{x}_B - \bar{x}_A = 7 - 5/4 = 1/6$

$\bar{a}_a = 7 \qquad \bar{x}_B - \bar{x}_c = 7 - 1/8 = 5/2$

$\bar{x}_c = 1/8 \quad \bar{x}_A - \bar{x}_c = 5 / 4 - 1/8 = 3/6$

برای این جا باید S_d را با تفاوت میانگین ها مقایسه کنیم. اگر اختلاف دو به دوی میانگین هر گروه از میزان S_d کمتر باشد تفاوت معنا دار نیست مثلاً بین دو گروه AوB تفاوت معناداری نیست اما بین گروه B و C چون ۵/۲ از $S_d = 3 / 87$ بزرگ تر است تفاوت قطعاً معنادار است و درباره تفاوت AوC نیز چون عدد نزدیک به ۳/۸۷ است می توان گفت احتمالاً تفاوت معنادار است.

*** نکته کنکوری دکتری** $f=t^2$

جلسه دهم – جلسه آخر ٩٣/٩/٢٣

شرایط آزمون t و F:

١- توزیع متغیر وابسته تقریباً نرمال باشد.

٢- واریانس بین طبقات متغیر مستقل تقریباً برابر باشد.

٣- متغیر وابسته فاصله ای

۴- متغیر مستقل دو شقی باشد (از t) و اگر چند شقی باشد (از F استفاده می کنیم).

* **آزمون های پارامتریک**: آزمون هایی که توزیع دارند و دارای متغیر فاصله ای هستند. با این تعریف آزمون های F و t یعنی آزمون هایی که مقایسه میانگین می کنند پارامتریک هستند و آزمون هایی که توزیع ندارد (متغیر فاصله ای نباشد) قطعا ناپارامتریک است و آزمون توزیع هایی که توزیع دارند و فاصله نباشند برخی می گویند پارامتریک و برخی ناپارامتریک.

*آزمون کروسکال والیس

١- **مبنای محاسباتی آزمون χ^2 :** تفاضل فراوانی مشاهده شده و مورد انتظار

٢- **مبنای محاسبه آزمون های ترتیبی:** تفاضل جفت های موافق و مخالف

٣-**آزمون های مقایسه میانگین (t و f):** تفاضل میانگین هاست

*** در آزمون کروسکال والیس مبنای محاسباتی، تفاوت رتبه هاست**

$$H_{\chi^2} = H = \frac{12}{N(N-1)} \Sigma_{\alpha=1}^{w} \left[\frac{R_j^2}{N_j}\right] - 3(N-1)$$

تعداد گروه و نه فراوانی = *j*

$$H = \frac{12}{\text{تعداد کل}(\text{تعداد کل} - 1)} \times \text{مجموع} \left[\frac{\text{رتبه های گروه ها}}{\text{تعدادهمان گروه}}\right] - 3\left(\text{تعداد} - 1\right)$$

A	*AR* (نمرات را مرتب می کنیم و در نتیجه گروه های *B* و *C* نیز تغییر می کند)	رتبه	*B*	رتبه *B* براساس *A*	رتبه	*C*	رتبه *C* بر اساس *C*	رتبه
۲	۲	۱	۷	۷	۳	۱	۱	۱/۵
۵	۴	۲	۸	۹	۵	۲	۳	۵
۴	۵	۳	۹	۸	۴	۳	۲	۳/۵
۷	۷	۴	۵	۵	۱	۲	۲	۳/۵
۹	۹	۵	۶	۶	۲	۱	۱	$\frac{(2+1)=3}{2} = 1.5$

در ستون *C* برای رتبه اعداد یک چون ۲ عدد یک داریم و یکی رتبه اول و دیگری رتبه دوم است در نتیجه اعداد ۲ و یک یعنی رتبه های ۲ عدد یک را با هم جمع و تقسیم بر ۲ می کنیم به این ترتیب رتبه اعداد یک هم کلاً ۱/۵ می شود.

رتبه *A*	رتبه *B*	رتبه *C*	تفاضل *A-B*	تفاضل *A-C*	تفاضل *B-C*	$(A-B)^2$	$(A-C)^2$	$(B-C)^2$
۱	۳	۱/۵	−۲	۰۰/۵	+۱/۵	۴	۰/۲۵	۲/۲۵
۲	۵	۵	−۳	۳−	۰	۹	۹	۰

۰/۲۵	۰/۲۵	۱	+۰/۵	−۰/۵	−۱	۳/۵	۴	۳
۶/۲۵	۰/۲۵	۹	−۲/۵	+۰/۵	+۳	۳/۵	۱	۴
۰/۲۵	۱۲/۵	۹	+۰/۵	+۳/۵	+۳	۱/۵	۲	۵
۹	۲۲	۳۲	۰	۰	۰	۱۵	۱۵	۱۵

اکنون میانگین رتبه ها را محاسبه می کنیم.

۱/۸= ۵÷۹ و ۴/۴= ۵÷۲۲ و ۶/۴=۵÷۳۲

$$H = \frac{12}{15(15-1)} \times [(6 / 4 + 4 / 4 + 1 / 8)] - 3(15-1) =$$

$\Longrightarrow H= 0/72-42=41/28$

$\Longrightarrow Hc = -41 / 28$

H_t $(\propto=0/05, f=14) =21/064$

$\Longrightarrow |H_c| > H_t \Rightarrow$ تفاوت معنا دار است

و H_0 رد می شود و H_1 تأیید می شود و گروه A با گروه C تفاوت معنادار دارد.

*اگر تعداد گروه ها نابرابر باشد، مثلاً گروه A ۴ نفر و گروه B ۶ نفر باشد، آنگاه رتبه ها را به توان ۲ می رسانیم و به جای تفاضل ها استفاده می کنیم. اما اگر تعداد گروه ها برابر باشند همین تفاضل استفاده می کنیم.

– اگر بخواهیم از مجذور رتبه ها استفاده کنیم باید مشابه تفاضل ها میانگین مجذورها را نیز حساب کنیم و آنگاه در فرمول بگذاریم.

* نکته: اگر در آزمون کروسکال والیس گره داشته باشیم از فرمول اصلاح شده آن استفاده می کنیم.

t: تعداد گره ها که مثال ۲ است. $T= t^2- t$

گره: رتبه مشترک ها باید $\frac{H}{T}$ را حساب کنیم.

تفسیر۴۱-؛ منفی یعنی تفاضل Aیا C زیاد است. چرا A را مثبت می دانم چون A مبنای رتبه گذاری بود. از طرفی میانگین تفاضل A-B مساوی ۶/۴ است، و میانگین تفاضل B-C نیز ۱/۸ است پس تفاوت بین C,A است.

امتحان ۴ سوال ۱۲ نمره امتحان – ۸ نمره تکلیف

*هر کدام از طریق یک آزمون قابل حل است.

*با توجه به سطح متغیر باید تشخیص دهم از کدام آزمون استفاده کنیم.

*جدول توزیع آزمون ها همراه مان باشد.

-فرضیه ها را باید بنویسیم؛آزمون مربوط را باید انجام دهیم؛ تعداد آزمون باید محاسبه شود؛ مقدار تفاوت ها بیان شود؛ قبول یار و فرضیه ها بیان شود ؛ تفسیر هر سوال.