

1696

دعای پیش از مطالعه

اَللهُمَ اَخرِجنی مِن ظُلُماتِ الوَهمِ وَاَکرِمنی بِنُورِ الفَهمِ
اَللهُمَ افتِح عَلَینا اَبوابَ رَحمتَکَ وَ انشُر عَلَینا خَزائِنِ عُلُومِکَ
بِرَحمَتِکَ یا اَرحَمَ الرّاحِمینَ

پروردگارا، خارج کن مرا از تاریکی‌های فکر و گرامی بدار به نور فهم
پروردگارا، بگشای بر ما درهای رحمتت را و بگستران گنج‌های دانشت را به امید رحمت
تو ای مهربان‌ترین مهربانان

بیایید به حقوق دیگران احترام بگذاریم

دوست عزیز، این کتاب حاصل دسترنج چندین ساله‌ی مؤلف، مترجم و ناشر آن است. تکثیر و فروش آن به هر شکلی بدون اجازه از پدیدآورنده کاری غیراخلاقی، غیرقانونی، غیرشرعی و کسب درآمد از دسترنج دیگران است، نتیجه‌ی این عمل نادرست، موجب رواج بی‌اعتمادی در جامعه و بروز پی‌آمدهای ناگوار در زندگی و محیط ناسالم برای خود و فرزندانمان می‌گردد.

آناهید گستر خلیلی

AGK

Anahid Gostar Khalili

نکات داغ آمار حیاتی

ویژه کارشناسی ارشد رشته‌ی اقتصاد بهداشت

تألیف و گردآوری:

امین علیزاده

«رتبه 14 کارشناسی ارشد اقتصاد بهداشت دانشگاه علوم پزشکی ایران»

سرشناسه	: علیزاده، امین، 1368-
عنوان و نام پدیدآور	: AGK نکات داغ آمار حیاتی ویژه کارشناسی ارشد رشته‌ی اقتصاد بهداشت/ تألیف و گردآوری امین علیزاده.
مشخصات نشر	: تهران: آناهید گستر خلیلی، 1404.
مشخصات ظاهری	: 68ص.: جدول، نمودار؛ 14/5×21/5 س‌م.
شابک	: 978-622-4744-58-6
وضعیت فهرست‌نویسی	: فیپا
موضوع	: آمار حیاتی -- آمار -- راهنمای آموزشی -- مسائل، تمرین‌ها و غیره (غیره) Vital statistics -- Study and teaching -- Problems, exercises, etc (Higher)
رده‌بندی کنگره	: HA35
رده‌بندی دیویی	: 001/422076
شماره کتابشناسی ملی	: 10005307

انتشارات آناهید گستر خلیلی

نام: AGK نکات داغ آمار حیاتی

تألیف و گردآوری: امین علیزاده

ناشر: آناهید گستر خلیلی

نوبت و سال چاپ: اول . 1404

شمارگان: 1000

چاپ و صحافی: گیلان

مدیر تولید: امیرحسین خلیلی

مدیر فنی و هنری: مریم آرده

تایپ و صفحه‌آرایی: آرزو راضی

بهاء: 110000 تومان

آموزشگاه/ فروشگاه: 02166568621

مرکز پخش: ضلع جنوب‌غربی میدان انقلاب . جنب بانک پارسیان . مجتمع تجاری پارس . طبقه اول

مدیر فروش: 09191685970 (ارتباط از طریق پیامک)

agkpublisher@gmail.com **www.agkins.com**

طلیعه سخن مدیر تولید:

سپردن کار به اهلش نشانه‌ی خردمندی است.

برای تولید بسته‌های آموزشی تمام تلاش خود را به کار گرفته‌ایم تا محصولی شایسته را تقدیم شما عزیزان نماییم. مجموعه‌ای که پیش روی شماست براساس معتبرترین و به‌روزترین منابع کنکور می‌باشد و به قلم مجرب‌ترین اساتید تألیف شده است و مطالعه آن موجب صرفه‌جویی در زمان و هزینه شما خواهد شد.

ارائه بالاترین آمار قبولی در کنکور کارشناسی ارشد و دکتری در دهه اخیر بیان‌گر توانایی این گروه در راستا اهداف خود می‌باشد. اینک با اتکا به تجربه‌ای 17 ساله می‌توانیم ادعا کنیم که مجموعه حاضر، بی‌شک از جمله‌ی کامل‌ترین و تضمینی‌ترین بسته‌های آموزشی در عرصه کنکور می‌باشد.

برخی از ویژگی‌های بسته‌های آموزشی:

- منطبق با معتبرترین و به‌روزترین منابع کنکور
- گردآوری و تألیف توسط مجرب‌ترین اساتید کنکور
- تدریس مطالب از سطح پایه تا پیشرفته به بیان ساده و روان
- انتقال بهتر مفاهیم با استفاده از شکل، جدول و نمودار
- تألیف مبتنی بر نحوه‌ی طراحی سوالات کنکور
- مطالعه یک منبع منسجم به‌جای مطالعه چندین منبع مختلف
- منبعی خودآموز و بدون نیاز به مدرس

اینک علاوه بر مطالعه بسته‌های آموزشی استفاده از خدمات زیر را به عنوان ضمانتی برای قبولی به شما داوطلبان عزیز توصیه می‌کنم؛

- نکات داغ (Hot Points)

این‌بار زمان به نفع شما در حال گذر است؛ دیگر نیازی به خلاصه‌نویسی برای دوران مرور و جمع‌بندی نخواهید داشت ...

اکثر داوطلبان پس از مطالعه‌ی منابع برای خلاصه‌نویسی دچار سردرگمی می‌شوند یا به خلاصه‌های خود اطمینان ندارند، حتی گاهی حجم خلاصه‌ها بیش از حجم منبع می‌شود!!! گاهی هم به اشتباه تصور می‌کنند که نکات و خلاصه‌هایشان در چارچوب کنکور است یا نکاتی را که گلچین کرده‌اند مهم‌اند.

توجه کنیم که خلاصه‌نویسی مطالب برای کنکور دارای اصول و قاعده است که اغلب داوطلبان پس از مطالعه قادر به انجام صحیح آن نیستند. به همین منظور تمام تلاش خود را به‌کار گرفته‌ایم که استانداردترین منبع نکات را در چارچوب کنکور تقدیم شما نماییم.

هدف از تولید نکات داغ فراهم کردن یک منبع برای دوران مرور و جمع‌بندی است.

- ویژگی‌های نکات داغ:

برگرفته از معتبرترین و به‌روزترین منابع کنکور

مجموعه‌ای از نکات کنکوری به صورت طبقه‌بندی موضوعی

پوشش قابل ملاحظه سوالات کنکور در کم‌ترین حجم ممکن

مناسب‌ترین منبع برای دوران مرور و جمع‌بندی

- آزمون‌های آزمایشی: خوش بود گر محک تجربه آید به میان

داوطلب شجاع کیست؟ داوطلبی است که خود را مورد سنجش و ارزیابی قرار دهد تا پیش از آزمون سراسری به نقاط ضعف خود پی برده و آن‌ها را تقویت نماید. صادقانه به شما توصیه می‌کنیم که در آزمون‌های آزمایشی شرکت کنید تا بتوانید خود را با خود (سنجش) و با دیگران (ارزیابی) مقایسه و نقاط ضعف و قوت خود را شناسایی نمایید.

- مشاوره و برنامه‌ریزی: برای موفقیت هوش کافی نیست، جسارت لازم دارد.

در صورتی که نمی‌توانید با توجه به زمان آزادتان برنامه‌ی مطالعاتی استاندارد برای خود تدوین کنید حتماً از افراد با تجربه که خود در این مسیر موفق بوده‌اند کمک بخواهید. افرادی که تجربیات خود را صادقانه در اختیار شما قرار می‌دهند و برنامه‌ریزی مناسب با شرایط زمانی و ویژگی‌های روحی و جسمی مختص شما انجام می‌دهند.

- کتاب تست:

کوچک‌ترین واحد یادگیری در کنکور، تست است.

شما می‌توانید برای این‌که سطح یادگیری خود را پس از مطالعه بسنجید از کتاب‌های تست مجموعه (تست کنکور) و ماطراحان (تست‌های تألیفی) استفاده کنید. این کتاب‌ها به‌صورت موضوعی طبقه‌بندی شده‌اند و می‌توانند نحوه‌ی طرح سوالات کنکور و سطح آن‌ها را به‌صورت واضح به شما نشان دهند.

- درسنامه‌های نکات داغ و کتاب‌های الگوریتم مناسب‌ترین منابع برای دوران مرور و جمع‌بندی:

بیش‌تر داوطلبان حین مطالعه نگران خلاصه‌نویسی و مرور آموخته‌های خود هستند و برای خلاصه‌نویسی زمان زیادی را صرف می‌کنند. اینک به شما مژده می‌دهیم که این‌بار زمان به نفع شما در حال گذر است و دیگر با استفاده از درسنامه‌های نکات داغ و کتاب‌های الگوریتم نیاز به خلاصه‌نویسی ندارید.

شما می‌توانید پیشنهاد و یا انتقاد سازنده خود را به شماره تلفن همراه 09191685970 فقط و فقط پیامک نمایید.

مشاور و مدیر تولید

امیرحسین خلیلی

طلیعه سخن مؤلف:

شاید به جرأت بتوان گفت، درس آمار حیاتی در اکثر رشته‌هایی که از این درس سوال می‌دهند، یک تنه می‌تواند باعث کسب یک نتیجه ایده‌آل برای شما شود.
اما این درس اولاً سخت است و دوماً منابع معرفی شده معمولاً حجیم و زمان‌بر هستند، به همین دلایل تصمیم گرفتیم تا با تألیف این کتاب کمک‌آموزشی که حاصل بررسی کنکورهای اخیر و منابع معرفی شده می‌باشد در راستای سرعت بخشیدن به یادگیری و کسب نتیجه مطلوب شما عزیزان است گامی مهم برداریم.
برای این‌که از این کتاب بهتر نتیجه بگیرید ما راهکار زیر را به شما پیشنهاد می‌دهیم:
شما ابتدا نکات داغ را برای هر فصل مطالعه کنید و بلافاصله بعد از اتمام مطالعه نکات هر فصل اقدام به تست زدن کنید و پاسخ‌های تشریحی را به دقت مطالعه کنید و سپس در زمان‌های مختلف بعدی دوباره تست بزنید و پاسخ‌های تشریحی و نکات داغ را مرور کنید.
نگران کتاب تست آمار حیاتی نباشید چون‌که به زودی کتاب‌های تست با پاسخ‌های کاملاً تشریحی منتشر می‌کنیم. امیدواریم که این کتاب کمک‌آموزشی بتواند به شما کمک کند تا در مسیر تحقق اهداف بهترین‌ها را برای خودتان رقم بزنید.
در پایان جا دارد از همکاران گرامی جناب آقای امیرحسین خلیلی و سرکار خانم مریم آرده که با زحمات شبانه‌روزی خود در انتشارات آناهید گستر خلیلی، اینجانب را در تألیف این کتاب کمک‌آموزشی صبورانه یاری رساندند، کمال تقدیر و تشکر صمیمانه را داشته باشم و برای ایشان در تمام مراحل زندگی آرزوی بهترین‌ها را دارم.

با تشکر
امین علیزاده

فهرست مطالب

فصل و عنوان **صفحه**

فصل اول: آمار توصیفی 9

فصل دوم: آنالیز ترکیبی و احتمال 17

فصل سوم: متغیر تصادفی و تابع احتمال 22

فصل چهارم: توزیع‌های احتمال گسسته و پیوسته 28

فصل پنجم: قضایای حدی تئوری احتمالات 39

فصل ششم: نمونه‌گیری و توزیع‌های نمونه‌گیری 41

فصل هفتم: تئوری برآورد 45

فصل هشتم: آزمون فرض آماری 51

فصل نهم: تحلیل (آنالیز) واریانس 58

فصل دهم: همبستگی و رگرسیون 62

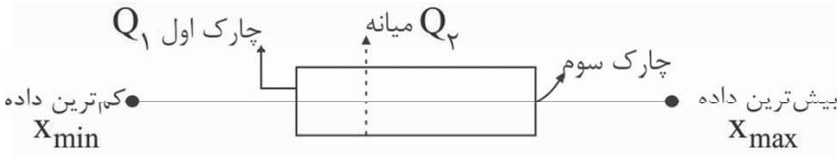
فصل

1

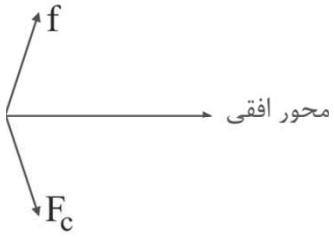
آمار توصیفی

- F_i : فراوانی مطلق/ f_i : فراوانی نسبی/ P_i : درصد فراوانی نسبی/ FC_i : فراوانی تجمعی
- X_i : مرکز دسته/ C: طول دسته
- نمودارها
  - نمودار بافت‌نگار یا هیستوگرام: برای رسم روی محور xها حدود طبقات با طول‌های مساوی و روی محور yها فراوانی مطلق یا فراوانی نسبی یا چگالی مطلق $(d_i = \frac{F_i}{C})$ و یا چگالی نسبی $(d'_i = \frac{f_i}{C})$
    - $x \Leftarrow$ حدود طبقات
    - $y \Leftarrow$ فراوانی یا چگالی
  - نمودار چند بر فراوانی یا پلی‌گون: محور xها مرکز طبقات و محور yها فراوانی مطلق یا نسبی هستند.
    - $x \Leftarrow$ مرکز طبقات
    - $y \Leftarrow$ فراوانی
  - نمودار میله‌ای یا ستونی: محور xها مرکز طبقات و محور yها فراوانی مطلق یا نسبی هستند.
  - نمودار فروانی تجمعی: محور xها حد بالای طبقات و محور yها فراوانی تجمعی هستند.
    - $x \Leftarrow$ حد بالای طبقات
    - $y \Leftarrow$ فراوانی تجمعی
  - نمودار دایره‌ای: سهم هر طبقه بر حسب درجه برابر است با: $S_i = f_i \times 360$

- نمودار شاخه و برگ: ارقام داده‌های آماری به دو بخش تقسیم می‌شوند. شاخه شامل یک یا چند رقم اولیه (مثل دهگان) و برگ شامل ارقام باقیمانده (مثل یکان).
- نمودار جعبه‌ای:



- نمودار پاره‌تو: محور افقی موضوع یا صفت مورد بررسی است، محور عمودی سمت چپ فراوانی است که به ترتیب نزولی ترسیم می‌شود و محور سوم یک محور عمودی در سمت راست است که بیان‌گر فراوانی نسبی تجمعی است.



مشخص کننده‌های عددی (پارامترها)

- میانگین حسابی

$$\bar{x}=\frac{\sum x_i}{N}$$ میانگین داده‌های خام / $$\bar{x}=\frac{F_1X_1+\ldots+F_kX_k}{F_1+\ldots+F_k}$$ میانگین در جدول فراوانی طبقه‌بندی نشده

مرکز طبقات

$$\bar{x}=\frac{\sum F_iX_i}{N}=\sum f_iX_i$$ میانگین در جدول فراوانی طبقه‌بندی شده

هر عدد دلخواه

$$\sum X_i=N\bar{X} \quad / \sum(X_i-\bar{X})=0 \quad / \sum(X_i-\bar{X})^2<\sum(X_i-a)^2$$

هر کدام یک مجموعه هستند

$$Z_i=X_i+Y_i \Rightarrow \bar{Z}=\bar{X}+\bar{Y} \quad / \bar{a}=a \quad / \overline{X_i+a}=\bar{X}+a$$

$$\overline{aX}=a\bar{X}$$

- میانگین هندسی: اگر داده‌های ما نسبت‌ها، درصدها، شاخص‌ها، نرخ رشد و... باشد
  - برای محاسبه متوسط آن‌ها از میانگین هندسی استفاده می‌کنیم و داریم:

$$\bar{X}_G=\sqrt[n-1]{\frac{\text{داده سال آخر}}{\text{داده سال اول}}}$$

  - و اگر بخواهیم متوسط درصد رشد سالانه را به‌دست آوریم داریم:

$$\bar{g}=(\bar{X}_G-1)\times 100$$

  - در میانگین هندسی اگر داده‌ها دارای وزن باشند داریم:

$$\bar{X}_G=\sqrt[n]{X_1^{w_1}.X_2^{w_2}.....X_n^{w_n}}$$

فصل

2

آنالیز ترکیبی و احتمال

- آنالیز ترکیبی: دسته‌بندی چیزها به گروه‌های چندتایی بر طبق قراردادهای خاصی است.
- اصل ضرب: اگر عملی به m راه بعد از آن عمل دیگری به n راه و پس از آن عمل دیگری به k راه مختلف صورت گیرد تعداد حالت‌هایی که این اعمال می‌توانند با هم انجام شوند برابر است با:

$m\times n\times k\times ...$

  - نشانه‌ی اصل ضرب حرف ربط «و» است. (مثل پیراهن و شلوار)
- اصل جمع: اگر هدفی را به دو روش a_1 و a_2 بتوان انجام داد به‌طوری که a_1 به m و a_2 به n طریق صورت گیرند تعداد راه رسیدن به هدف برابر $m+n$ است. (نشانه «یا») (مثل پیراهن و تی‌شرت)
- جایگشت یا تبدیل: به چندین طریق می‌توان n شی را در کنار هم قرار داد که برابر است با: $n!$
  - اگر بخواهیم n شی را در کنار هم قرار دهیم به‌طوری که r شی آن کنار هم باشند تعداد روش‌ها می‌شود: $r!(n-r+1)!$
  - تعداد روش‌هایی که می‌توان n شی را در کنار هم قرار داد به‌طوری که r شی آن کنار هم نباشند برابر است با: $n!-r!(n-r+1)!$
  - اگر m تا شی 1 تا 2 داشته باشیم و بخواهیم آن‌ها را یک در میان بچینیم داریم: 1- اگر $m=n$ باشد می‌شود $m!\times n!$(●●●●●) 2- اگر $m=n+1$ باشد می‌شود $\begin{pmatrix}\bullet\bullet\bullet\bullet\bullet\bullet\\ \bullet\bullet\bullet\bullet\bullet\bullet\end{pmatrix}$
  - اگر بخواهیم 5 مداد، 6 خودکار و 7 خودنویسی را به‌طوری کنار هم قرار دهیم که مدادها، خودکارها و خودنویس‌ها کنار هم باشد تعداد راه‌ها برابر است با: $3!\times(5!\times6!\times7!)$

قـدر مطلـق تفاضـل «میـانگین ایـن n متغیـر تصـادفی» از «میـانگین امیـدهای ریاضـی آن‌هـا» از عدد کوچک و مثبت ε بزرگ‌تر یا مساوی باشد برابر است با:

قضیه: $P\left(\left|\frac{\Sigma xi}{n}-\frac{\Sigma ai}{n}\right|\geq\varepsilon\right)=\frac{Varx}{n\varepsilon^2}$ عکس قضیه و: $P\left(\left|\frac{\sum xi}{n}-\frac{\sum ai}{n}\right|\leq\varepsilon\right)=1-\frac{Varx}{n\varepsilon^2}$

- قانون اعداد بزرگ به‌صورت چه بی‌شف: با توجه به نامساوی دوم چه بی‌شف داریم:

$$\lim_{n\to\infty}\left(P\left(\left|\frac{\Sigma xi}{n}-\frac{\Sigma ai}{n}\right|\geq\varepsilon\right)\right)=0 \quad / \quad \lim_{n\to\infty}\left[\left(\left|\frac{\Sigma xi}{n}-\frac{\Sigma ai}{n}\right|\leq\varepsilon\right)\right]=1$$

فصل

6

نمونه‌گیری و توزیع‌های نمونه‌گیری

- نمونه تصادفی: نمونه‌ای است که نحوه‌ی انتخاب آن به گونه‌ای باشد که هر یک از عناصر جامعه، شانس مشخص و غیرصفر برای انتخاب شدن در آن نمونه داشته باشد.
- **روش‌های نمونه‌گیری:**
  - نمونه‌گیری تصادفی ساده: زمانی استفاده می‌شود که جامعه شماره‌گذاری شده باشد یا به آسانی شماره‌گذاری شود و عناصر نمونه به آسانی قابل دسترس باشند (مثل قرعه‌کشی و...)
  - نمونه‌گیری منظم یا سیستماتیک: این‌گونه است که مثلا ابتدا یک عضو مثلا عضو دوم از جامعه انتخاب می‌شود و بعد از آن عناصر دیگر به صورت iتا، iتا مثلا 5تا، 5تا انتخاب می‌شوند. زمانی کاربرد دارد که نوعی ترتیب در جامعه وجود داشته باشد. این روش آسان و کم‌هزینه است اما مشکلش این است که چند عنصر پشت سر هم، در نمونه نمی‌آیند.
  - نمونه‌گیری طبقه‌بندی شده: ابتدا جامعه را طبقه‌بندی می‌کنیم، سپس نمونه‌ای تصادفی از هر طبقه انتخاب می‌نماییم. مزیت این روش:
    - از هر طبقه نمونه‌ای اندازه‌ی مناسب در نظر می‌گیرد.
    - دقیق‌ترین تخمین را دارد.
    - زمانی استفاده می‌شود که بعد از طبقه‌بندی طبق یک ویژگی خاص، اختلاف بین طبقات از اختلاف درون هر طبقه بیش‌تر باشد.
  - نمونه‌گیری خوشه‌ای: زمانی استفاده می‌شود که عناصر جامعه همگن باشد. در این روش ابتدا جامعه را به خوشه‌ها (گردهمایی) مختلف تقسیم می‌کنیم، بعد از میان این

حال به کمک جدول زیر که جدول آنالیز واریانس نام دارد به محاسبه‌ی آماره‌ی F می‌پردازیم و آن را با $F_{\alpha,K-1,n-K}$ مقایسه می‌کنیم.

منبع تغییرات	مجموع مجذور تغییرات	درجه آزادی	میانگین مجموع مجذور خطا (واریانسی)	
واقعی	SSR	$K-1$	$MSR=\frac{SSR}{K-1}$	$\Rightarrow F=\frac{MSR}{MSE}=\frac{\frac{SSR}{K-1}}{\frac{SSE}{N-K}}$
شانسی (خطا)	SSE	$N-K$	$MSE=\frac{SSE}{N-K}$	
کل	SST	$N-1$		

$$SST=\sum_{i=1}^{K}\sum_{j=1}^{n}x_{ij}^2-\frac{1}{N}T^2 \quad \text{و} \quad x_{ij}\ T=\sum_{i=1}^{K}\sum_{j=1}^{n}X_{ij}$$

$$SSR=\sum_{i=1}^{K}\frac{T_i^2}{n_i}-\frac{1}{N}T^2 \quad \text{و} \quad N=n_1+n_2+\ldots+n_K \Rightarrow N=nK$$

$SSE=SST-SSR$ و $SST=SSR+SSE \Rightarrow$ اثر کل = اثر تیماری + اثر خطا

آنالیز واریانس دو عاملی: در این‌گونه سوالات نمونه‌ها براساس دو ویژگی (دو عامل) طبقه‌بندی می‌شوند. در این صورت علاوه بر اثر تیماری و اثر خطا، یک اثر دیگر به نام اثر بلوکی (β_j) وجود دارد که با توجه به عامل دوم شکل می‌گیرد.

در این صورت هر مشاهده را می‌توان به صورت زیر نشان داد.

میانگین کل، اثر تیماری، اثر بلوکی، اثرخطا

$$X_{ij}=\mu+\alpha_i+\beta_j+e_{ij}$$

در این صورت برای آزمون ادعای میانگین‌ها برابر نیستند: H_1 باید به بررسی

$$\begin{cases}H_0:\mu_1=\mu_2=\ldots=\mu_k\\ H_1:\text{میانگین‌ها برابر نیستند}\end{cases}$$

دو ادعای زیر بپردازیم که برای این کار از جدول آنالیز واریانس دو عاملی بهره می‌گیریم:

$$\begin{cases}H_0:\alpha_1=\ldots=\alpha_k=0\\ H_1:\alpha_i \text{ برابر صفر نیستند}\end{cases} \qquad \begin{cases}H_0:\beta_1=\ldots=\beta_n=0\\ H_1:\beta_j \text{ برابر صفر نیستند}\end{cases}$$

منبع تغییرات	مجموع مجذورات SS	درجه آزادی	واریانس MS
اثرتیماری	$SSR = \frac{1}{n}\sum_{j=1}^{K} T_j^2 - \frac{1}{nk}T^2$	$K-1$	$MSR = \frac{SSR}{K-1}$
اثربلوکی	$SSB = \frac{1}{K}\sum_{j=1}^{n} T_j^2 - \frac{1}{nk}T^2$	$N-1$	$MSB = \frac{SSB}{n-1}$
اثر خطا	$SSE = SST - SSR - SSB$	$(K-1)(n-1)$	$MSE = \frac{SSE}{(n-14)(K-1)}$
اثر کل	$SST = \sum\sum x_{ij}^2 - \frac{1}{nk}T^2$	$nK-1$	

$F = \frac{MSB}{MSE} \Rightarrow$ با $F_{\alpha,(n-1),(n-1)(K-1)}$ مقایسه می‌شود.

$F = \frac{MSR}{MSE} \Rightarrow$ با $F_{\alpha,(K-1),(n-1)(K-1)}$ مقایسه می‌شود.

چنانچه هر دو ناحیه عدم رد قرار گیرند ادعای برابری میانگین جوامع تایید می‌شود و اختلاف بارزی بین میانگین‌ها وجود ندارد.

فصل

10

همبستگی و رگرسیون

- در این فصل به‌طور همزمان به بررسی دو متغیر می‌پردازیم و روابط بین این دو متغیر را مدنظر قرار می‌دهیم.
- ضریب همبستگی: درجه یا شدت همبستگی بین دو متغیر را نشان می‌دهد. به‌عبارتی بیانگر رابطه‌ی خطی بین دو متغیر مثل x و y است. (ضریب همبستگی به روابط غیرخطی توجهی ندارد) ضریب همبستگی جامعه (P) را می‌توان به کمک ضریب همبستگی نمونه D برآورد کرد (برابر هم فرض کرد).

ضریب همبستگی

$$P=\frac{Cov(x,y)}{\delta x.\delta y} \quad و \quad Cov(x,y)=\sum_{i=1}^{K}\frac{(x_i-\bar{x})(y_i-\bar{y})}{N}$$

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}=\frac{n\sum xy-(\sum x)(\sum y)}{\sqrt{\left[n\sum x^2-(\sum x)^2\right]\left[n\sum y^2-(\sum y)^2\right]}}=\frac{\sum xy-n\bar{x}\bar{y}}{\left[\sum x^2-n\bar{x}^{-2}\right]\left[\sum y^2-n\bar{y}^{-2}\right]}$$

r ضریب همبستگی نمونه

- $r_{y,x}=r_{x,y}$
- $-1\le r\le+1$ می‌باشد. اگر x و y مستقل از هم باشند $r=0$ است. اما اگر $r=0$ شد فقط می‌گوییم بین x و y رابطه‌ی خطی وجود ندارد (در حالی که ممکن است رابطه‌ی غیرخطی وجود داشته باشد).

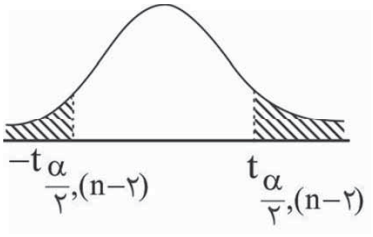
اگر $r=1$ باشد می‌گوییم همبستگی خطی بین x و y کامل و مثبت است یعنی y فقط

تابعی از x است و x و y هم جهت تغییر می‌کنند. اگر $r=-1$ باشد می‌گوییم همبستگی خطی بین x و y کامل و منفی است. یعنی y فقط تابعی از x است و x و y در خلاف جهت هم تغییر می‌کنند.

- وجود همبستگی به معنی رابطه‌ی علت و معلولی نیست.
- ضریب همبستگی مستقل از مبدأ و مقیاس اندازه‌گیری است.
- علامت ضریب همبستگی نشان‌دهنده‌ی جهت تغییرات است و شدت همبستگی با توجه به قدرمطلق ضریب همبستگی مشخص می‌شود البته نمی‌توان گفت $r=0/9$ نشان‌دهنده‌ی سه برابر قوی‌تر بودن همبستگی نسبت به $r=0/3$ است.
- آزمون فرض در مورد همبستگی: اگر ادعایی به‌صورت زیر مطرح شود که به آزمون معنی‌دار بودن ضریب همبستگی معروف است، به صورت زیر عمل می‌کنیم:

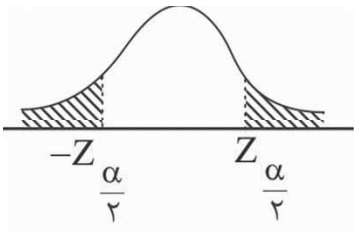
$$\begin{cases} H_0: P=0 \\ H_1: P\neq 0 \end{cases} \Rightarrow t=\frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \quad \text{و} \quad t_{\frac{\alpha}{2},(n-2)}$$

- رد H_0 به معنی معنی‌داری ضریب همبستگی و عدم رد آن به معنی عدم معنی‌داری ضریب همبستگی بین x و y است.



- اگر ادعایی به‌صورت زیر مطرح شد که به آزمون یکسان بودن همیشگی ضریب همبستگی با عددی ثابت است، از Z استفاده می‌کنیم:

$$\begin{cases} H_0: P=0 \\ H_1: P\neq 0 \end{cases} \qquad Z=\frac{r-P_0}{\sqrt{\frac{1-P_0^2}{\sqrt{n}}}} \quad , \quad Z_{\frac{\alpha}{2}}$$



- رگرسیون: در رگرسیون دو متغیره هدف ما به‌دست آوردن یک معادله به فرم $y=\alpha+\beta x$ است تا بررسی کنیم متغیر مستقل (x) چه تاثیری بر متغیر وابسته (y) دارد. اگر اطلاعات مربوط به جامعه را داشته باشیم و به ازای هر x یک y منحصر به فرد به دست آید، می‌توانیم به‌راحتی معادله‌ای به شکل بالا بنویسیم. اما معمولا به ازای هر x ما چند y خواهیم داشت در این صورت از μ_y به جای y_i استفاده می‌کنیم. اما با این کار بین هر y_i و μ_y اختلافی به

اندازه‌ی $y_i - \mu_y = u_i$ شکل می‌گیرد که به آن جمله‌ی اخلال یا خطا می‌گوییم. u_i دارای میانگین صفر و واریانس δ^2 است. با این تفاسیر خط رگرسیون مورد نظر به صورت زیر خواهد بود:

$$\mu_y = \alpha + \beta x \Rightarrow y = \alpha + \beta x + u_i$$

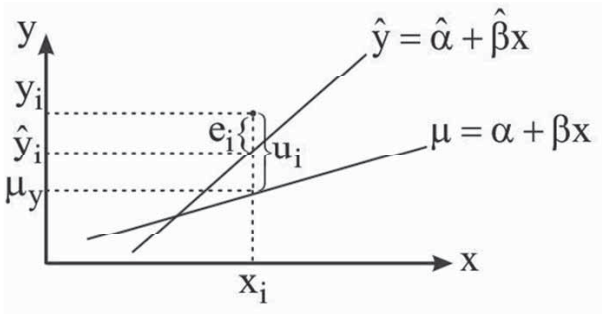
- خط رگرسیون به‌دست آمده خط رگرسیون جامعه است. اما از آن‌جا که معمولا ما اطلاعات مربوط به جامعه را به‌طور کامل نداریم، از یک نمونه برای برآورد خط رگرسیون کمک می‌گیریم. مطابق خط رگرسیون زیر که خط رگرسیون برآورد شده است، $\hat{y}$، $\hat{\alpha}$ و $\hat{\beta}$ به ترتیب برآوردهایی از β, α, μ_y هستند.

$$\hat{y} = \hat{\alpha} + \hat{\beta} x$$

همچنین می‌توان $y - \hat{y} = e_i$ که خطای نمونه نام دارد به‌عنوان برآوردی از جمله‌ی اخلال دانست. در این صورت خط رگرسیون برآوردی را می‌توان به‌صورت زیر هم نوشت.

$$y = \hat{\alpha} + \hat{\beta} x + e_i$$

- حال با توجه به مطالب گفته شده خط رگرسیون جامعه و خط رگرسیون برآوردی و u_i و e_i در شکل روبه‌رو نشان داده شده است:



- بنابراین برای به‌دست آوردن خط رگرسیون ما نیاز به محاسبه‌ی $\hat{\alpha}$ و $\hat{\beta}$ (برآوردگرها) داریم. برای این کار روش‌های مختلفی وجود دارد که بهترین و مهم‌ترین آن روش OLS است (روش حداقل مربعات معمولی)

- روش حداقل مربعات معمولی (OLS) یا کم‌ترین مجذورات: در این‌جا با مینیمم کردن $\sum e_i^2$ خواهیم داشت:

$$\beta = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{\sum xy - nxy}{\sum x^2 - n\bar{x}^2} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2} = \frac{SPD_{x,y}}{SSD_x} = \frac{SS_{x,y}}{SS_x}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \Rightarrow \hat{y} = \hat{\alpha} + \hat{\beta} x$$

- خط به‌دست آمده دارای ویژگی‌های زیر است:
  - مقادیر $\bar{x}$ و $\bar{y}$ در آن صدق می‌کنند.
  - $\bar{e}_i = 0 \leftarrow \frac{\sum e_i}{n} = 0$
  - $\bar{y} = \hat{\bar{y}}$ $4-$ e_iها با x_iها و y_i ها همبستگی ندارند یعنی: $\sum y_i e_i = 0 \, , \, \sum x_i e_i = 0$
- فروض مدل رگرسیون خطی (فروض نرمال):
  - میانگین هر یک از uiها برابر صفر است یعنی: $E(ui) = 0$
  - مقادیر uiها مستقل از همند و همبستگی ندارند یعنی: $v : i \neq j \Rightarrow Cov(ui, uj) = 0$
  - Uiها دارای واریانس یکسان هستند یعنی: $Var(u_i) = \delta_u^2$ این فرض بیان می‌کند که جامعه y متناظر با مقادیر مختلف X، واریانس یکسان دارند.
  - توزیع هر یک از uiها نرمال و مستقل از یکدیگر است. می‌توانیم بنویسیم: $ui \sim N(0, \delta_u^2)$
  - xiها غیرتصادفی‌اند و مقدار xi برای کلیه yi ثابت است.
  - Xiها و uiها همبستگی ندارند یعنی: $Cou(Xi, ui) = 0$ است.

- خطای معیار برآورد (انحراف معیار خطای نمونه) (Se): گاهی برای خط رگرسیون از استنباط آماری (فاصله اطمینان و آزمون فرض) استفاده می‌کنیم برای این کار لازم است انحراف معیار خطای جامعه (δ_u) را داشته باشیم، اما از آن‌جا که δ_u معمولا مشخص نیست ما آن را توسط Se که انحراف معیار خاطی نمونه یا خطای معیار برآورد نامیده می‌شود، برآورد می‌کنیم. Se پراکندگی نقاط نمونه را در بالا و پایین خط رگرسیون نمونه اندازه‌گیری می‌کند ($S_e = \hat{\delta}_u$)

$$Se = \sqrt{\frac{\sum y^2 - \hat{\alpha}\sum y - \hat{\beta}\sum xy}{n-2}} = \sqrt{\frac{\sum (y-\bar{y})^2 - \hat{\beta}^2 \sum (x-\bar{x})^2}{n-2}} = \sqrt{\frac{SSD_y - \hat{\beta}^2 SSD_x}{n-2}}$$

- خصوصیت تخمین زننده‌های OLS تحت فرض نرمال بودن:
  - بدون تورش هستند یعنی:

$$E(\hat{\alpha}) = \alpha \, , \, E(\hat{\beta}) = \beta \qquad \hat{\delta}_u^2 = S_e^2, \hat{\beta}, \hat{\alpha}$$

  - دارای حداقل واریانس هستند و داریم:

$$Var(\hat{\alpha}) = \delta_{\hat{\alpha}}^2 = \frac{S_e^2 \sum x^2}{n\sum x^2 - (\sum x)^2} / \delta_{\hat{\beta}}^2 = \frac{nS_e^2}{n\sum x^2 - (\sum x)^2}$$

  - سازگارند. یعنی اگر حجم نمونه به سمت بی‌نهایت میل کند. تخمین زننده‌ها به مقادیر جامعه‌شان نزدیک می‌شوند.

- $\hat{d}$ و $\hat{\beta}$ به‌طور نرمال توزیع شده‌اند بنابراین می‌توان آن‌ها به فرم استاندارد نرمال (z) درآورد.

$$\hat{\alpha} \sim N(\alpha, \delta^2_{\hat{\alpha}}) \Rightarrow Z = \frac{\hat{\alpha}-\alpha}{\delta_{\hat{\alpha}}} \sim (0,1) \quad \text{و} \quad \hat{\beta} \sim N(\beta, \delta^2_{\hat{\beta}}) \Rightarrow Z = \frac{\hat{\beta}-\beta}{\delta_{\hat{p}}} \sim (0,1)$$

- عبارت $\frac{(n-2)S_e^2}{\delta_u^2}$ از توزیع χ^2 با درجه‌ی آزادی n-2 تبعیت می‌کند: $\chi^2_{n-2} = \frac{(n-2)S_e^2}{\delta_u^2}$
- $\hat{\alpha}$ و $\hat{\beta}$ مستقل از S_e^2 ($\hat{\delta}_u^2$) توزیع شده‌اند.

استنباط‌های آماری در مورد پارامترهای خط رگرسیون جامعه:

- β: اگر آزمون فرضیه‌ای در مورد β مطرح گردد می‌توانیم از آماره آزمون زیر استفاده کنیم.

$$\begin{cases} H_0: \beta = \beta_0 \\ H_1: \beta \neq \beta_0 \end{cases} \qquad t = \frac{\hat{\beta}-\beta_0}{S_{\hat{\beta}}} \quad , \quad t_{\frac{\alpha}{2},(n-2)}$$

$$S_p = \sqrt{\frac{\sum(y-\bar{y})^2 - \hat{\beta}^2\sum(x-\bar{x})^2}{(n-2)\sum(x-\bar{x})^2}} \leftarrow S_{\hat{\beta}} = \sqrt{\frac{nSe^{12}}{n\sum x^2 - (\sum x)^2}} = \frac{Se}{\sqrt{\sum(x-\bar{x})^2}}$$

- یک روش برای آزمون وجود یا عدم وجود رگرسیون (عدم معناداری رگرسیون)، استفاده از آزمون فرضیه‌ی زیر است:
  - اگر H_0 رد شود، رگرسیون معنادار است در غیر این صورت رگرسیون وجود ندارد.

$$\begin{cases} H_0: \beta = 0 \\ H_1: \beta \neq 0 \end{cases} \Rightarrow t = \frac{\hat{\beta}}{S_{\hat{\beta}}} \quad , \quad t_{\frac{\alpha}{2},(n-2)}$$

- همچنین فاصله‌ی اطمینان C درصدی برای β نیز به صورت زیر خواهد بود.

$$P(\hat{\beta} - t_{\frac{\alpha}{2},(n-2)}.S_{\hat{\beta}} \leq \beta \leq \hat{\beta} + t_{\frac{\alpha}{2},(n-2)}.S_{\hat{\beta}}) = C$$

- α:

$$\begin{cases} H_0: \alpha = \alpha_0 \\ H_1: \alpha \neq \alpha_0 \end{cases} \qquad t = \frac{\hat{\alpha}-\alpha_0}{S_{\hat{\alpha}}} \quad , \quad t_{\frac{\alpha}{2},(n-2)}$$

$$S_{\hat{\alpha}} = \sqrt{\frac{S_e^2\sum x^2}{n\sum x^2 - (\sum x)^2}} = \frac{Se}{\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x-\bar{x})^2}}}$$

$$P(\hat{\alpha} - t_{\frac{\alpha}{2},(n-2)}.S_{\hat{\alpha}} \leq \alpha \leq \hat{\alpha} + t_{\frac{\alpha}{2},(n-2)}.S_{\hat{\alpha}}) = C$$

δ_α^2 ⇐ :

$$\begin{cases} H_0 : \delta_u^2 = \delta_0^2 \\ H_1 : \delta_u^2 \neq \delta_0^2 \end{cases} \qquad \chi^2 = \frac{(n-1)Se^2}{\delta_0^2} \ , \ \chi^2_{\frac{\alpha}{2},(n-2)}$$

$$P\left(\frac{(n-2)S_e^2}{\chi^2_{\frac{\alpha}{2},(n-2)}} \leq \delta_u^2 \leq \frac{(n-2)S_e^2}{\chi^2_{(1-\frac{\alpha}{2}),(n-2)}} \right) = C$$

- μ_y: برآورد فاصله‌ی اطمینان C درصدی برای μ_y به ازای مقدار ثابت $x = x_0$ به صورت زیر انجام می‌شود.

$$P(\hat{y} - t_{\frac{\alpha}{2},(n-2)} . S_{(\mu_y|x_0)} \leq \mu_y \leq \hat{y} + t_{\frac{\alpha}{2},(n-2)} . S_{(\mu_y|x_0)}) = C$$

$$\hat{y} = \hat{\alpha} + \hat{\beta} x_0 \ , \ S_{(\mu_y|x_0)} = Se\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - n\bar{x}^2}}$$

- Y: برآورد فاصله‌ی اطمینان C درصدی برای y به‌ازای مقدار ثابت $x = x_0$ نیز به‌صورت زیر انجام می‌گیرد:

$$P(\hat{y} - t_{\frac{\alpha}{2},(n-2)} . S_{(y|x_0)} \leq y \leq \hat{y} + t_{\frac{\alpha}{2},(n-2)} . S_{(y|x_0)}) = C$$

$$S_{(y|x_0)} = Se\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - n\bar{x}^2}}$$

- یک روش دیگر برای آزمون عدم معناداری رگرسیون (با همان وجود رابطه‌ی خطی بین x و y یا عدم وجود این رابطه) استفاده از جدول آنالیز واریانس است. به‌طوری که در آن:

$$SST = SSR + SSE$$

تغییرات کل: $SST = \sum (y - \bar{y})^2 = \sum y^2 - \frac{1}{n}(\sum y)^2$

تغییرات توضیح داده شده: $SSR = \sum (\hat{y} - \bar{y})^2 = \hat{\beta}^2 \sum (x - \bar{x})^2 = \hat{\alpha} \sum y + \hat{\beta} \sum xy - \frac{1}{n}(\sum y)^2$

تغییرات توضیح داده نشده: $SSE = \sum (y - \hat{y})^2 = \sum y^2 - \hat{\alpha} \sum y + \hat{\beta} \sum xy$

و مقدار به‌دست آمده برای F را به صورت راست دامنه با $F_{\alpha,1,n-2}$ مقایسه می‌کنیم.

منبع تغییرات	درجه آزادی	مجموع مجذورات	میانگین مربعات	آماره آزمون F
X	1	SSR	$MSR = SSR$	$F = \frac{MSR}{MSE}$
خطا	n-2	SSE	$MSE = \frac{SSE}{n-2}$	
کل	n-1	SST		

- ضریب تعیین: مجذور ضریب همبستگی است و به ما می‌گوید که چند درصد از تغییرات y توسط x توضیح داده می‌شود.

$$R^2 = r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = \frac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2} = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}$$

$$R^2 = \frac{(n\sum xy - \sum x \sum y)^2}{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]} = \frac{\hat{\alpha}\sum y + \hat{\beta}\sum xy - \frac{1}{n}(\sum y)^2}{\sum y^2 - \frac{1}{n}(\sum y)^2}$$

- ضریب تعیین مهم‌ترین معیار خوبی برازش خط رگرسیون است ضمناً برای مقایسه‌ی همبستگی باید از ضریب تعیین استفاده کنیم. یعنی اگر $r_1 = 0/8$ و $r_1 = 0/4$ باشد می‌گوییم همبستگی بین x و y وقتی $r = 0/8$ است. $\left(\frac{r_1^2}{r_2^2} = \frac{(0/8)^2}{(0/4)^2} = 4\right)$ 4 برابر وقتی است که $r = 0/4$ است. همچنین همیشه برای تغییر همبستگی از ضریب تعیین استفاده می‌کنیم.
- همواره $r = \pm\sqrt{R^2}$ است و علامت r با توجه به علامت $\hat{\beta}$ تعیین می‌شود.

$$r = \hat{\beta}\frac{S_x}{S_y} \quad , \quad \hat{\beta} = r\frac{S_y}{S_x} \quad / \quad r = \frac{Cov(x,y)}{\delta x.\delta y} \quad , \quad \hat{\beta} = \frac{Cov(x,y)}{\delta_x^2}$$

- اگر معادلات رگرسیون خطی y برحسب x و x برحسب y به صورت زیر باشد داریم: $R^2 = \hat{\beta}_1.\hat{\beta}_2$

$$\hat{y} = \hat{\alpha}_1 + \hat{\beta}_1 x \quad , \quad \hat{x} = \hat{\alpha}_2 + \hat{\beta}_2 y$$

- ضرب کردن متغیرها در اعداد مثبت تاثیری روی ضریب همبستگی ندارد.
- معادله خط رگرسیون را می‌توان به صورت زیر نیز نوشت: $\hat{y} = \bar{y} + \hat{\beta}(x - \bar{x})$
  - گاهی پیش می‌آید که در برخی سوالات α را بنا به فرمول فوق برابر $\bar{y}$ فرض کنند.
- می‌دانیم (برای x هم همین‌طور است):

$$\sum(y-y)^2 = n\sum y^2 - (\sum y)^2 = \sum y^2 - \frac{1}{n}(\sum y)^2 = \sum y^2 = \sum y^2 - n\bar{y}$$

جامــع

حاوی تمامی مطالب و نکات لازم
برای کنکور براساس منابع

بانک سوالات

جمع‌آوری سوالات کنکور کاردانی به کارشناسی
کارشناسی ارشد و دکتری به‌صورت فصل‌بندی شده

الگــوریتم

چکیده‌ی تمامی مطالب و نکات لازم
برای کنکور براساس منابع

سوالات تألیفی

تألیف سوالات مشابه کنکور



دریافت نمونه‌ی کتاب به‌صورت رایگان

FREE
DOWNLOAD

www.agkins.com

جامـــع

حاوی تمامی مطالب و نکات لازم
برای کنکور براساس منابع

بانک سوالات

جمع‌آوری سوالات کنکور کاردانی به کارشناسی
کارشناسی ارشد و دکتری به‌صورت فصل‌بندی شده

الگـــوریتم

چکیده‌ی تمامی مطالب و نکات لازم
برای کنکور براساس منابع

سوالات تألیفی

تألیف سوالات مشابه کنکور



دریافت نمونه‌ی کتاب به‌صورت رایگان

FREE
DOWNLOAD

www.agkins.com

شماره تماس با نمایندگی‌های فعال و رسمی

آناهیـد گستـر خـلیـلی

نمایندگی	شماره تماس	نمایندگی	شماره تماس
اراک (آقای حسینی)	09913713750	شیراز (آقای فروردین/ خانم هوشمندی)	09195730152
اردبیل (خانم عاصمی‌زاده)	09199101248	فسا (خانم شعبانی)	09913713742
ارومیه (آقای آرمیون)	09199101241	قزوین (خانم چگینی)	09195730149
اسفراین (خانم اسماعیل‌زاده)	09196321852	کرج (آقای دکتر علیرضاپور)	09197781947
اصفهان (آقای نصری)	09195730150	کاشان (خانم صادق‌پور)	09913713746
ایلام (آقای دکتر بوچانی)	09190980318	کردستان (آقای ابراهیمی)	09195730148
اهواز (آقای امین‌فر)	09195730154	کرمانشاه (آقای ابراهیمی)	09195730148
بیرجند (آقای دکتر بهروان)	09195907203	کهکیلویه (آقای فروردین/ خانم هوشمندی)	09195730152
تبریز (خانم عاصمی‌زاده)	09195730147	گرگان (آقای شاه‌علی)	09199101247
جهرم (آقای یاعلی جهرمی)	09013737896	مراغه (آقای صمدی)	09195733179
خرم‌آباد (خانم میر)	09192705878	مشهد (خانم دکتر شجاعی)	09913713744
زنجان (خانم دکتر هوشیار)	09192705871	میاندوآب (آقای صمدی)	09195733179
ساری (آقای دکتر اکبری)	09197781944	هرمزگان (آقای زارعی)	09195733284
سمنان (آقای دکتر سلمانی)	09913713741	همدان (آقای سوری)	09195730155
شهرکرد (خانم تقی‌پور)	09199101249	یزد (خانم مهندس آزاد)	09199101243

بانک کتاب ناهید

«هر کتابی، از هر انتشاراتی را از ما بخواهید»



- ✓ جامع‌ترین بانک کتاب
- ✓ تحویل روزانه
- ✓ ارسال به تمامی نقاط کشور
- ✓ سفارش کتاب به‌صورت تلفنی و آنلاین

www.NIBS.ir

کتب دانشگاهی، فنی‌و مهندسی، علوم پزشکی، علوم انسانی، عمومی، ادبی، مذهبی، کمک آموزشی، کودک و نوجوان و کتب نفیس

فروشگاه: تهران – خیابان انقلاب- روبه‌روی درب اصلی دانشگاه تهران

پاساژ فروزنده – طبقه همکف – پلاک ۳۳۱

تلفن: ٦٦٤٨٩٣٧٥ – ۰۲۱ – ٦٦٤٨٩٣٤٩ – ۰۲۱