

درس طایفه جدید آمار پیشرفته ۳ واحدی هستند. اگر خواستیم درس را برای بالابردن معدل، بسته آزمونهای چهار
نیمه سیستم فصول با برطرف شدن این درس را بر داریم. برای ادامه تحصیل مان یا برای این نام داشته باشیم یا یک مقاله به یک جایی مقیم
نویسیم والا امکان نیست برای ادامه تحصیل. مقاله از این نام ارزشش نیست است چون علاوه بر این، ارزش حقوق را به عنوان
یا به عنوان یک نیمه ای هم نوشته و این بهره.

به نام آنکه به شماری موجودات آگاه است

دوره مجازی کارشناسی ارشد آمار ریاضی با گرایش های آمار رسمی و داده کاوی

گروه آمار، دانشکده ریاضی و علوم کامپیوتر، مرکز آموزش الکترونیکی دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

این دوره برای تربیت کارشناسان ارشد آمار مورد نیاز ادارات و سازمانها و تربیت آنها برای ادامه تحصیل در سطوح دکتری جهت انتقال علم آمار، آموزش آمار و به کارگیری آنها در سطوح بالاتر می باشد. این رشته در صدد است، کارشناسان ارشدی تربیت کند که در مراکز تولید آمار، نقش موثری داشته و در جهت بهبود کیفیت داده های آماری و تحلیل آنها از دانش خود بهره گیرند.

با توجه به کاربرد این رشته در گردآوری، پردازش، استنباط، مدل سازی و تجزیه و تحلیل انواع سیستم های تولید اطلاعات و همچنین کاربرد در مسایل صنعتی، کشاورزی، پزشکی، اجتماعی اهمیت این رشته به خوبی مشخص می شود. مراکزی نظیر سازمان مدیریت و برنامه ریزی کشور (استانداری ها)، مرکز آمار ایران، بانک ها، شرکت های صنعتی و بازرگانی و سایر مراکز دولتی و خصوصی از عمده مراکز جذب دانش آموختگان این دوره می باشند.

شایان ذکر است که این گرایش ها، هیچ تفاوتی با رشته های آمار ریاضی مصوب وزارت علوم، که هم اکنون در گروه آمار دانشکده ریاضی و علوم کامپیوتر دانشگاه صنعتی امیرکبیر برقرار می باشد، ندارد. صرفاً دروس اختیاری، سمینار و پروژه تحقیق (بجای پایان نامه) به صورت جهت دار در زمینه های آمار رسمی یا داده کاوی، یا استنباط آماری، احتمال و فرایندهای تصادفی (یعنی گرایش آمار ریاضی)، اخذ می شوند.

با این وجود دروس اجباری تا حد امکان به صورت کاربردی ارایه می شود و نتایج ریاضی بدست آمده با کمک نرم افزارهای آماری شبیه سازی خواهد شد. تا ضمن درک بهتر از مسایل طرح شده به نرم افزارهای آماری، که جزء لاینفک آمار می باشند، تسلط لازم پیدا کنند.

جزئیات بیشتر، نظیر شرایط عمومی و اختصاصی را در سایت آموزش الکترونیکی دانشگاه ملاحظه کنید: <http://vu.aut.ac.ir>

آمار رسمی Official Statistics : داده های متوسط دولتی تولید می شود

آمار واژه ای است که معنای دوگانه ولی مرتبط به هم دارد. در معنای اول، آمار به اطلاعات عددی یعنی اعداد و ارقامی که به منظور خاصی تهیه و به شکل های مختلفی (نمودار، جدول و...) نمایش داده می شوند، اطلاق می شود. در معنای دوم، واژه ای آمار عنوانی برای یک شاخه ی علمی است؛ مجموعه فنون گردآوری داده ها، طبقه بندی و تلخیص آنها و سرانجام نتیجه گیری از داده ها. ارتباط این دو در این واقعیت نهفته است که اگر اعداد و ارقام پیش گفته با روش های علمی به دست آمده باشند باید با کمک علم آمار باشد. به عبارت دیگر، استفاده از

درس این سلب در مورد حقایق دارد است یا داده کاوی. ما میخواهیم یکی کنیم تا در این تحلیل داده ها و داده کاوی را در این سلب بیان کنیم. این سلب که در اینجا هست برای سلب سلب است که میخواهند بر سر آن داده کاوی ها در دانشگاه را بر سر داده کاوی، البته دوتا گزاش بود یک گزاش که آمار رسمی بود و یک گزاش هم هست بان داده کاوی که بعضی بگویند است که ما داریم. همه چیزهایی که لازم است در قسمت طبع درس است و در اردن (اسم درس) - سرفصل درس - چه سلب ای هم چیزی میخوانیم درس بر همین؟! (همین سلب در قسمت طبع در این سلب)

چیزی که ما اینجا می بینیم مخصوصاً آماری ها عبارت از Data Analiziz و Data mining را بخوبی متوجه نمی شویم. میخواهم ببینیم تحلیل داده ها و حفاری داده ها که بخاطر بخش صحن داده ها می چرخد. مقدس داده های چهلوی پس آرد!! همه آمار کارش با داده ها نبود!! چرا عجب داده های را مطرح کردند!! با چند تا مثال ساده که در کتاب Tan در فصل اولی آمده گفته، سعی کنید که این کتاب را به دست بگیرید. چند تا مثال را هم ما در اینجا به اختصار می نویسیم.

علم آمار شرط لازم برای تهیه آماری قابل استناد است. واژه ای آمار در عبارت آماری رسمی به معنای اول آن است ولی واژه ای رسمی، آن را به اطلاعات عددی خاصی محدود می کند. آماری رسمی به اطلاعات عددی گفته می شوند که توسط دولت یا مراجع صلاحیت دار که در قوانین و مقررات مشخص هستند، تولید و منتشر می شوند و اطلاعاتی را در مورد وضعیت عمومی کشور برای امور مدیریتی (برنامه ریزی، سیاست گذاری، و تصمیم گیری) به دست می دهند.

در تعریف واژه ای آمار به دو معنای آن و ارتباط آن ها با یکدیگر اشاره شد. این ارتباط در نقشی نهفته بود که علم آمار در تولید آماری ایفا می کرد زیرا بخشی از مطالعات کاربردی این علم به حل مسائل واقعی و اجرایی در گردآوری داده ها و تولید آماری اختصاص دارد. معمولاً هدف این مطالعات ابداع روش هایی است که آماری با کیفیت بالا و با کمترین هزینه را به دست دهند. بنابراین باید در نظام آماری، تولید آماری به متخصصان آماریشناسان واگذار شود تا استفاده از روش های علمی در تولید آمار نمود پیدا کند. در این خصوص نیاز به تاسیس گرایشی با نام آمار رسمی احساس می شود. تولید آماری رسمی نتیجه فعالیت های مختلفی است: وضع قوانین و مقررات، تهیه تعاریف و استانداردها، تعیین نیازها و اولویت ها، به کار گیری روش های علمی، طراحی ابعاد اجرایی، استخراج و تحلیل داده ها و سرانجام انتشار و اطلاع رسانی اطلاعات آماری. مطالعه این فرایند آن چیزی است که در رشته آمار رسمی مورد بررسی قرار می گیرد.

معرفی برخی از درس های گرایش آمار رسمی

گرایش آمار رسمی، تلفیقی از چندین مبحث آمار، اقتصاد، جمعیت شناسی، مدیریت و ... می باشد. در ابتدای ورود به رشته ای آمار رسمی، قبل از هر چیز باید شناخت مقدماتی از محیطی که در آن آماری رسمی تولید می شوند، پیدا کرد. در درس **مقدمه ای بر آماری رسمی** جایگاه اطلاعات آماری در برنامه ریزی ها و در پی آن به نقش نظام های آماری برای تولید و عرضه اطلاعات آماری و به تشریح وظایف واحدهای مرکز آمار ایران و نقش هر یک در تولید آماری پرداخته می شود.

آمارهای ثبتي که بخش عمده ای از آماری رسمی را در دل خود دارند و نیز کاربردهای آماری رسمی در تحلیل های اجتماعی - اقتصادی مانند حساب های ملی معرفی می شوند. اما تولید آماری رسمی به تنهایی کافی نیست و بررسی کیفیت این داده ها نقش مهم تری را ایفا می کند. به همین دلیل، در رشته آمار رسمی، شناخت روش های مختلف آمارگیری، تفاوت آن ها با یکدیگر، ابزارهای گردآوری داده ها و مسائلی پیرامون آن، بخش مهمی از مطالعات را به خود اختصاص می دهد که در درس **روش های آمارگیری** به بحث گذاشته می شوند.

یکی دیگر از مباحثی که کیفیت آماری رسمی را تحت تاثیر خود قرار می دهد، طرح های نمونه گیری می باشد. به همین جهت درس نمونه گیری کاربردی برای شناسایی انواع طرح های نمونه گیری، برآوردها، نحوه برخورد با بی پاسخی و ... در نظر گرفته شده است.

مثال «فرض کنید نزدیک مسجد مسعود، در سمت راست که تمام ترانس های ایرانی ایران، از آن سمت رد میشوند، ما بایستیم تعداد ترانس ها را بشماریم و این کار را در فرام انجام داده اند، و به این ترتیب ترانس ها مثلاً در یک روز مثلاً ۱۰۰ ترانس است، در یک روز مثلاً یک ترانس این بود که یک ترانس یک آدمی فریاد و ... پس معلوم می شود که این ترانس ها یک مقدار خیلی کم یا یک مقدار خیلی زیاد را نشان می دهند و پس ممکن است بماند، حالا فرض کنید ما ۱۰۰ میلیون عدد داریم میخوانیم از آنها می بینیم، با چه نرم افزاری میتوانیم این کار را انجام دهیم؟ ۱۱۲ قوی ترین نرم افزار به ما داده است، Matlab است، اما با Matlab نمیتوانیم کار کنیم، اما با Excel میتوانیم.

که ممکن است یک اطلاعات دیگری درون این داره ما نفهمه باشد که اصلاً ما نتوانیم در مورد پادشاه یا خصم این که در نظرمان
است و به اصطلاح بیاییم یک آنوقت آمار این انجام دهیم. به جای این که آنوقت آمار انجام ^{آن} دهیم، بیاییم بر یکم داخل دل داره ما
بیاییم چه اطلاعاتی میتوانیم از آن دل داره ما بدست آوریم یا چهاره نشان کنیم و به آن بیاییم که مضواصم بریم.
پس تولید داره ما سرعت خیلی زیاد می‌کنیم، از آن طرف سفت افراد و نرم افراد ما که ما داریم خیلی پیشرفت کرده اند.

گیرنده باشند که این به صورت خودکار و بر اساس تجربه‌ای که از طریق شبکه‌های عصبی یا درخت‌های تصمیم‌گیری به دست می‌آورند، بهبود می‌بخشد. داده کاوی منحصر به گردآوری و مدیریت داده‌ها نبوده و تجزیه و تحلیل اطلاعات و پیش‌بینی را نیز شامل می‌شود برنامه‌های کاربردی که با بررسی فایل‌های متن یا چند رسانه‌ای به کاوش داده‌ها پرداخته، و پارامترهای گوناگونی را در نظر می‌گیرد، که اساس الگو یابی و مدل‌سازی در داده کاوی هستند:

← forecasting (برای آینده)
← prediction (برای حالت)
← interpolation (برای کمبود)

پیوند (Association): الگوهایی که بر اساس آن یک رویداد به دیگری پیوند داده می‌شود مثلاً خرید قلم به خرید کاغذ.

ترتیب (Sequence): الگویی که به تجزیه و تحلیل توالی رویدادها پرداخته و مشخص می‌کند کدام رویداد رویدادهای دیگری را در پی دارد.

رده بندی: (Classification) شناسایی الگوهای جدید.

خوشه بندی (Clustering): کشف و مستند سازی مجموعه‌ای از حقایق ناشناخته.

پیش بینی (Forecasting): کشف الگوهایی که بر اساس آنها پیش بینی قابل قبولی از رویدادهای آتی آرایه می شود.

اختلاف‌ها و تشابه‌های داده‌کاوی با آمار

داده‌کاوی (کشف دانش در پایگاه داده‌ها)، هم پوششی زیادی با آمار کلاسیک دارد، به طوری که ممکن است برخی از مردم داده‌کاوی را زیرمجموعه‌ای از آمار تلقی کنند. برخی از آمارشناسان محافظه‌کار نیز چنین عقیده‌ای دارند ولی واقع‌گرایانه نیست. داده‌کاوی از ابزار و روش‌هایی از سطوح دیگر مخصوصاً یادگیری ماشین، نظریه و فناوری پایگاه داده‌ها و... استفاده می‌کند و ضرورتی ندارد که به برخی از سطوحی که مد نظر آمار است مربوط شود. همچون فرض‌هایی که در آمار وجود دارند خیلی زیاد است.

اخيراً علاقه‌مندی برخی از آمارشناسان به رشته‌های گوناگونی از دانش، از جمله داده‌کاوی، یادگیری ماشین و ... به صورت شگفت‌انگیزی افزایش یافته است. داده‌کاوی به خلاصه‌های مؤثر از مقادیر بزرگ داده‌ها، شناخت ساختار مورد علاقه، رابطه‌های درون مجموعه داده‌ها و استفاده از مجموعه داده‌های قبلی برای پیشگویی مشاهده‌ها آینده می‌پردازد. آمارشناسان نیز فنون خوبی برای پرداختن به همه این مسایل را دارند. مدل‌های آماری زیادی برای روشن کردن رابطه‌های یک مجموعه از داده‌ها یا برای پیشگویی‌ها وجود دارد. تحلیل خوشه‌ای، تحلیل تشخیصی، رگرسیون نا پaramتری و ... در بسیاری از مسایل داده‌کاوی به کار می‌روند. برای یک آمارشناس ملاحظه داده‌کاوی تنها به عنوان شاخه‌ای از آمار گمراه کننده است. با وجود این مسایل و روش‌های داده‌کاوی ویژگی‌های متفاوتی دارند.

در تحلیل داده ها به شکل مسئله فرای داریم یا نرم افزار؟!! ما مشکل نرم افزاری داریم، ما در سخت افزار حل می توانیم به نسبت نرم افزار - بران این ما بتوانیم کارهای داده های را با کامپیوتر انجام دهیم، معمولاً کنار هر نرم افزار یک نرم افزار Data mining هم ارائه میشود مثلاً نرم افزار SPSS که به PS با تغییر بسته آورده و آموخته یک نرم افزار بنام clementine ارائه داده که با ماشین کارهای داده ای جدید اضافه شده است

مثلاً انجام داد. قوی ترین نرم افزار که فعلاً وجود دارد نرم افزار SAS است که این هم یک نرم افزار ارائه داده SAS Enterprise و البته که این نرم افزارها در بعضی رده (این) است که این نرم افزار که در سطح بسیار بالا قرار دارد و نیاز به آموزش خاص دارد و تخصص خاص دارد.

به نام دانشگاه های دیگر از نرم افزار Veka استفاده میکنند. مرکز آمار که صندوق داده های ما را به شما می دهد. SAS استفاده میکند. نرم افزار ریمپل بر روی R یا استاتیک matlab و silab که matlab پولی است اما طاب صابنی است. معمولاً هر نرم افزار که پولی است یک نرم افزار مجانی هم کنارش میزنند که بتواند به نوعی آن را برسد. R یک قس داره برای داده ها و R محدودیت حجم داده ندارد مثل SAS، SPSS دارد محدودیت مثل clementin.

داده کاوی و آمار هر دو به طرح استنباط از داده ها مربوط می شوند. هدف از استنباط ممکن است درک الگوهای همبستگی و پیوندهای سببی بین مقادیر داده ها (توصیف)، یا ساخت پیشگویی ها از مقادیر داده های قبلی است. آمار کلاسیک دربرگیرنده مدلی برای توزیع احتمال داده ها و ساخت استنباط به صورت بیان احتمال است. روش های داده کاوی در بسیاری از موارد مسایلی را در برمی گیرد که به سادگی در چارچوب آمار کلاسیک برآزش داده نمی شوند و از آمار فاصله گرفته اند. در آمار ما با جامعه سر و کار داریم در حالی که داده کاوی با پایگاه داده ها سرو کار دارد. داده کاوی با نمونه گیری مانند آمار سرو کار ندارد، یعنی دوباره می توان از داده ها استفاده کرد. داده کاوی از میان مجموعه مدل های برآزش داده شده، بهترین مدل را انتخاب می کند، ولی در آمار تنها یک مدل برآزش می شود. آمار مدل برآزش داده شده را برای تعمیم به جامعه به کار می برد، ولی داده کاوی مدل برآزش داده شده انتخابی را برای پیشگویی داده های جدید به کار می برد.

مجموعه داده ها در داده کاوی می تواند خیلی وسیع تر و گسترده تر از مجموعه داده های معمول در آمار، شامل چند صد میلیون یا هزار میلیارد رکورد باشد.

تفاوت آمار و داده کاوی؟

نوع داده ها در آمار کمی است ولی داده کاوی داده های کمی، کیفی و متنی را پوشش می دهد.

نوع متغیرهای ورودی در آمار عددی ولی در داده کاوی متغیرها از نوع عددی، طبقه ای و متنی می باشند.

نوع تمرکز در آمار روی مدل ولی در داده کاوی تمرکز روی الگو است.

زیر بخش های اصلی آمار برآورد، توزیع های احتمال، آزمون فرضیه، امتیازبندی مدل و پیشگویی است، ولی زیر بخش های داده کاوی مدل بندی پیشگویی ها، بخش بندی پایگاه داده ها و ساخت فرضیه است.

هدف اصلی از به کار گیری روش ها در آمار استفاده از برآوردها و توزیع ها برای ادغام اطلاعات، ولی در داده کاوی هدف اصلی کشف دانش مورد علاقه است.

شکل نتایج در آمار به صورت مدل های کلی برآورد می شود ولی در داده کاوی مدل های موضعی محاسبه می شوند.

ارزش اطلاعاتی نتایج در آمار معلوم و محدود است در حالی که در داده کاوی نامعلوم و نامحدود می باشد.

در آمار جستجو برای دستیابی به نتایج محدود به جستجوهای جهت دار بوده و با نتایج نیز آشنا هستیم. ولی در داده کاوی جستجوها اکثراً از طریق کاربر تعیین شده و ممکن است جهت دار باشند، اما اصولاً روش خودکار است و نوع نتایج نامعلوم است.

خرید لباس های ما چه کاری انجام می دهیم؟! به عنوان مثال به ما اینطور در آمار مردم در استان ما این شهرهای مختلف چه است؟! عدم شهرها را با هم میزنیم و در آن شهرها را هم با هم میزنیم، این اعداد خیلی قابل استفاده نیست، مقداری با این یک نقشه GIS این استان را میزنند و بعد شهرهای که در آن هستند را هم میزنند، این یک خوش قرار دهند، یک رنگ بزنند، خیلی واضح تر میشود. خوب این در آن افراد خاص بود در آن ما میفهمیم استان ها را که در آن است، مثلاً آن شخص را با آن ملل متعدد با هم مقایسه کنیم، وضعیت بهداشت، وضعیت رانندگی، در آن خانوار، بعد خانوار، سواد خانوار و... آن در مقایسه میکنیم، مثلاً آن را در جدول نگاه کنیم، الان در آن غیر ممکن است که از میان یک هست جدول عددی ها را بگیریم و در آن در آن clustering این کار را با احتیاط میکنیم انجام دهیم. هم تعداد خوشه ها را مشخص می کند هم یک

البته classification و clustering فقط تفاوت اسی در این است که تعداد مشخص نیستند classification اصطلاحاً unsupervised است و supervised clustering است یعنی راهنمایی شده است \Rightarrow classification
 راهنمایی شده است \Rightarrow clustering
 برای اینکه سیر خلاصه‌تری به آمار تفاوت آمار داده‌کاران را می‌بینیم؟!! که فکر نکنند داده‌کاری یک زیر شاخه‌ای از آمار است.

در مقایسه با آمار، داده‌کاوی توجه کمتری به ویژگی‌های جانبی استنباط‌های بزرگ نمونه‌ای دارد و فلسفه کلی یادگیری، بیشتر شامل ملاحظه پیچیدگی مدل‌ها و محاسباتی که آن‌ها نیاز دارند، است.

برای اطلاعات بیشتر به مرجع‌های زیر مراجعه کنید:

حائری مهریزی، علی اصغر (۱۳۸۲) «داده‌کاوی: مفاهیم، روش‌ها و کاربردها» پایان‌نامه کارشناسی ارشد آمار اقتصادی و اجتماعی، دانشکده اقتصاد، دانشگاه علامه طباطبائی.

Hand. D.J. (1998) Review of Data mining, The American Statistician, 52, 112-118

<http://fa.wikipedia.org/wiki>

<http://www.creem.st-and.ac.uk/datamining>

تعداد واحدهای درسی این دوره، ۳۲ واحد به صورت زیر است:

- دروس گروه ۱ (اصلی-اجباری): ۱۶

- دروس گروه ۲ (تخصصی-اختیاری): حداکثر ۱۶ واحد

- روش تحقیق و سمینار ۲ واحد

دروس گروه ۱ (اصلی-اجباری)

ردیف	عنوان	تعداد واحد
۱	استنباط آماری ۱	۴
۲	آنالیز ریاضی ۲	۴
۳	استنباط آماری ۲	۴
۴	احتمال ۱ (نظریه احتمال ۱)	۴

عنوان بندی تکمیلی است برای خیر مقدم پس به خوش بینی یک تکی از داده‌کاران است.
 کتاب TAN و مدل‌های آنش است.

دروس گروه ۲ (تخصصی-اختیاری)

ردیف	عنوان	تعداد واحد	گرایش
۱	داده کاوی	۳	داده کاوی، آمار ریاضی، آمار رسمی
۲	پروژه تحقیق	۳	آمار ریاضی، آمار رسمی، داده کاوی
۳	مقدمه‌ای بر آمار رسمی	۳	آمار رسمی
۴	خوشه بندی و تصویرسازی	۳	داده کاوی، آمار رسمی، آمار ریاضی
۵	فنون آماری	۴	داده کاوی، آمار ریاضی
۶	مدل‌های خطی ۱	۴	آمار ریاضی، آمار رسمی، داده کاوی
۷	فرآیندهای تصادفی ۲	۴	آمار ریاضی
۸	سری‌های زمانی ۲	۴	آمار ریاضی
۹	آنالیز چند متغیره	۴	آمار ریاضی، داده کاوی
۱۰	نظریه صف	۴	آمار ریاضی
۱۱	روش‌های آمارگیری	۳	آمار رسمی
۱۲	نمونه گیری کاربردی ۱	۳	آمار رسمی
۱۳	خطای کل آمار گیری	۳	آمار رسمی
۱۴	مباحث مخصوص در آمار	۳ یا ۴	داده کاوی، آمار ریاضی، آمار رسمی
۱۵	محاسبات آماری و تکنولوژی داده‌گان	۳	داده کاوی، آمار ریاضی، آمار رسمی
۱۶	نمونه گیری کاربردی ۲	۳	آمار رسمی
۱۷	آمارهای ثبتي	۳	آمار رسمی
۱۸	تحلیل داده‌های نمونه‌گیری‌های پیچیده	۳	آمار رسمی
۱۹	مدیریت آمار گیری	۳	آمار رسمی
۲۰	حساب‌های ملی	۳	آمار رسمی
۲۱	روش تحقیق آماری	۲	داده کاوی، آمار ریاضی، آمار رسمی

بسم تعالی

تاریخ:
شماره:
پیرست:

دانشگاه صنعتی امیر کبیر
(بلی تکنیک تهران)

دانشکده ریاضی
علوم کامپیوتر



در تمرین‌هایی را باید حل کنید:
از هر فصل، قسم اول از سمت راست
شماره دانشجو می‌خورد + ۱۰ آده را حل کنید
به عبارت دیگر از هر فصل دو تمرین باید حل
آگر این اعداد از تعداد تمرین‌ها بزرگتر بود، از
اینکه از فصل به عنوان ادامس آن در نظر بگیریم
و حل کنید.

۳. شماره شماره تمرین‌هایی باشد. در صورتی
که ۱. تمرین کامل می‌شود و حل شده باشد
شماره آن را خواص بگیرید و با توجه به
شماره سوالات امتحان به تمرین‌ها
کتاب آگهی توصیف می‌دهد و تمرین‌ها را حل کنید
در صورتی که شماره کامل اخذ نشود شماره گرفته شده

تهران - خیابان حافظ - روبروی خیابان سیمه - تلفن: ۶۶۴۰۶۳۲۲ و ۶۶۴۰۶۳۲۰ - فاکس: ۶۶۴۹۷۹۲ - پست الکترونیک: math-os@aut.ac.ir
ب. شماره آزمون پایان نامه خواهد بود که در آن