

به نام آنکه به شماره‌ی موجودات آگاه است

## دوره مجازی کارشناسی ارشد آمار ریاضی با گرایش‌های آمار رسمی و داده کاوی

گروه آمار، دانشکده‌ی 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 به‌کارگیری آن‌ها در سطوح بالاتر می‌باشد. این رشته در صدد است، کارشناسان ارشدی تربیت کند که در مراکز تولید آمار، نقش موثری داشته و در جهت بهبود کیفیت داده‌های آماری و تحلیل آن‌ها از دانش خود بهره گیرند.

با توجه به کاربرد این رشته در گردآوری، پردازش، استنباط، مدل‌سازی و تجزیه و تحلیل انواع سیستم‌های تولید اطلاعات و همچنین کاربرد در مسایل صنعتی، کشاورزی، پزشکی، اجتماعی اهمیت این رشته به خوبی مشخص می‌شود. مراکز نظیر سازمان مدیریت و برنامه‌ریزی کشور (استانداری‌ها)، مرکز آمار ایران، بانک‌ها، شرکت‌های صنعتی و بازرگانی و سایر مراکز دولتی و خصوصی از عمده مراکز جذب دانش‌آموختگان این دوره می‌باشند.

شایان ذکر است که این گرایش‌ها، هیچ تفاوتی با رشته‌ی آمار ریاضی مصوب وزارت علوم، که هم اکنون در گروه آمار دانشکده‌ی ریاضی و علوم کامپیوتر دانشگاه صنعتی امیرکبیر برقرار می‌باشد، ندارد. صرفاً دروس اختیاری، سمینار و پروژه تحقیق (بجای پایان نامه) به صورت جهت‌دار در زمینه‌ی آمار رسمی یا داده کاوی، یا استنباط آماری، احتمال و فرایندهای تصادفی (یعنی گرایش آمار ریاضی)، اخذ می‌شوند.

با این وجود دروس اجباری تا حد امکان به صورت کاربردی ارائه می‌شود 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 می‌باشند، 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

آمار واژه‌ای 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 شکل‌های مختلفی (نمودار، جدول و...) نمایش داده می‌شوند، اطلاق می‌شود. در معنای دوم، واژه‌ی آمار عنوانی برای یک شاخه‌ی علمی است؛ مجموعه فنون گردآوری داده‌ها، طبقه بندی و تلخیص آن‌ها و سرانجام نتیجه‌گیری از داده‌ها. ارتباط این دو در این واقعیت نهفته است که اگر اعداد و ارقام پیش گفته با روش‌های علمی به دست آمده باشند باید با کمک علم آمار باشد. به عبارت دیگر، استفاده از

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

در تعریف واژه‌ی آمار به دو معنای آن و ارتباط آن‌ها با یکدیگر اشاره شد. این ارتباط در نقشی نهفته بود که علم آمار در تولید آمارها ایفا می‌کرد زیرا بخشی از مطالعات کاربردی این علم به حل مسائل واقعی و اجرایی در گردآوری داده‌ها و تولید آمارها اختصاص دارد. معمولاً هدف این مطالعات ابداع روش‌هایی است که آمارهایی با کیفیت بالا و با کم‌ترین هزینه را به دست دهند. بنابراین باید در نظام آماری، تولید آمارها به متخصصان و آمارشناسان واگذار شود تا استفاده از روش‌های علمی در تولید آمار نمود پیدا کند. در این خصوص نیاز به تاسیس گرایشی با نام آمار رسمی احساس می‌شود. تولید آمارهای رسمی نتیجه‌ی فعالیت‌های مختلفی است: وضع قوانین و مقررات، تهیه‌ی تعاریف و استانداردها، تعیین نیازها و اولویت‌ها، به کار گیری روش‌های علمی، طراحی ابعاد اجرایی، استخراج و تحلیل داده‌ها و سرانجام انتشار و اطلاع رسانی اطلاعات آماری. مطالعه این فرایند آن چیزی است که در رشته آمار رسمی مورد بررسی قرار می‌گیرد.

### معرفی برخی از درس‌های گرایش آمار رسمی

گرایش آمار رسمی، تلفیقی از چندین مبحث آمار، اقتصاد، جمعیت‌شناسی، مدیریت و ... می‌باشد. در ابتدای ورود به رشته‌ی آمار رسمی، قبل از هر چیز باید شناخت مقدماتی از محیطی که در آن آمارهای رسمی تولید می‌شوند، پیدا کرد. در درس **مقدمه‌ای بر آمارهای رسمی** جایگاه اطلاعات آماری در برنامه ریزی‌ها و در پی آن به نقش نظام‌های آماری برای تولید و عرضه‌ی اطلاعات آماری و به تشریح وظایف واحدهای مرکز آمار ایران و نقش هر یک در تولید آمارها پرداخته می‌شود.

آمارهای ثبتي که بخش عمده‌ای از آمارهای رسمی را در دل خود دارند و نیز کاربردهای آمارهای رسمی در تحلیل‌های اجتماعی - اقتصادی مانند حساب‌های ملی معرفی می‌شوند. اما تولید آمارهای رسمی به تنهایی کافی نیست و بررسی کیفیت این داده‌ها نقش مهم‌تری را ایفا می‌کند. به همین دلیل، در رشته آمار رسمی، شناخت روش‌های مختلف آمارگیری، تفاوت آن‌ها با یکدیگر، ابزارهای گردآوری داده‌ها و مسائلی پیرامون آن، بخش مهمی از مطالعات را به خود اختصاص می‌دهد که در درس **روش‌های آمارگیری** به بحث گذاشته می‌شوند.

یکی دیگر از مباحثی که کیفیت آمارهای رسمی را تحت تاثیر خود قرار می‌دهد، طرح‌های نمونه گیری می‌باشد. به همین جهت درس نمونه گیری کاربردی برای شناسایی انواع طرح‌های نمونه گیری، برآوردها، نحوه برخورد با بی پاسخی و ... در نظر گرفته شده است.

بدیهی است آنچه کیفیت داده‌ها را کاهش می‌دهد، بروز خطا است. در این رشته، چگونگی جزئیات آمارگیری‌ها که بر کیفیت نتایج آن‌ها اثر می‌گذارند و آن چه که در آمارگیری‌ها خطا شناخته می‌شود، مورد بحث قرار می‌گیرد و به بررسی جنبه‌های مختلف طراحی و اجزای آمارگیری‌ها می‌پردازد. شناسایی انواع خطاهای آمارگیری اعم از خطاهای نمونه‌گیری و غیر نمونه‌گیری در درس **خطای کل آمارگیری** بررسی می‌شوند.

برای اطلاعات بیشتر به سایت‌های زیر که مراجع‌های متن فوق نیز می‌باشند مراجعه کنید:

<http://www.srtc.ac.ir/~navvabpour>

<http://www.southampton.ac.uk/socsci/socstats/moffstat>

## داده کاوی Data Mining

با گسترش روزافزون فناوری رایانه‌ای و اینترنت، و مجهز شدن سازمان‌ها، شرکت‌های دولتی، خصوصی و مراکز دیگر به این ابزار امکان جمع آوری اطلاعات دقیق و به روز با قیمت ارزان برای آن‌ها فراهم آمده است، به طوری که این امکان منجر به ایجاد مجموعه وسیعی از داده‌ها شده است که شامل مقادیر زیادی از سوابق ثبت شده هستند. این سوابق مد نظر مدیران، برنامه‌ریزان و پژوهشگران بوده و از آن‌ها برای تهیه گزارش‌های مختلف و تصمیم‌گیری‌های روزمره و راهبردی استفاده می‌شود. از سوی دیگر، در قرن حاضر اطلاعات جمع آوری شده با رشد حیرت آوری مواجه شده به گونه‌ای که روش‌های مرسوم تحلیل داده‌ها امکان پذیر نیست و داده‌ها نیاز به حفاری برای استخراج اطلاعات از آن‌ها دارند. استخراج دانش از مجموعه‌ای بزرگ از داده‌ها، داده کاوی گویند. داده کاوی فرایندی است که با استفاده از تکنیک‌های هوشمند، دانش را از مجموعه‌ای از داده‌ها استخراج می‌کند. داده کاوی مجموعه‌ای از فنون برای کشف سریع و دقیق اطلاعات با ارزش از پایگاه‌های حجیم داده‌ای در هر کشور، سازمان و شرکتی به منظور توسعه علمی، فنی و اقتصادی است. دو روش برای داده کاوی وجود دارد یکی الگو یابی و دیگری مدل‌سازی. هدف گرایش داده کاوی آموزش کلاسیک روش‌های داده کاوی است.

داده‌کاوی پایگاه‌ها و مجموعه‌های حجیم داده‌ها، در پی کشف و استخراج دانش مورد تحلیل و کند و کاوهای ماشینی (و نیمه ماشینی) قرار می‌دهد. این گونه مطالعات و کاوش‌ها را به واقع می‌توان همان امتداد و استمرار دانش کهن و همه جا گیر آمار دانست. تفاوت عمده در مقیاس، وسعت و گوناگونی زمینه‌ها و کاربردها، و نیز ابعاد و اندازه‌های داده‌های امروز است که شیوه‌های ماشینی مربوط به یادگیری، مدل‌سازی، و تعلّم را طلب می‌نماید.

اصلاح Data Mining همان طور که از ترجمه آن به معنی داده کاوی مشخص می‌شود به مفهوم استخراج اطلاعات نهان و یا الگوها و روابط مشخص در حجم زیادی از داده‌ها به یک یا چند بانک اطلاعاتی بزرگ است. داده کاوی به بهره‌گیری از ابزارهای تجزیه و تحلیل داده‌ها به منظور کشف الگوها و روابط معتبری که تا کنون ناشناخته بوده‌اند اطلاق می‌شود. این ابزارها ممکن است مدل‌های آماری الگوریتم‌های ریاضی و روش‌های یاد

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 پیش بینی 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 مدل سازی در داده کاوی هستند:

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 ارایه می‌شود.

## 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.

اخیراً 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 پارامتری 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.

داده‌کاوی و آمار هر دو به طرح استنباط از داده‌ها مربوط می‌شوند. هدف از استنباط ممکن است درک الگوهای همبستگی و پیوندهای سببی بین مقادیر داده‌ها (توصیف)، یا ساخت پیشگویی‌ها از مقادیر داده‌های قبلی است. آمار کلاسیک دربرگیرنده مدلی برای توزیع احتمال داده‌ها و ساخت استنباط به صورت بیان احتمال است. روش‌های داده‌کاوی در بسیاری از موارد مسایلی را در برمی‌گیرد که به سادگی در چارچوب آمار کلاسیک برآزش داده نمی‌شوند و از آمار فاصله گرفته‌اند. در آمار ما با جامعه سر و کار داریم در حالی که داده‌کاوی با پایگاه داده‌ها سرو کار دارد. داده‌کاوی با نمونه‌گیری مانند آمار سرو کار ندارد، یعنی دوباره می‌توان از داده‌ها استفاده کرد. داده‌کاوی از میان مجموعه مدل‌های برآزش داده شده، بهترین مدل را انتخاب می‌کند، ولی در آمار تنها یک مدل برآزش می‌شود. آمار مدل برآزش داده شده را برای تعمیم به جامعه به کار می‌برد، ولی داده‌کاوی مدل برآزش داده شده انتخابی را برای پیشگویی داده‌های جدید به کار می‌برد.

مجموعه داده‌ها در داده‌کاوی می‌تواند خیلی وسیع‌تر و گسترده‌تر از مجموعه داده‌های معمول در آمار، شامل چند صد میلیون یا هزار میلیارد رکورد باشد.

**نوع داده‌ها** در آمار کمی است ولی داده‌کاوی داده‌های کمی، کیفی و متنی را پوشش می‌دهد.

**نوع متغیرهای ورودی** در آمار عددی ولی در داده‌کاوی متغیرها از نوع عددی، طبقه‌ای و متنی می‌باشند.

**نوع تمرکز** در آمار روی مدل ولی در داده‌کاوی تمرکز روی الگو است.

**زیر بخش‌های اصلی آمار** برآورد، توزیع‌های احتمال، آزمون فرضیه، امتیازبندی مدل و پیشگویی است، ولی زیر بخش‌های داده‌کاوی مدل بندی پیشگویی‌ها، بخش بندی پایگاه داده‌ها و ساخت فرضیه است.

**هدف اصلی** از به کار گیری روش‌ها در آمار استفاده از برآوردگرها و توزیع‌ها برای ادغام اطلاعات، ولی در داده‌کاوی هدف اصلی کشف دانش مورد علاقه است.

**شکل نتایج** در آمار به صورت مدل‌های کلی برآورد می‌شود ولی در داده‌کاوی مدل‌های موضعی محاسبه می‌شوند.

**ارزش اطلاعاتی** نتایج در آمار معلوم و محدود است در حالی که در داده‌کاوی نامعلوم و نامحدود می‌باشد.

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

در مقایسه با آمار، داده‌کاوی توجه کمتری به ویژگی‌های مجانبی 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

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 دادگان	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