چکیده

استفاده از داده‌های پروب در بهینه‌سازی زمان‌بندی چراغ راهنمایی

وسایل نقلیه‌ی پروب به عنوان یک راه جمع‌آوری اطلاعات ترافیکی زمان- حقیقی مورد نیاز سیستم‌های بهینه‌سازی، بیشتر توجهات را به خود جلب کرده است. در این پایاننامه یک سری فرمول‌های ریاضی یکپارچه، برای اعمال کنترل (بروی) چراغ راهنمایی ارائه شده است. در ابتدا، الگوریتم تشخیص زمان پیش‌روی خودروها موجود در گروه، برای شناسایی گروه‌ها با استفاده از اطلاعات بهنگام گرفته شده از وسایل نقلیه‌ی پروب، ارائه شده است. این الگوریتم صف‌های موجود و گروه‌هایی که در حال نزدیک شدن به هر تقاطع می‌باشند را شناسایی می‌کند.در ادامه یک برنامه‌ریزی خطی آمیخته با اعداد صحیح برای بدست آوردن جواب‌های بهینه‌ی چراغ مبتنی بر کنترلگرهای ترافیک موجود و داده‌های بهنگام گروه که از یک سری وسایل نقلیه خاص مانند اتوبوس‌های درون‌شهری دریافت می‌شود، حل شده است. شبیه‌سازی میکروسکوپیک با استفاده از $AIMSUN$ نشان می‌دهد، این فرمول‌های ریاضی می‌تواند تأخیر را برای همه‌ی وسایل نقلیه در شرایط غیر اشباع و فوق اشباع ترافیکی در مقایسه با عملکرد کنترل قدیمی هماهنگ- فعال چراغ راهنمایی با بهینه‌سازی زمان‌بندی نرم‌افزار $SYNCHRO$، کاهش دهد.

فهرست مطالب

عنوان صفحه

فصل ۱: مقدمات ۱
۱–۱ تاریخچه ۲
۱–۲ نظریه‌ی احتمال ۴
۱–۲–۱ امید ریاضی ۸
۱–۲–۲ واریانس ۸
۱–۲–۳ رگرسیون و همبستگی ۱۰
۱–۳ ساختار مختلف انتخاب ۱۱
۱–۳–۱ مقدمه ۱۱
۱–۳–۲ انتخاب در ساختار همزمان ۱۳
۱–۳–۳ انتخاب در ساختار مرحله‌ای ۱۳
۱–۳–۴ نظریه انتخاب ۱۴
۱–۳–۵ مدلهای انتخاب گسسته ۱۵

فصل ۲: معرفی وسایل نقلیه‌ی پروب و قابلیت‌های آنها ۱۸
۲–۱ مقدمه ۱۹
۲–۲ جمع‌آوری اطلاعات ترافیکی زمان–حقیقی ۲۰
۲–۲–۱ بدست آوردن اطلاعات ترافیکی زمان–حقیقی با استفاده از تلفن همراه ۲۳
۲–۳ شناسایی گروه‌هایی از وسایل نقلیه و تخمین پارامترهای مربوط به آن . . . ۲۷
۲–۴ تخمین طول صف وسایل نقلیه در یک تقاطع ایزوله با استفاده از داده‌های پروب ۳۴

فصل ۳: یک مدل برنامه‌ریزی خطی آمیخته برای زمان‌بندی چراغ راهنما با استفاده از داده‌هاب پروب ۴۵
۳–۱ برنامه‌ریزی خطی آمیخته با اعداد صحیح ۴۶
۳–۱–۱ استاندارد $NEMA$ ۴۷
۳–۱–۲ محدودیت‌های بدیهی مسأله ۴۹

۳–۲ دسته‌بندی و ارزیابی تأخیر گروه‌های نزدیک‌شونده به تقاطع ۵۶
۳–۲–۱ انتخاب چرخه‌های لازم برای عبور یک گروه از تقاطع ۵۸
۳–۲–۲ ارزیابی تأخیر گروه‌های وسایل نقلیه ۶۳

فصل ۴: مثال کاربردی و نتایج حاصل از شبیه‌سازی ۷۳
۴–۱ معرفی نرم‌افزار ۷۴
۴–۲ مثال ۷۹

فصل ۵: نتیجه‌گیری و پیشنهادات ۸۲

آ پیوست ۸۵
آ–۱ مدل‌های استفاده شده در نرم‌افزار $AIMSUNNG$ ۸۶
آ–۱–۱ مدل تولید وسایل ۸۶
آ–۱–۲ مدل تعقیب وسایل ۸۶
آ–۲ معرفی محیط اصلی نرم‌افزار $AIMSUNNG$ ۸۷
آ–۲–۱ پنجره صلی ۸۷
آ–۲–۲ نوار فهرست ۸۷
آ–۲–۳ محیط ترسیم دوبعدی ۸۸
آ–۲–۴ نوار ابزار ۸۸
آ–۳ رسم و ویرایش شبکه معابر ۸۸
آ–۳–۱ اجزاء شبکه ۸۹
آ–۳–۲ اطلاعات تقاضای ترافیک ۹۱
آ–۳–۳ اجرای مدل شبیه‌سازی ۹۲
آ–۳–۴ خروجی شبیه‌سازی ۹۲

مراجع ۹۵

واژه‌نامه فارسی به انگلیسی ۹۸

فهرست جدول‌ها

عنوان صفحه

جدول ۱–۱: پارامترهای توزیع هندسی ۱۰
جدول ۲–۱: کاربردهای اطلاعات ترافیکی زمان‌ـحقیقی ۲۱
جدول ۲–۲: مقایسه بین شناساگرها و وسایل نقلیه پروب ۲۲

فهرست شکل‌ها

عنوان صفحه

شکل ۲–۱: نمونه‌ای از دکل مخابراتی راه اندازی شده برای استفاده از تلفن همراه . ۲۳
شکل ۲–۲: بافت سلولی شبکه‌ی تلفن همراه ۲۴
شکل ۲–۳: تخمین مکان گوشی با استفاده از چندین دکل مخابراتی ۲۶
شکل ۲–۴: تصویری از توزیع وسایل نقلیه‌ی پروب در جریان ترافیک ۲۹
شکل ۲–۵: شناسایی پروفایل گروه‌ها تحت شرایط مختلف ۳۱
شکل ۲–۶: وسایل نقلیه در حال نزدیک‌شدن به تقاطع هستند ۳۴
شکل ۲–۷: امید ریاضی طول صف وسایل نقلیه با استفاده از وضعیت مکانی آخرین وسیله‌ی نقلیه پروب ۳۹
شکل ۲–۸: واریانس طول صف وسایل نقلیه با استفاده از وضعیت مکانی آخرین وسیله‌ی نقلیه پروب ۴۰
شکل ۲–۹: تابع احتمال مرزی L_p ۴۱
شکل ۲–۱۰ واریانس وزنی L_p ۴۲
شکل ۲–۱۱ واریانس خطا با استفاده از توزیع پواسون و دوجمله‌ای منفی ۴۳

شکل ۳–۱: استاندارد $NEMA$ برای یک تقاطع موجود در شبکه حمل و نقل شهری . . . ۴۸
شکل ۳–۲: کنترلگری با دو حلقه و هشت فاز ۴۹
شکل ۳–۳: گراف مربوط به کنترل دو حلقه‌ای در تقاطع i ۵۰
شکل ۳–۴: فازهای فعال در اولین چرخه از تقاطع iام ۵۲
شکل ۳–۵: انتخاب چرخه برای عبور کردن گروه از تقاطع iام ۵۹
شکل ۳–۶: گروهی از وسایل نقلیه که با تأخیر صف مواجه شده است ۶۴
شکل ۳–۷: تأخیر سیگنال برای گروه $(m, i, p, ۲)$ ۶۶
شکل ۳–۸: نمودار تجمعی رسیدن و خروج گروه (m, i, p, j)ام ۶۷
شکل ۳–۹: تأخیر سیگنال و صف در تقاطع بالا دستی ۷۰

شکل ۴–۱: مقایسه بین میانگین طول چرخه در روش پایاننامه با روش هاسچ و آلبک . . ۸۰

شکل ۴–۲: مقایسه‌ی تأخیر حاصل از روش پایاننامه با روش‌های $ASC\ Free$ و $ASC\ Coord$ ۸۱

شکل آ–۱: پنجره‌ی اصلی نرم افزار $AIMSUN$ ۸۷
شکل آ–۲: بازیابی فایل پس زمینه ۸۹
شکل آ–۳: تصویری از یک قطعه انتخاب شده ۹۰
شکل آ–۴: نماد گذاری مورد استفاده در قطعه ها ۹۰
شکل آ–۵: انواع خطوط جانبی ۹۱
شکل آ–۶: پنجره ویرایش $TrafficState$ ۹۲
شکل آ–۷: تنظیم فایل خروجی در پنجره‌ی سناریو ۹۴

–

فصل ۱

مقدمات

۱-۱ تاریخچه

امروزه به دلیل افزایش تقاضای سفر، مسأله‌ی ترافیک به عنوان یکی از عمده‌ترین مشکلات شهری به شمار می‌رود. به همین دلیل اقدام مناسب برای بهبود شرایط ترافیک ضروری به نظر می‌رسد. از دیدگاه مدیران ترافیک یکی از مسائل مهم برای بهبود شرایط ترافیک، یافتن زمان سفر مسیرهای مختلف شبکه می‌باشد. با توجه به اینکه مدت زمان سفر نه تنها برای کاربران شبکه‌ی حمل و نقل شهری جهت یافتن کوتاه‌ترین مسیرها ضروری است، بلکه بر روی تصمیمات مدیران ترافیک نیز تأثیر مستقیم دارد، بنابراین اولین تحقیقات در این زمینه توسط ووتن و همکاران [۱]، در سال ۱۹۹۵[۹]، ادبل و همکاران [۲]، در سال ۱۹۹۵ [۱]و دیا [۳] در سال ۲۰۰۲ انجام شده است. پس از آن مطالعات زیادی در این خصوص به عمل آمده است که ایده کلی آن‌ها بر اساس دو روش زیر است:

۱) سیستم‌های پیشرفته اطلاعات مسافر$(ATIS)$ [۴]: در این روش سعی می‌شود اطلاعات مربوط به شرایط موجود در شبکه را در اختیار مسافران قرار دهند تا بر اساس آن مسافرین بتوانند (انتخاب مسیر کوتاه‌تر، انتخاب اتوبوس یا دیگر وسایل نقلیه) داشته باشند. بر اساس مطالعات انجام شده توسط یانگ و منگ [۵] و [۶] تغییرات برخی از فاکتورها مانند زمان سفر، می‌تواند موجب تغییر

[۱] $Wootton\ et\ al.$
[۲] $Adbel-Aty\ et\ al.$
[۳] Dia
[۴] $Advanced\ Traveller\ Transportation\ Systems$
[۵] $Yang\ Meng$
[۶] $Lo\ Szeto$

انتخاب مسافران گردد.

۲) سیستم‌های پیشرفته مدیریت ترافیک ($ATMS$) [۱]: از پیشگامان این روش هورداکس [۲] در سال ۲۰۰۲ و بارسلو و همکاران [۳] در سال ۲۰۰۵ می‌باشند. این روش مبتنی بر مدیریت بهینه سیستم حمل و نقل، با داشتن اطلاعات شبکه است. محققین در این بخش تلاش کرده‌اند با در اختیار داشتن اطلاعات لازم از شرایط شبکه‌ی حمل و نقل، سیستم را به صورت بهینه اداره کنند؛ به عبارت دیگر با مدیریت منابع و امکانات موجود اهداف خاصی را برآورد کنند (مانند کمینه کردن زمان سپری شده در خیابان). برای بدست آوردن اطلاعات دقیق‌تر از شرایط ترافیک لازم است که این اطلاعات توسط وسایل نقلیه مجهز به GPS [۴]، که روی مسیرهای غیر از ویژه حرکت می‌کنند، جمع‌آوری گردد؛ به عنوان مثال اگر یک اتوبوس از یک مسیر آمیخته استفاده کند (قسمتی از مسیر منحصر به اتوبوس و قسمت دیگری از مسیر بین اتوبوس و دیگر خودروها مشترک باشد) تنها داده‌های مسیر مشترک برای برآورد شرایط ترافیک مورد استفاده قرار می‌گیرند.

همچنین الگوریتم‌ها و نرم‌افزارهای شناخته شده زیادی برای کنترل جریان ترافیک وجود دارند که بر اساس گروه‌های در حال حرکت کار می‌کنند. از جمله‌ی این نرم‌افزارها می‌توان به $VISSIM$، $SYNCHRO$، $TRANSYT-7F$، $AIMSUN$ و ... اشاره کرد. در نرم‌افزار شبیه ساز ماکروسکوپیک TRANSYT-7F است که پراکندگی گروه‌ها و جریان‌های ثانویه در آن در نظر گرفته می‌شود[۲۴]. این نرم‌افزار، سیستم کنترل تطبیقی جریان ترافیکی SCOOT را تکامل می‌بخشد[۱۶]. در سال ۱۹۹۵ دل اولمو [۵] و میرچاندانی [۶] الگوریتمی به نام REALBAND پیشنهاد کردند که هدف آن برطرف کردن ناسازگاری (بی‌نظمی حرکت) وسایل نقلیه بود.

[۱] *Advanced Traffic Management Systems*
[۲] *Hourdakis*
[۳] *Barcelo et al.*
[۴] *Global positioning System*
[۵] *Dell'Olmo*
[۶] *Mirchandani*

REALBAND به جمع‌آوری اطلاعات وسایل نقلیه در شبکه‌ی حمل و نقل شهری بر اساس داده‌های به دست آمده از شناساگرها می‌پردازد و از آن داده‌ها برای کمینه کردن مجموع کل تاخیرها استفاده می‌کند[۱۱]. اخیراً الگوریتمی برای کنترل جریان ترافیک با استفاده از چراغ راهنما توسط جیانگ [۱] و همکارانش برای تقاطع‌هایی با حجم ترافیک سنگین و سبک طراحی شده است[۱۸]. در گذشته مشخص کردن پارامترهای ترافیک با استفاده از شناساگر، از نظر اقتصادی بسیار هزینه‌بر بود. اما امروزه با استفاده از وسایل نقلیه مجهز به دستگاه‌های ارتباطی، فرستنده‌های الکترونیکی در سیستم‌های الکترونیکی اخذ عوارض (ETC) [۲] و گوشی‌های تلفن همراه مجهز به GPS این کار راحت‌تر و بسیار سریع‌تر و ارزان‌تر نسبت به روش استفاده از شناساگرها قابل انجام می‌باشد[۱۵].

۱-۲ نظریه‌ی احتمال

در زندگی روزمره با تجربیات و آزمایش‌هایی روبرو می‌شویم که نتایج حاصل از تکرارها همواره یکسان نمی‌باشد؛ به عبارت دیگر نتیجه (برآمد) حاصل از یک آزمایش دقیقاً قابل پیش‌بینی نیست. با این حال مجموعه‌ی نتایج حاصل از آن تجربه، بطور کامل مشخص است؛ این مجموعه را فضای نمونه می‌گویند و با حرف S نشان می‌دهند. هر زیر مجموعه از فضای نمونه یک پیشامد نامیده می‌شود. هر عملی که تصادف در آن دخالت داشته باشد، در اصطلاح تئوری احتمال، آزمایش تصادفی نامیده می‌شود. بطور دقیق آزمایش تصادفی عبارت است از عملی تحت شرایطی ثابت و برحسب موضوعی مورد علاقه، هر بار تکرار شود، منجر به یکی از اعضای مجموعه S گردد.

تعریف ۱.۱: یک فضای نمونه را که شامل تعداد متناهی یا نامتناهی ولی شمارش‌پذیر عضو

---

[۱] *Jiang*

[۲] *Electronic Toll Collection*

است، فضای نمونه‌ی گسسته گویند؛ مانند فضای نمونه پرتاب یک سکه یا یک تاس که تعداد عناصر فضای نمونه آنها متناهی است.

تعریف ۲.۱: اگر فضای نمونه شامل مجموعه‌ی تمام اعداد بین دو حد مشخص باشد، اصطلاحاً آن را فضای نمونه پیوسته گویند، مانند مدت زمانی که کارگری برای کار روی قطعه‌ای صرف می‌کند.

تعریف ۳.۱: هر زیر مجموعه از فضای نمونه را یک پیشامد می‌گویند.

برای محاسبه میزان شانس یک پیشامد از مفهوم نظریه احتمال استفاده می‌شود. احتمال A را با نماد $P(A)$ نمایش می‌دهیم. احتمال‌ها علاوه‌بر اینکه مقادیر یک تابع مجموعه‌ای هستند باید در اصول موضوعه‌ی احتمال نیز صادق باشند. در زیر اصول موضوعه‌ی احتمال، زمانی که فضای نمونه گسسته باشد، آورده شده است[۲۷].

اصل ۱) احتمال هر پیشامد، عددی حقیقی و نامنفی است؛

اصل ۲) $P(S = ۱)$

اصل ۳) اگر $\{A_i\}_{i\in I}$ دنباله‌ی متناهی یا نامتناهی از پیشامدهای دوبه‌دو ناسازگار از S باشند آنگاه $P(\cup_{i\in A_i}) = \sum_{i\in I} P(A_i)$.

در بسیاری از مسائل مربوط به احتمال، ما تنها به جنبه‌ای خاص (یا دو یا چند تا) ا نتیجه‌های یک آزمایش توجه داریم. برای مثال، در برخی از موارد وقتی یک جفت تاس پرتاب می‌کنیم، تنها مجموع دو شماره‌ای که ظاهر می‌شوند مورد توجه است. در این مثال توجه ما به اعدادی است که با نتیجه‌ی یک آزمایش شانسی همراه هستند. بنابراین، توجه ما به مقادیری است که آنچه اصطلاحاً متغیر تصادفی خوانده می‌شود، اختیار می‌کند.

در بسیاری از موارد هنگام تکرار یک آزمایش تصادفی بیشتر به تعداد دفعاتی که یک پیشامد رخ داده است علاقمند هستیم؛ به عنوان مثال در آزمایش‌های تکراری پرتاب یک سکه، تعداد

دفعاتی که شیر رؤیت شده مهم باشد نه ترتیب آن. چنین کمیتی را یک متغیر تصادفی می‌نامند. باید توجه داشت که مقادیر این کمیت قبل از انجام آزمایش مشخص نیست. متغیرهای تصادفی بر اساس مقادیری که می‌توانند اختیار کنند به دو دسته‌ی گسسته و پیوسته تقسیم می‌شوند.

تعریف ۴.۱: متغیر تصادفی X را یک متغیر تصادفی گسسته گویند، هرگاه تعداد مقادیری که می‌تواند اختیار کند متناهی یا نامتناهی شمارش پذیر باشد.

تعریف ۵.۱: متغیر تصادفی X را یک متغیر تصادفی پیوسته گویند، هرگاه مقادیری که می‌تواند اختیار کند از یک مجموعه‌ی نامتناهی باشد؛ به عبارت دیگر مقادیر متغیر تصادفی پیوسته X در یک فاصله یا بازه قرار دارد.

مثال ۶.۱: سکه ناآریبی را دو بار به تصادف پرتاب می‌کنیم. تعداد شماره شیرهای رو شده در این آزمایش تصادفی را با X نشان می‌دهیم. واضح است که از یک پیشامد به پیشامد دیگر، مقدار X تغییر می‌کند. مثلاً اگر پیشامد HH رخ دهد X برابر ۲ و اگر پیشامد HT رخ دهد X برابر ۱ می‌شود. افزون بر آن در این تغییر، عامل شانس حاکم می‌باشد. مثلاً احتمال اینکه X برابر ۲ شود $\frac{1}{4}$ و احتمال اینکه برابر ۱ شود $\frac{1}{2}$ می‌باشد. از اینرو X را یک متغیر تصادفی می‌نامند.

تعریف ۷.۱: مجموعه تمام مقادیری که متغیر تصادفی X اختیار می‌کند، تکیه‌گاه X نامیده می‌شود و با S_x نمایش داده می‌شود.

تعریف ۸.۱: اگر X یک متغیر تصادفی گسسته باشد، تابعی که برای هر مقدار x متعلق به تکیه‌گاه X به صورت $f(x) = P(X = x)$ نشان داده می‌شود را تابع چگالی احتمال متغیر تصادفی X می‌نامیم.

تعریف ۹.۱: تابع $f(x)$، که به ازای هر x متعلق به R مقدار آن نامنفی است، چگالی احتمال یک متغیر تصادفی پیوسته‌ی X خوانده می‌شود اگر و تنها اگر :

الف) به ازای هر دو عدد حقیقی a و b داشته باشیم $P(a \leq x \leq b) = \int_a^b f(x)dx$

۲) $\int_{-\infty}^{\infty} f(x)dx = ۱$.

تعریف ۱.۱۰: توزیع برنولی: فرض کنید متغیر تصادفی X، تعداد موفقیت‌ها در یک آزمایش برنولی باشد آنگاه متغیر تصادفی X دارای توزیع برنولی با تکیه‌گاه $S_x = \{۰, ۱\}$ دارای توزیع برنولی با توزیع احتمال روبه‌رو می‌باشد: $P(x = ۱) = p \quad , \quad P(x = ۰) = q$

تعریف ۱.۱۱: توزیع برنولی(دو نقطه‌ای): آزمایش برنولی را یک بار انجام می‌دهیم . هدف، بررسی تعداد موفقیت‌های حاصل است که می‌تواند ۰ یا ۱ باشد.

تعریف ۱.۱۲: توزیع دوجمله‌ای: متغیر تصادفی X توزیع دوجمله‌ای دارد، و به آن عنوان متغیر تصادفی دوجمله‌ای داده می‌شود، اگر و تنها اگر توزیع احتمال آن به صورت زیر باشد:

$$P(x, n, p) = \begin{pmatrix} n \\ \\ k \end{pmatrix} p^x (۱ - p)^{n-x} \quad x = ۰, ۱, ۲, ..., n. \qquad (۱-۱)$$

پس، تعداد پیروزی‌ها در n امتحان، متغیری تصادفی است که توزیع دو جمله‌ای با پارامترهای n و p دارد. نام توزیع دوجمله‌ای از این واقعیت نتیجه می‌شود که مقادیر $P(x, n, p)$ به ازای $x = ۰, ۱, ۲, ..., n$، جملات متوال بسط دو جمله‌ای $[p + (۱ - p)]^n$ هستند؛ این مطلب نشان می‌دهد که مجموع احتمال‌ها، همان‌گونه که باید، مساوی ۱ است.

تعریف ۱.۱۳: توزیع دوجمله‌ای منفی : متغیر تصادفی X، توزیع دوجمله‌ای منفی دارد و به آن، عنوان متغیر تصادفی دوجمله‌ای منفی داده می‌شود، اگر و تنها اگر توزیع احتمالش به ازای، $x = k, k + ۱, k + ۲, ...$ به صورت زیر باشد:

$$P(x, k, p) = \begin{pmatrix} x - ۱ \\ \\ k - ۱ \end{pmatrix} p^x (۱ - p)^{n-x}. \qquad (۲-۱)$$

تعریف ۱۴.۱: توزیع هندسی: متغیر تصادفی X، توزیع هندسی دارد، و به آن، عنوان متغیر تصادفی هندسی داده می‌شود، اگر و تنها اگر توزیع احتمال آن به صورت زیر باشد:

$$P(x,p) = p(۱-p)^{x-۱} \quad x = ۱,۲,۳,... \tag{۳-۱}$$

تعریف ۱۵.۱: توزیع پواسون: متغیر تصادفی گسسته X دارای توزیع پواسون با پارامتر λ است، اگر و تنها اگر تابع چگالی احتمال آن به صورت زیر

$$P(x,\lambda) = \frac{\lambda^x e^{-}\lambda}{x!} \quad x = ۰,۱,۲,۳,..., \tag{۴-۱}$$

باشد. امید ریاضی و واریانس توزیع پواسون، هر دو برابر با λ می‌باشند.

۱-۲-۱ امید ریاضی

اگر X یک متغیر تصادفی گسسته با تکیه‌گاه $S_x = \{x_۱, x_۲, x_۳, ...\}$ باشد آنگاه امید ریاضی X را با $E(X)$ نشان داده و به صورت زیر تعریف می‌شود:

$$E(x) = \sum_{x_i \in S_x} P(X = x_i). \tag{۵-۱}$$

همچنین برای متغیر تصادفی پیوسته‌ی X داریم:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx. \tag{۶-۱}$$

۱-۲-۲ واریانس

همان‌گونه که مشاهده می‌شود امید ریاضی یک متغیر تصادفی، میانگین وزنی مقادیر متغیر تصادفی را محاسبه می‌کند. بنابراین از آن به عنوان معیاری مناسب برای تمرکز آن متغیر

می‌توان استفاده کرد. این در حالی است که در مقایسه‌ی بین دو متغیرهای تصادفی گاهی لازم است میزان پراکندگی یک متغیر، حول امید ریاضی آن را بدست آوریم. یکی از مهم‌ترین شاخص‌های پراکندگی حول میانگین، واریانس است.

تعریف ۱۶.۱: متغیر تصادفی X مفروض است، واریانس X از رابطه زیر محاسبه می‌شود:

$$Var(X) = \sigma^2 = E(X^2) - (E(X))^2 \qquad (۷-۱)$$

اگر واحد اصلی مقادیر مشاهده شده برای متغیر تصادفی X، m باشد، آن‌گاه واحد اندازه‌گیری واریانس، مجذور واحد اصلی یعنی m^2 است؛ از همین رو جذر مثبت واریانس را انحراف معیار (σ) می‌نامند.

آزمایش تصادفی را در نظر بگیرید که نتیجه‌ی آن دو پیشامد موفقیت S و یا شکست F به ترتیب با احتمال $P(S) = p$ و $P(F) = q = ۱ - p$ باشد. این آزمایش را یک آزمایش برنولی می‌نامند. بر اساس این آزمایش متغیرهای تصادفی مختلفی را می‌توان تعریف کرد که در ادامه توزیع احتمال مربوط به بعضی از آنها را بررسی می‌کنیم. ضریب پراکندگی یکی از مهم‌ترین معیارهای پراکندگی نسبی در هنگام بررسی جوامع و اختلاف آن‌ها با یکدیگر است که در صورت بروز شرایط زیر مورد استفاده قرار می‌گیرد.

- میانگین دو جامعه مورد مقایسه، با هم متفاوت باشد؛
- واحد اندازه‌گیری دو جامعه مورد مقایسه، با هم متفاوت باشد.

تعریف ۱۷.۱: حاصل تقسیم انحراف معیار به میانگین را ضریب پراکندگی می‌نامند و آن را با CV نشان می‌دهند.

با توجه به کاربرد گسترده‌ی توزیع هندسی در این تحقیق پارامترهای این توزیع را که شامل

تابع احتمال، امید ریاضی، واریانس و انحراف معیار می‌باشد در جدول زیر آورده شده است.

جدول ۱-۱: پارامترهای توزیع هندسی

احتمال وقوع اولین موفقیت در x امین آزمایش	$p(x) = q^{x-1}$	تابع احتمال
میانگین تعداد آزمایش مورد انتظار برای رسیدن به اولین موفقیت	$\frac{1}{p}$	امید ریاضی
واریانس تعداد آزمایش لازم برای رسیدن به اولین موفقیت	$\sigma_x^2 = \frac{q}{p^2}$	واریانس
انحراف معیار	$\sigma_x = \sqrt{\frac{q}{p^2}}$	

۱-۲-۳ رگرسیون و همبستگی

مقدمه

یکی از اهداف اصلی بسیاری از پژوهشهای آماری ایجاد وابستگی‌هایی است تا پیش‌بینی یک یا چند متغیر را بر حسب سایرین میسر گرداند.به عنوان مثال مطالعاتی انجام شود تا فروشهای بالقوه یک محصول جدید را بر حسب قیمت آن، مخارج سرگرمی‌های خانواده را بر حسب درآمد آن و مصرف سرانه‌ی برخی مواد غذایی را بر حسب ارزش غذایی آنها پیش‌بینی کند.

البته گرچه مطلوب آن است که بتوان کمیتی را بر حسب سایرین دقیقاً پیش‌بینی کرد، ولی این کار به ندرت میسر است.

به طور صوری، اگر توزیع توأم دو متغیر تصادفی X و Y را داشته باشیم و بدانیم که X مقدار x را اختیار می‌کند، مسأله‌ی اصلی رگرسیون دو متغیره عبارت است از تعیین میانگین Y به ازای مقدار مفروضی از X است.

اگر $f(x,y)$ مقدار چگالی توأم دو متغیر تصادفی X و Y در (x,y) باشد، مسأله رگرسیون دو متغیره صرفاً عبارت است از تعیین چگالی شرطی X به شرط $X = x$ و سپس با محاسبه انتگرال

$$E(Y|_x) = \int_{-\infty}^{\infty} y.w.(x|_y)dy \qquad (۸-۱)$$