دانشگاه تهران

جزوه کلاسی درس:

الگوسازی یا کاربرد مدل های کمّی در شهرسازی

مدرّس: دکتر اسفندیار زبردست

گردآورنده: خباز

مقدمه:

مدل ها جهان واقعی را به شرایط ساده تر تبدیل می کنند تا آن را درک و مدیریت کنند. مدل نمودی از واقعیت است.

تأکید بر مدل های کمی و ریاضی می باشد.

ساده ترین، مدل هنسن (Hansen) می باشد از ۱۹۵۹ تاکنون. این مدل یک مدل مکانی، برای تخصیص جمعیت می باشد.

چه عواملی باعث جذب جمعیت به مناطق می شوند؟

۱ـ دسترسی به اشتغال

۲ـ امکان توسعه از نظر اسکان (زمین بایر با کاربری مسکونی)

ترکیب این دو شاخص پتانسیل توسعه را نشان می دهد: (شاخص دسترسی: Accessibility index، شاخص Holding capacity : ظرفیت نگه داشت یا امکان رشد)

به دلیل تفاوت رفتار انسانها و ...، دقت مدل در شهرسازی نمی تواند خیلی بالا باشد. مثلاً دقت بالای ۴۰٪ خوب است.

هرجا عامل تولید، جذب و فاصله در کار باشد از عامل جاذبه استفاده می شود.

مدل جاذبه محدودیت تک قیدی (۱ـ محدودیت جذب ۲ـ محدودیت تولید)

در بحث های توسعه مراکز خرید در شهر: مرکز خرید بعدی کجا احداث شود.

مدل استیوارد، مبتنی بر قانون جاذبه نیوتن: تبیین میان کنش فضایی

در قاعده رتبه اندازه، هر چه شیب از 45° انحراف داشته باشد توزیع نامتعادلتر است.

مدل های خطی:

رگرسیون ساده و چندمتغیره: رابطه کمی بین دو یا چند متغیر را بیان می کند.

مدل های بهینه یابی یا بهینه سازی

مدل های Integrated از ۱۹۹۰ به بعد مورد توجه هستند. با نرم افزارهای GIS

<u>مدل ها، انواع مدل ها و تاریخ استفاده از آنها:</u>

اساساً مدل یک بازسازی ساده واقعیت است که سعی می کند پیچیدگی آشکار جهان واقعی را از آن گرفته، آن را به حالت ساده قابل مدیریت تبدیل کند. مدل نمودی از واقعیت است. کار مدل تبدیل متغیرهای مختلف به متغیرهای اندک قابل مدیریت است. دسته بندیهای مختلف مدلها به صورت زیر است:

مدلها
- تصویری
- قیاسی
- ریاضی

(تصویری، قیاسی، ریاضی) × (ایستا، پویا)

- ایستا
- پویا

(ایستا، پویا) × (احتمالی، معین)

- احتمالی: $\text{if } A \xrightarrow{\text{احتمالاً}} B$
- معین: $\text{if } A \longrightarrow B$

عمده تأکید بر مدلهای ریاضی می باشد.

مدلهای تصویری: **Iconic Models**

مدلهای کوچک مقیاس برای نشان دادن توسعه شهری. این مدلها صرفنظر از مقیاس شبیه به آن چیزی هستند که نمایش می دهند مانند عکس های هوایی، نقشه ها، ماکت ها (۳D) ،نقشه هوایی از طریق استروسکوپ

مدلهای قیاسی: **Analog Models**

قابل مقایسه با (شبیه به) اجسام یا موقعیتهای واقعی هستند اما هیچ همشکلی با آنها ممکن است نداشته باشند. مثل خطوط تراز، تونالیته های رنگ برای شدت تراکم.

برخلاف مدلهای تصویری، این مدلها نیاز به لژاند دارند.

مدلهای ریاضی یا انتزاعی: **Abstract Models**

(مدلهای نمادین یا سمبولیک) symbolic

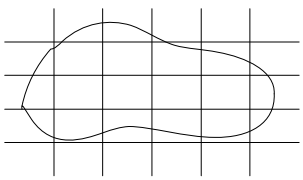
با افکار انتزاعی شروع شده و با استفاده از علائم، رمزها و سمبولها ثبت می شوند. مثل سمبول H_2O برای فرمول شیمیایی یا معادلات ریاضی x و y و ... و ممکن است تنها یک رابطه نباشد. از طریق معادلات ریاضی سیستم ساده می شود.

مدلهای ریاضی	مدلهای توصیفی	Descriptive Models
	مدلهای پیش بینی کننده	Predictive Models
	مدلهای برنامه ریزی	Planning Models
	(یا سنجش یا ارزیابی یا تجویزی)	evaluative

مدلهای توصیفی برای بیان یک سری روابط در یک مقطع زمانی به کار می روند. عمدتاً در تحلیل وضع موجود از آنها استفاده می شود.
می توان از تحلیل رگرسیونی برای مدل توصیفی استفاده کرد.
از این مدل برای پیش بینی نیز می توان استفاده کرد. در مدلهای پیش بینی کننده یکسری روابط توصیفی را با در نظر گرفتن عامل زمانی به آینده می بریم.
در مدلهای برنامه ریزی معمولاً معیارهایی را انتخاب می کنیم تا از بین گزینه های موجود در شهر گزینه مطلوب را انتخاب کنیم.
مثلاً مدل گام یا AHP یا Smart
این مدلهای ریاضی ساخت اولیه شان سخت است ولی به راحتی قابل جرح و تعدیل و اصلاحند که از محسنات این مدل محسوب می شود.
مثل مدل Meplan که در جاهای مختلف استفاده شده است.
Integrated land use transportation Models
ساختار مدل همان ساختار دهه ۷۰ است با ساختار داده- ستانده ولی خود را با تحولات روز هماهنگ کرده، نوع داده ها متفاوت شده است.
مدلها عمدتاً از دهه ۱۹۵۰ به بعد در شهرسازی مورد توجه قرار گرفتند. در جنگ جهانی دوم آمریکاییها با مشکل جدیدی به نام ترافیک، ازدحام، آلودگی هوا و از دست رفتن فرصتها در ترافیک مواجه شدند. پس به دنبال راه حل رفتند. سه راه حل مورد نظر بود:

۱. ساخت خیابان ها و اتوبانهای جدید
۲. ساخت پل روی رودخانه های درون شهری
۳. سیستم حمل و نقل عمومی کارآمدی را طراحی و اجرا کنند.

هر سه گزینه هزینه بر هستند. پس به فکر جمع آوری و تحلیل اطلاعات و انتخاب بهترین گزینه افتادند. اما با حجم زیادی از اطلاعات مواجه شدند که تحلیل آنها خود نیاز به یک مدل داشت. کمک مالی دولت آمریکا به نهادهای تحقیقاتی و پژوهشی برای یافتن راه حل صورت گرفت.
ظهور کامپیوتر در آن زمان و اجرای تحلیلهای ریاضی در مدت زمان کوتاه و فیدبک زدن اتفاق دوم بود.
اتفاق سوم پیشرفت در علوم بود. قبل از آن محاسبات به صورت کاملاً سنتی بود. تحلیلهای ماتریسی به وجود آمد. علوم تحقیق در عملیات و ... نیز به وجود آمد.
مطالعات متعددی همزمان راجع به تراکم و دسترسی و فاصله از مرکز شهر و ... شروع شد. مثلاً با فاصله گرفتن از مرکز شهر تراکم کمتر می شود، مثل نظریه دسترسی هنسن. نظریاتی نیز راجع به نحوه توسعه شهری ارائه شد و اینها باعث شکل گیری مدلهای اولیه شد.
یکی از مدلها برای بررسی مسائل حمل و نقل در شیکاگو مدل CATS بود که در ۱۹۵۵ شروع شد.
The Chicago Area Transportation Study
می خواستند نرخ تولید سفر را به دست آورده و برای پیش بینی تولید سفر در ۱۹۸۰ استفاده کنند. تیمی متشکل از ریاضیدانان، حقوقدانان، متخصصین کامپیوتر، شهرسازی، مهندسان و ... تشکیل شده و به سرپرستی Donglas carroll شروع شد.



از ابتدا وابستگی زیاد بین حمل و نقل و کاربری اراضی را می دانستند پس نمی توانستند بدون یکی برای دیگری برنامه ریزی کنند. کل شهر را به صورت گرید در نظر گرفتند. که مساحت هر شبکه ۱ مایل مربع بود و اسم آنها را زونهای ترافیکی گذاشته بودند.
کاربری اراضی در ۵۰۰۰۰ بلوک شهری را مورد بررسی قرار داده و به ۱۴ نوع کاربری طبقه بندی کرده و برداشت کردند.
۱۹۶۲ خاتمه مدل.
در مدل CATS، پیک کار بر روی مدل: ۱۹۵۶
پس از آن به دنبال PATS رفتند که در ۱۹۵۶ شروع شد:

The Pittsburgh Area Transportation Study

و به دنبال آن (DMATS)

Detroit Metropolitan Area Transportation Study

نتایج آنها در ۱۹۶۳ منتشر شد.

۱۹۶۳ خانم لاری مدل خود را بر پیتزبورگ می دهند.

۱۹۵۹ مدل مکانی هنسن

مدل UNC توسط آقای Chapin و همکاران Uunversity of North Carolina

این مدل با رگرسیون چند متغیره کار می کرد. کارکرد مدل برای پیش بینی این بود که کدام زمین شهری زودتر به زیر ساخت می رود. (پیش بینی جهات توسعه)

شهر را گریدبندی می کند. گریدهای کوچکتر با مساحت ۱ آکر

acre ≃ 0.405 Hec هکتار

برای هر گرید شاخصی به نام attractiveness index (شاخص جذابیت)طراحی کردند که شاخص جذابیت هر cell ترکیب خطی از ۷ متغیر می باشد:

۱. ارزش اولیه زمین
۲. وجود فاضلاب عمومی
۳. دسترسی به نزدیکترین خیابان اصلی
۴. دسترسی به نزدیکترین مدرسه ابتدایی
۵. دسترسی به نزدیکترین نواحی کار

دو شاخص دیگر دارند که اهمیت کمتری دارند.

Cell ای که شاخص جذابیت بیشتر دارد زودتر به زیر ساخت می رود. Cell ها در محدوده قانونی شهر هستند.

حریم برای توسعه آتی شهر است. هر نوع ساخت و ساز در آن ممنوع است به جز گسترش روستای واقع در آن. حد عرف نقشه landuse پیشنهادی برای افق طرح و ضوابط و مقررات ساخت و ساز در طرح جامع تثبیت می شوند. که در آن ساخت و ساز در حریم مشخص می شود.

مدل لاری ۱۹۶۳ The Model of Metropolice Airo & Lary

مدل لاری: پایه بیش از ۹۰٪ مدل هایی است که بعدها ساخته شده و می شود. به بیان دیگر تقریباً ۹۰٪ مدلهای کنونی منطق مدل لاری را دارند (مدلهای integrated)

در مدل لاری به سراغ نظریه اقتصاد پایه می رویم تا شهر را ساده کنیم (ساختار مبتنی بر نظریه اقتصاد پایه).

فعالیتها در شهر
- پایه: موتور توسعه
- غیرپایه: تبعی

اقتصاد شهر را بر مبنای تئوری اقتصاد پایه که ۱۰ سال از آن می گذشت بررسی کرد. فعالیت های تبعی به دنبال فعالیتهای پایه حرکت می کنند. سه روش برای تشخیص فعالیتهای پایه و غیرپایه داریم.

تشخیص بخشهای پایه و غیر پایه
- مستقیم
- غیرمستقیم
  - روش فرضی
  - روش ضریب مکانی
  - روش حداقل نیاز

شاغلین پایه به مناطق اسکان هدایت شده سپس مشاغل خدماتی را پیدا کرده و تخصیص می دهد. جمعیت وابسته را پیدا و ... همه از خدمات هم بهره مند می شوند.

Input-output : اصلاح مدل لاری: اطلاعات بیشتر و دقت بیشتر

در ۱۹۶۶ گرین، مدل لاری را اصلاح می کند. به جای هنسن برای تخصیصها از مدل جاذبه استفاده می کند: مدل گرین ـ لاری

برای اصلاح عملکرد مدل لاری:

مایکل هرین ـ مارشال اشنیکاس ـ مدلهای TOM۱ و TOM۲ ـ دیوید فود ـ الن ویلسون

الن ویلسون خانوارها با درآمدهای مختلف را در نظر گرفت. سطح دستمزد را با محل اشتغال در نظر گرفت و انواع واحدهای مسکونی و تغییرات در واحدهای مسکونی. اما مدل پیچیده تر شد. برای به دست آوردن اطلاعات باید فرض کرد.

مدلهای دیگر: لیزبون، توپاز (روی برنامه ریزی خطی کار می کند در استرالیا) TOPAZ در دهه ۱۹۶۰

هزینه های حمل و نقل را کمینه و دسترسی را بیشینه می کند. برای پیش بینی توسعه شهری در ملبورن

مدل کاراکاس، مدل بث، MBER(۷۲-۷۰)، استکهلم

تا در ۱۹۷۳، Douglass lee مقاله ای تحت عنوان فاتحه خوانی مدلهای بزرگ مقیاس نوشت و در آن ۷ دلیل برای کارآمد نبودن آنها آورد.

که تأثیر زیادی روی روند مدلها گذاشت و نقطه عطفی برای تجربه مدلسازی تلقی می شود.

Jornel of American Planning Association

۱. Hyper comprehensiveness	بیش از حد جامع بودن
۲. Grossness	کلی بودن (به جزئیات نمی توانند بپردازند)
۳. Data hungriness	داده خوار بودن (اطلاعات زیادی طلب می کنند)
۴. Wrong headedness	بی جهت و به بیراهه رفتن (جهت مناسب نداشتن)

مثلاً اطلاعات در شهر جمع آوری شده و output همسایگی می خواهند.

۵. Complicatedness	پیچیده بودن

استفاده از آنها راحت نبود

۶. Mechanicalness	ایده ها فازی را در نظر نمی گیرند، فقط صفر و یکند.
۷. Expensiveness	بیش از حد گران بودن

بعد از این مقاله یک دوران افول داشتیم یعنی دهه ۱۹۷۰ مدلسازی جدید نداشتیم. ولی باعث شد که نوع مدلها تغییر کند و همزمان شد با پیدایش کامپیوترهای شخصی که این امکان را فراهم می کردند در مدت کوتاه و هزینه کم مدل طراحی شود.

مدلها از بزرگ مقیاس تبدیل به کوچک مقیاس و موضوعی شدند.

رساله دکتری Lee (معمار با فوق لیسانس و دکترای شهرسازی ۱۹۶۰ دانشگاه کرنر): مدل TOM (Time Oriented Metropolitan Model)

ایراد مدل لاری (ایستا): قبلاً این مدل را برطرف کرده بود

دهه ۱۹۸۰، PC ها نیز معرفی شدند. در همه جا در دسترس بودند حتی مؤسسات خصوصی کوچک. پس به صورت بهینه، Run و آزمون مدل با هزینه کم در فضای کوچک اتفاق می افتاد.

اواخر ۱۹۸۰ و ۱۹۹۰ : معرفی GIS

تا آن زمان مدلها خود را با جداول تطبیق داده و جدول را به نقشه تبدیل می کردند. ولی با GIS مدلها خود را با آن هماهنگ کرده و به جای جدول تحلیل روی نقشه انجام می شد.

مدل CUF۱ و CUF۲ : John Landis که بر روی Arc view کار می کند. California Urban Future model

Mikhail wegener آلمانی در مدلها به طبقه بندی Operational urban models (مدلهای کاربردی) رسیده است و مدلهای دیگر آکادمیک هستند. مدلهایی که حداقل برای یک شهر یا یک منطقه شهری به کار گرفته شده باشند عملکردی هستند.

MEPLAN (Marcial Tchenique)در تهران نیز در دهه ۱۹۷۰ به کار گرفته شد، ولی اکنون GIS based شده است.

What if : ریچارد پلاسترمن

۱۹۹۴ از Lee خواستند که نظرات کنونی‌اش را راجع به مدلها بنویسد.

Lee عذرخواهی کرد بابت دو چیز ۱ـ زبان تند ۲ـ هنوز پاسخی به سوالها نیافتم.

به گفته او: یک تئوری باید پشت مدلها باشد و هنوز تئوریهایی که عملکرد مدلها را تغییر دهد ندیده‌ام.

در اواخر دهه ۱۹۹۰ مدلسازان به نظریات و روشهای نوین روی آوردند مثل استفاده از هوش مصنوعی، رشد سلولی (رشد شهر شبیه به رشد سلولی) (سلولهای خودکار) CA: Cellular Automata

مدل SELUTH: آقای Clark از دانشگاه واشنگتن

رشد شهر را بر اساس نظریه از علوم دیگر شبیه‌سازی می‌کند. این مدل در لیزبون و پورتو، ران شده است و نتیجه مطلوبی گرفته‌اند. براساس مدل و مقایسه واقعیت (با اطلاعات ماهواره‌ای). نسل جدید مدلها از تفکر لاری فاصله گرفتند.

مدل TOPAZ از مدلهای integrated جالب: Technique for Optimum Placement of Activites in Zons

که Brotchie و Sharpe استرالیایی آن را ساختند. اولین کاربردهای مدل در ملبورن بوده است و سپس بقیه شهرهای آن. از ۱۹۶۰ به بعد خود را به روز کرده و اکنونGIS Based است. می‌خواستند ببینند توسعه شهری ملبورن از نظر شکل شهر چگونه است. چند گزینه در نظر گرفته و تأثیر شکل شهر روی آلودگیهای زیست محیطی را بررسی کردند.

تأثیر شکل شهر روی سه بعد زیست محیطی
- استفاده از سوخت حمل و نقل
- گازهای گلخانه‌ای
- کیفیت هوا

پیش‌بینی: در سال ۲۰۱۱، نیم میلیون نفر به جمعیت ملبودن اضافه خواهد شد.

گزینه‌های مختلف شکل شهر روی مدل بررسی شده:

۱. ادامه وضع موجود (هیچ کار نکنیم چه می‌شود)
۲. شهر را فشرده کنیم.
۳. بخشهایی از شهر را به گوشه‌ها ببریم.
۴. شهر را حاشیه‌ای کنیم.
۵. سه راهرو برای شهر طراحی کنند و شکل شهر corridor شود.
۶. توسعه در بیرون شهر و شکل شهر دست نخورد.

این کار را با TOPAZ ۲۰۰۰ انجام دادند.

در دانشگاه نیومکزیکو از مدل جاذبه برای میزان دسترسی به تسهیلات بهداشتی داخل شهر استفاده کردند. تلفیق سه نرم‌افزار: SAS، Excel ، Arc view

John otensman از Excel و مدلهای جاذبه برای یافتن مکانهای مناسب برای خدمات خرده‌فروشی در شهر استفاده کردند.

مدل هنسن: ۱۹۵۹

یک مدل مکانی است که برای پیش‌بینی مکان جمعیت ساخته شده است. این روش برای تخصیص جمعیتی که در افق شهر قرار است به جمعیت آن اضافه شود به زونهای مختلف شهر به کار گرفته می شود.

فرضهای هنسن برای این تخصیص:

۱. دسترسی به اشتغال عامل اصلی در تعیین مکان جمعیتی است.

برای تعریف رابطه ایندو شاخصی را به نام Accessibility Index تعریف کرد.

A_{ij} : دسترسی به اشتغال زون i نسبت به زون j

E_j : اشغال در زون j

d_{ij}: فاصله زون i از زون j

b: پارامتری است که شرایط مدل را فراهم می‌کند و از طریق کالیبره کردن یا تنظیم مدل به دست می‌آید.

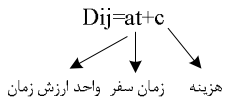
با فاصله از مرکز شهر تراکم کمتر می‌شود. هرچه فاصله بیشتر شود، دسترسی به اشتغال کمتر خواهد شد.

این روش می تواند برای منطقه، محله یا ناحیه باشد.

معمولاً Centroid یا مرکز ثقل زونها را در نظر گرفته و فاصله را از آن می‌سنجند. مرکز ثقل می‌تواند مرکز ثقل عملکردی یا مرکز ثقل هندسی باشد. ممکن است همه جا مرکز ثقل عملکردی نداشته باشیم ولی ساده‌ترین روش، درنظر گرفتن مرکز ثقل هندسی است.

فاصله ۱ Zone از ۲Zone :
- مسافت خیابانی (کوتاهترین مسیر)
- فاصله فضایی
- زمان طی مسافت + هزینه طی مسافت (ساعت پیک انتخاب نمی‌شود)

چون هدف ساده کردن است فاصله هوایی انتخاب می‌شود.



واحد ارزش زمان بر اساس متوسط درآمد افراد می‌باشد.

$d_{ii} = \frac{1}{4} \times$ (کوتاهترین فاصله منتهی به i)

برای یافتن b، اگر مطالعات مشابه انجام شده باشد می‌توان از آنها استفاده کرد. این پارامتر با گذشت زمان خیلی تغییر نمی‌کند. اگر نداشتیم باید در دو مقطع زمانی تمامی اطلاعات وجود داشته باشند. از داده‌های یک مقطع زمانی مقدار عددی پارامتر را به دست آورده و پیش‌بینی کرده و مقایسه می‌نماییم. با سعی و خطا فاصله را از هم کم کرده، مقادیر عددی پارامترها به دست می‌آید.

برای به دست آوردن دسترسی به اشتغال زون i باید دسترسی به اشتغال زون i با تمامی زونهای دیگر با هم جمع شوند.

دسترسی به اشتغال زون i

$$A_{11} = A_{12} + A_{13} + A_{14} \rightarrow \boxed{A_i = \sum_{j=1}^{n} A_{ij}}$$

هنسن عامل دیگری علاوه بر دسترسی به اشتغال را برای جمع‌آوری جمعیت لازم دانست:

میزان زمینهای خالی دارای کاربری مسکونی (چون جاذب جمعیت است) در هر Zone

H_i :Holding Capacity امکان رشد یا ظرفیت جذب

از ترکیب ایندو شاخصی به نام development potential به دست آورد.

پتانسیل توسعه: (باید یکسان باشد) هر واحد سطح، واحد آن است $D_i = A_i \times H_i \rightarrow$

هر زونی بر مبنای پتانسیل توسعه نسبی‌اش سهمی از مازاد جمعیت را به خود جذب می‌کنند.

G_i: سهم زون i از آن جمعیت G_t مازاد جمعیت در افق طرح

پتانسیل توسعه نسبی $G_i = G_t \times \frac{D_i}{\sum D_i}$

$$G_i = G_t \times \frac{A_i H_i}{\sum A_i H_i}$$ یا:

هنسن برای H_i، واحد آکر را در نظر گرفت.

Planning process:

شناخت، هدف‌گذاری، گزینه‌ها، انتخاب گرینه، Monitoring فیدبک می‌خورد. کاملاً پویا می‌باشد.

مثال عددی: در شهری که از ۳ زون تشکیل شده است، اطلاعات مربوط به جمعیت، کل اشتغال و امکان رشد (میزان زمینهای بایر دارای کاربری مسکونی) در هر یک از زونها و ماتریس فواصل بین آنها در جدول زیر ارائه شده است، طرح توسعه شهر که در دست بررسی است، افزایش جمعیت در افق طرح (۱۰ ساله) را ۳۲۰۰۰ نفر پیش‌بینی کرده است. با توجه به اطلاعات فوق جمعیت اضافه شده به مناطق سه‌گانه شهر را محاسبه کنید. فرض شود $b=2$

Zone	جمعیت	کل اشتغال	امکان رشد (هکتار)
1	42000	10000	220
2	53000	13000	150
3	65000	18000	120

ماتریس فاصله (d_{ij}):

i \ j	1	2	3
1	2	3	4
2	3	2	5
3	4	5	3

۱. به دست آوردن ماتریس A_{ij} $A_{ij} = \frac{E_j}{d_{ij}^{b=2}}$

ردیف i ⟶ ↓ j

Zonei \ j	1	2	3
1	$\frac{10000}{2^2}=2500$	$\frac{13000}{3^2}=1444.44$	$\frac{18000}{4^2}=1125$
2	$\frac{10000}{3^2}=1111.11$	$\frac{13000}{2^2}=3250$	$\frac{18000}{5^2}=720$
3	$\frac{10000}{4^2}=625$	$\frac{13000}{5^2}=520$	$\frac{18000}{3^2}=2000$

$$A_i = \sum_{j=1}^{n} A_{ij}$$

i	1	2	3
A_i	5069.44	5081.11	3145

H_i	$D_i = A_i * H_i$	$\frac{D_i}{\sum D_i}$ پتانسیل توسعه نسبی	$G_i = G_t * \frac{D_i}{\sum D_i}$
220	1115278	0.4946	$32000(0.4946) = 15828$
150	762166.7	0.3380	$32000(0.3380) = 10816$
120	377400	0.1674	$32000(0.1674) = 5356$
	$\sum D_i = 2254844$	1.0000 ↘ جمع باید 1 شود	32000

این جوابها تنها ابزاری برای نتیجه‌گیری است.

محدودیت‌های مدل هنسن:

۱. در تلفیق شاخصهای دسترسی و امکان رشد برای رسیدن به پتانسیل توسعه نسبی، ضریب اهمیت این دو شاخص یکسان در نظر گرفته شده است.

۲. پتانسیل توسعه فقط در دو شاخص دیده شده است.

۳. در تعیین H_i برای مناطق، تراکم مختلف مناطق شهری و سطح اشتغال یکسان در نظر گرفته شده است.

۴. قیمت و ارزش زمین در مناطق مختلف شهر یکسان در نظر گرفته شده است.

۵. کل جمعیت دارای خصوصیات یکسانی در نظر گرفته شده‌اند.

با اعمال دو محدودیت آخر مدل پیچیده می‌شود. ولی ما کمبودها را به انحاء دیگری در مدل لحاظ می کنیم.

در جاهائیکه تراکم جمعیت و فعالیت وجود دارد Zone ها را کوچک و در تفرق آنها Zoneها را بزرگ می‌گیریم.

کاربردهای مدل هنسن:

ـ مطالعه بر روی کلیولند در آمریکا برای بررسی دسترسی به اشتغال و الگوی جرم و جنایت. یکی از مدل‌های مورد استفاده در دسترسی به اشتغال آن مدل هنسن است.

Job accessibility index

ارتباط منفی بین دسترسی به اشتغال و نرخ جرم و جنایت خصوصاً جرائمی که ریشه اقتصادی دارند، وجود دارد. (دزدی، ...)

همین مفهوم با مدل دیگری دنبال شد ولی مدل هنسن نتیجه بهتری دربرداشته است.

ـ تخصیص ۱۴۰ هتل به استان‌های کشور: بر اساس پتانسیل جذب توریسم هر استان

T_i پتانسیل جذب توریسم در استان i

تعداد هتل در زون $H_i = H_t \times \frac{T_i}{\sum T_i}$

آثار تاریخی ثبت شده، جاذبه‌های طبیعی، دسترسی، ... : بیش از ۴۰ شاخص بررسی شد. استفاده از تحلیل عاملی برای ساده‌تر کردن

ـ استفاده از شاخص دسترسی مدل هنسن برای شعاع‌های عملکردی شهرهای مرکزی و ارتباط یکپارچه با شهرهای پیرامونی.

$$A_{ij} = E_j e^{-\lambda t}$$

مدل‌های جاذبه: *Gravity Model*

۱. مدل جاذبه محدودیت تک قیدی
   - مدل جاذبه تک قیدی محدودیت تولید سفر
   - مدل جاذبه تک قیدی محدودیت جذب سفر

۲. مدل جاذبه محدودیت دو قیدی: هر دو محدودیت یکجا اعمال می‌شوند (محدودیت در تولید یا در جذب).

از قانون جاذبه نیوتن در فیزیک گرفته شده و با تشبیه به مسائل شهری از آن در شهرسازی استفاده می‌شود. اولین کسانی که متوجه شدند از این مدل می‌توان بین زونهای مختلف یک شهر یا سکونتگاههای یک منطقه فضایی استفاده کرد: Kins Lee

قانون جاذبه خرده‌فروشی رایلی قبل از آنها: به کارگیری در مباحث توسعه فضایی برای تعیین ناحیه حوزه نفوذ یک خرده‌فروشی

از محدودیت‌های جدی مدل: عدم وجود پایه نظری برای کارهای شهرسازی

$$T_{ij} = \overbrace{k_i l_j}^{\text{در جاذبه } G_i=9.81} O_i D_j F(d_{ij})$$

$\downarrow$ در شهرسازی تعیین می‌شوند.

T_{ij}: تعداد سفرهایی که از زون i شروع و به زون j ختم می‌شوند.

k_i و l_j: ضرایب تعدیلی هستند که شرایط مدل را فراهم می‌آورند.

O_i: تعداد سفرهایی که از زون i شروع می‌شوند.

D_j: تعداد سفرهایی که به زون j ختم می‌شوند.

$F(d_{ij})$: تابع فاصله

ساده‌ترین $F(d_{ij}) = \frac{1}{d^{\alpha}_{ij}}$ ۱) در اغلب موارد

بیشترین مورد استفاده را به خاطر سهولت دارد.

$$d_{ij} \to 0 \quad \Rightarrow \quad F(d_{ij}) \to \infty$$

توابع فاصله دیگر:

$$2) F(d_{ij}) = \frac{1}{e^{\beta d_{ij}}}$$

پارامترها بیشتر و تنظیم مدل سخت‌تر:

$$3) F(d_{ij}) = \frac{d^{\alpha}_{ij}}{e^{\beta d_{ij}}}$$

زونهای شهری هر چقدر کوچک باشند فاصله آن‌ها صفر نیست پس $F(d_{ij})$ در فرمول اول به ∞ نمی‌رسد.

مدل سوم نیز تلفیقی از دو مدل اول و دوم است.

در مدل جاذبه $\alpha = 2$ می‌باشد. ولی در شهرسازی α از تنظیم یا کالیبره کردن مدل به دست می‌آید.

اگر مطالعات قبلاً وجود داشته باشد، پارامترها به دست می‌آیند چون α و β با تغییر زمان چندان تغییر نمی‌کنند.

محدودیت تولید و جذب سفر اصطلاحات عمومی برای تعریف نوع میان کنشی هستند که در آن تعداد سفرهایی که از یک زون حرکت می‌کنند و تعداد سفرهایی که جذب یک زون می‌شوند از قبل معلوم است. (مقصد و مبدأ سفرها از ابتدا مشخص است)

در مدل جاذبه دوقیدی هم مبدأ و هم مقصد سفرها از ابتدا مشخص است و این زون بیشتر از این نمی‌تواند تولید یا جذب سفر کند.

میزان سفرهایی که جذب آن می‌شوند نباید از این محدودیت بالاتر رود.

کاربردها:

مدل جاذبه تک قیدی محدودیت تولید سفر: توسعه مراکز خرید

مدل جاذبه تک قیدی محدودیت جذب سفر: تخصیص یا توسعه مسکونی Residential Allocation

مدل جاذبه دوقیدی: برای بررسی سفرهای کاری (از منزل به کار یا از کار به منزل)

هرگاه در بحث‌های شهرسازی تعامل بین تولید و جذب را بخواهیم بدانیم می‌توانیم از مدل جاذبه استفاده کنیم.

دسترسی به خدمات و تسهیلات بهداشتی با استفاده از مدل جاذبه

هر مسئله شهری که تولید، جذب و عامل فاصله دارد یا هر عامل با نقش بازدارنده بین این دو می تواند از مدل جاذبه استفاده کند.

دسترسی به اطلاعات نیز تقش تعیین کننده ای دارد.

مدل جاذبه تک قیدی محدودیت تولید سفر:

$$T_{ij} = k_i l_j O_i D_j F(d_{ij})$$

O_i باید به T_{ij} محدود شود. ولی محدودیت جذب نداریم.

$$O_i = \sum_{j=1}^{n} T_{ij}$$

$$F(d_{ij}) = \frac{1}{d_{ij}^{\alpha}}$$

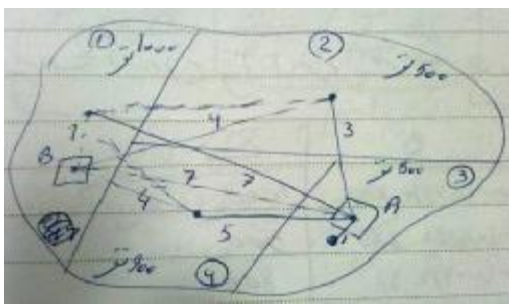
$$k_i = \frac{1}{\sum_{j=1}^{n} D_j F(d_{ij})} \qquad \ell_j = 1$$

$$\Rightarrow \quad T_{ij} = O_i \left[\frac{D_j}{d_{ij}^{\alpha}} \Big/ \sum_{j=1}^{n} \frac{D_j}{d_{ij}^{\alpha}} \right]$$

ماتریس احتمال Probability Matrix

ماتریس احتمال بین صفر تا یک متغیر است و نشان دهنده احتمال جذب به زون j است. سفر مبدأ و مقصد مشخص است.

مثال عددی: چهار ناحیه مسکونی فرضی که جمعیت هر یک از نواحی مشخص است را در نظر بگیرید که در یکی از آن‌ها مرکز خرید A با اندازه مشخص وجود دارد. (تعداد، مساحت زیربنای مغازه، میزان فروش، درآمد و گردش مالی ...) هر چه نشان دهنده داد و ستد باشد) که در حال حاضر به تمام چهار زون فوق سرویس می‌دهد. فرض کنید مرکز خرید جدیدی به نام B در زون 1 با اندازه‌ای دو برابر اندازه A قرار است ایجاد شود. چه تغییراتی در الگوی سفرهای خرید ممکن است اتفاق بیفتد؟



i \ j	A	B
1	7	1
2	3	4
3	1	7
4	5	4

ماتریس فاصله d_{ij}، $\alpha = 2$ در نظر بگیرید.

۱. به دست آوردن ماتریس $\frac{D_j}{d_{ij}^{\alpha}}$:

$$A = 100 \quad \rightarrow \quad B = 200$$

بهتر است از متغیرهای فیزیکی برای اندازه استفاده کنیم مثل تعداد مغازه یا مساحت.

i \ j	A	B	$\sum_{j=1}^{n} \frac{D_j}{d_{ij}^{\alpha}}$
1	$\frac{100}{7^2}=2.041$	$\frac{200}{1^2}=200$	202.041
2	$\frac{100}{3^2}=11.111$	$\frac{200}{4^2}=12.5$	23.61
3	$\frac{100}{1^2}=100$	$\frac{200}{7^2}=4.082$	104.08
4	$\frac{100}{5^2}=4$	$\frac{200}{4^2}=12.5$	16.5

j در ستون تغییر نمی‌کند: D_j : مرکز خرید A

۲. به دست آوردن ماتریس احتمال:

i \ j	A	B	$\sum$
1	$\frac{A_1}{\sum}=\frac{2.041}{202.04}=0.01$	$\frac{200}{202.041}=0.99$	1.00
2	$\frac{11.111}{23.61}=0.47$	$\frac{12.5}{23.61}=0.53$	1.00
3	$\frac{100}{104.08}=0.96$	$\frac{4.082}{104.08}=0.04$	1.00
4	$\frac{4}{16.5}=0.24$	$\frac{12.5}{16.5}=0.76$	1.00

احتمال اینکه از زون ۱ حرکت کرده و جذب مرکز خرید A شود بسیار کمتر از B است.

۳. به دست آوردن T_{ij} از طریق ضرب O_i در ماتریس احتمال:

i \ j	A	B	$\sum$
1	$1000(0.01)=10$	$1000(0.99)=990$	1000
2	$500(0.47)=235$	$500(0.53)=265$	500
3	$800(0.96)=768$	$800(0.04)=32$	800
4	$900(0.24)=216$	$900(0.76)=684$	900
$\sum$	1229	1971	3200

مراجعه به مرکز خرید A

می‌توان فقط جمعیت افراد در سنین سفر را بررسی کرد. به جای O_i می‌توان درآمد خانوار یا هزینه‌ها برای خرید را در نظر گرفت.

بخشی از درآمد که خرج می‌کنند. disposable income

هزینه مسکن جداست.

اطلاعات ما در ایران برای هزینه خانوارها استانی می‌باشد (شهری و روستایی در استان)

مدل جاذبه تک قیدی محدودیت جذب سفر:

هر زون فقط می‌تواند به اندازه ظرفیت جذب از T_{ij} جذب کند ولی تولید محدودیت ندارد.

هر زون فقط محدودیت جذب دارد.

$$\sum_{i=1}^{n} T_{ij} = D_j$$

$$F\left(d_{ij}\right) = \frac{1}{d_{ij}^{\alpha}}$$

$$k_i = 1 \qquad \ell_j = \frac{1}{\sum_{i=1}^{n} O_i F\left(d_{ij}\right)}$$

مدل محدودیت جذب سفر

$$\Rightarrow \quad T_{ij} = D_j \underbrace{\left[\frac{O_i}{d_{ij}^{\alpha}} \Big/ \sum_{i=1}^{n} \frac{O_i}{d_{ij}^{\alpha}} \right]}_{\text{ماتریس احتمال}}$$

جمع ستونی

عمدتاً برای تخصیص مسکونی به کار می‌رود.

مساحت کاربریهای مسکونی، تعداد واحدهای مسکونی و هر چه از نظر اسکان در آن زون جاذب جمعیت باشد یا کـاربری مسـکونی فـارغ از تراکم یا سطح اشغال که آسیب‌پذیرتر است.

۱. محاسبه $\frac{O_i}{d_{ij}^{\alpha}}$

۲. محاسبه ماتریس احتمال

۳. به دست آوردن T_{ij} از طریق ضرب ماتریس احتمال در D_j

مثال عددی: در اثر توسعه سه زون زیر در ۵ سال آینده پیش‌بینی می‌شود که تعداد شغلهای جدید در این زونها به شرح جدول زیـر باشـد. با استفاده از تجارب قبلی به نظر می‌رسد میان کنش فضایی بین این زونها با استفاده از مدل جاذبه مقدور باشد. $(\alpha = 1)$ با در نظر گـرفتن ماتریس فواصل اثرات این تعداد شغل جدید را برروی جمعیت زونهای سه‌گانه و توزیع آنها نشان دهید.

Zone	جمعیت	تعداد شغل جدید ایجاد شده
1	30000	200
2	20000	300
3	15000	400

i \ j	1	2	3
1	2	8	6
2	8	3	4
3	6	4	3

ماتریس فاصله d_{ij}

تعداد شغل‌های ایجاد شده مشخص است و شغل‌ها عامل جذب جمعیت هستند که محدودند. پس از محدودیت جذب استفاده می‌کنیم.

۱. محاسبه ماتریس $\frac{O_i}{d_{ij}^{\alpha}}$

i \ j	1	2	3
1	$\frac{30000}{2^1}=15000$	$\frac{30000}{8}=3750$	$\frac{30000}{6}=5000$
2	$\frac{20000}{8}=2500$	$\frac{20000}{3}=6666.67$	$\frac{20000}{4}=5000$
3	$\frac{15000}{6}=2500$	$\frac{15000}{4}=3750$	$\frac{15000}{3}=5000$
$\sum_{i=1}^{n}\frac{O_i}{d_{ij}^{\alpha}}$	20000	14166.67	15000

۲. محاسبه ماتریس احتمال (نسبت تولید را نشان می‌دهد) پتانسیل تولید زون‌ها را نشان می‌دهد.

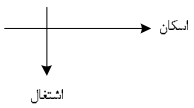
i \ j	1	2	3
1	$\frac{15000}{20000}=0.75$	$\frac{3750}{14166.67}=0.26$	$\frac{5000}{15000}=0.33$
2	$\frac{2500}{20000}=0.125$	$\frac{6666.67}{14166.67}=0.47$	$\frac{5000}{15000}=0.33$
3	$\frac{2500}{20000}=0.125$	$\frac{3750}{14166.67}=0.27$	$\frac{5000}{15000}=0.33$
$\sum$	1.0	1.0	1.0

→ یکی را 0.34 در نظر می‌گیریم

کدام زون پتانسیل تولید بیشتری برای گرفتن آن شغل‌ها دارد.

۳.

i \ j	1	2	3	$\sum$
1	$\boxed{0.75(200)=150}$	$0.26(300)=78$	$0.33(400)=132$	360
2	$0.125(200)=25$	$\boxed{0.47(300)=141}$	$0.33(400)=132$	298
3	$0.125(200)=25$	$0.27(300)=81$	$\boxed{0.34(400)=136}$	242
$\sum$	200	300	400	900



رابطه اسکان و اشتغال را نشان می‌دهد.

قطر ماتریس هم اسکان و هم اشتغال را نشان می‌دهد. مثلاً در زون ۱، ۱۵۰ نفر هم ساکنند و هم شاغل.

بعد خانوار ـ نسبت شاغلین در هر خانوار (بار تکفل) را باید بدانیم. برای پیش‌بینی جمعیت تعداد خانوار را نیز به دست می‌آوریم.

در مدل دقت بالای ۴۰٪ باشد کافیست. زیر ۴۰٪ رابطه علّی برقرار نمی‌شود.

مدل جاذبه محدودیت دوقیدی:

$$T_{ij}=k_i\ell_j O_i D_j F(d_{ij})$$

محدودیت جذب $\sum_{i=1}^{n} T_{ij} = D_j$ محدودیت تولید $\sum_{j=1}^{n} T_{ij} = O_i$

$$\ell_j = \frac{1}{\sum_{i=1}^{n} k_i o_i F(d_{ij})}$$

پیچیده‌ترین قسمت به دست آوردن ضرایب k_i و ℓ_j است

O_i و D_j را داریم چون هر دو محدودیت دارند.

$$k_i = \frac{1}{\sum_{j=1}^{n} \ell_j D_j F(d_{ij})}$$

ساده‌ترین شکل تابع فاصله $F(d_{ij}) = \frac{1}{d_{ij}^{\alpha}}$

این مدل واقع‌بینانه‌تر ولی پیچیده‌تر است پس کمتر از دوتای قبلی استفاده می‌شود.

مرحله ۱: به دست آوردن ماتریس $F(d_{ij})$

مرحله ۲: به دست آوردن k_i و ℓ_j از طریق تکرار (Iteration)

مرحله ۳: به دست آوردن T_{ij}

شروط: تعداد سفرهایی که از هر زون انجام می‌شود نباید بیش از جمعیت آن باشد و تعداد سفرهایی که جذب آن میشود نباید بیش از ظرفیت جذب آن باشد.

$$\Rightarrow \sum_{j=1}^{n} T_{ij} = O_i \qquad \sum_{i=1}^{n} T_{ij} = D_j$$

محدودیت‌های مدل:

۱. عمده‌ترین: عدم وجود پایه نظری منطقی. یعنی این مدل از پایه و اساس نظری منطقی درشهرسازی برخوردار نبوده و از علوم دیگر گرفته شده است.
۲. مدل، گروه‌های جمعیتی را به تفکیک در نظر نمی‌گیرد.
۳. قیمت انواع واحدهای مسکونی را بر حسب زونها در نظر نمی‌گیرد.
۴. تابع مسافتی که به کار می‌گیرد $(\frac{1}{d_{ij}})$ نارسا است.
۵. مشکلات زون‌بندی (چگونه زونها انتخاب می‌شوند) هرچه زون‌ها کوچکتر مشکل مدل کمتر
۶. تنظیم مدل سخت است (کالیبره کردن)

ولی با در نظر گرفتن همه اینها مدل بسیار پیچیده می‌شود.

مطالعاتی که با استفاده از مدل در سال‌های اخیر انجام شده:

۱. Thin و همکاران در ۲۰۰۲

هدف از مطالعه: بررسی میزان فشردگی شهرهای آلمان در ۱۱۶ شهر آلمان با استفاده از مدل جاذبه degree of compactness .

هر چه نواحی ساخته شده شهری پراکنده‌تر باشد فضاهای خالی بیشتری در آن موجود است. بنابراین میان کنش فضایی بین خوشه‌های شهر ضعیف‌تر است. هر چه میان کنش فضایی بین خوشه‌ها قویتر باشد (T بزرگتر) ساختار فضایی شهر فشرده‌تر است.

۲. در دانشگاه نیومکزیکو در آمریکا از مدل جاذبه برای اندازه‌گیری دسترسی جغرافیایی به امکانات و تسهیلات بهداشتی استفاده کردند. با سه نرم‌افزار SAS ،excel و Arc view

۳. ریچارد کلاسترمن (نویسنده نرم‌افزار Bodif) و آقای Xie (ژی) در سال ۱۹۹۷ از مدل جاذبه در دو محیط excel و Arc view استفاده کردند.

هدف: مطالعه الگوهای کنونی خریدهای خرده‌فروشی در منطقه کلان‌شهری اوهایو و پیش‌بینی اثرات ایجاد مراکز خرید جدید بر میزان فروش مراکز خرید موجود.

کارهای دیگری که انجام می‌دهد:

۱. تعیین مکان بهینه برای مراکز جدید
۲. اندازه بهینه مراکز
۳. برآورد اثرات افزایش یا تعطیل مرکز جدید
۴. تعیین حوزه نفوذ مراکز خرید موجود
۵. بررسی ویژگیهای بازار مراکز خرید موجود

مدل گرین ـ لاری:

بر مبنای تئوری اقتصاد پایه عمل می‌کند. کل فعالیت‌های اقتصادی را به دو نوع پایه و غیرپایه تقسیم می‌کنیم.

فرض: بخش پایه عامل رشد و توسعه است و بخش غیرپایه به دنبال بخش پایه حرکت می‌کند. یعنی موتور توسعه بخش پایه است.

برمبنای اقتصاد پایه یک مدل عمومی از شهر می‌سازد. مدل‌های جاذبه و هنسن مدل‌های میان کنش فضایی بودند و تلاش می‌کردند اتفاقات درون شهری را تبیین کنند.

ولی مدل لاری یک مدل عمومی توسعه شهری است: General Urban Model

پس در کل بحث توسعه شهری را در نظر می‌گیرد.

نظریه اقتصاد پایه را به دو نفر نسبت می‌دهند: داگلاس کورن، هومر هورت

اولین کار: شناسایی بخش پایه و غیرپایه اقتصادی:

مراحل بکارگیری مدل لاری:

۱. ابتدا شاغلین پایه را به مناطق مسکونی تخصیص می‌دهیم (با استفاده از مدل جاذبه تک‌قیدی محدودیت تولید سفر)
۲. جمعیت وابسته به شاغلین پایه را با استفاده از ضریب وابستگی جمعیت (α) به دست می‌آوریم. (بار تکفل)
۳. شاغلین خدماتی موردنیاز جمعیت پایه را براساس ضریب β یعنی نسبت شاغلین خدماتی به کل جمعیت به دست می‌آوریم.
۴. شاغلین خدماتی موردنیاز جمعیت پایه را با استفاده از مدل جاذبه تک‌قیدی محدودیت جذب سفر به مناطق تخصیص می‌دهیم.
۵. کل شاغلین خدماتی هر زون را محاسبه می‌کنیم.
۶. در تکرار بعدی از شاغلین خدماتی ۵ step استفاده می‌کنیم و تکرار دوم بر مبنای این تعداد شاغلین انجام شده و محاسبات ادامه پیدا می‌کند تا اینکه شاغلین خدماتی آخرین تکرار به عدد ناچیزی برسد.

Lee از جمعیت به عنوان اطلاعات جایگزین D_j استفاده کرد.

مزایا: اطلاعات جمعیتی به راحتی در اختیار است.

معایب: وقتی تعداد زونها زیاد باشد (مثل مناطق ۲۲ گانه تهران)، سرانه واحدهای مسکونی در منطق lee برای همه مناطق یکسان است، جمعیت خالی از اشکال نیست چون در این شهرها تفاوت جدی بین سرانه واحدهای مسکونی در این مناطق وجود دارد. پس استفاده از این تیپ داده ها مشکل جدی ایجاد می کند.

محدودیتهای مدل:

۱. به دلیل اینکه مدل از مدلهای جاذبه برای تخصیص استفاده می کند، بنابراین تمامی محدودیتهای مدل جاذبه شامل این مدل هم می شود.

۲. چون مدل گرین- لاری از نظریه اقتصاد پایه استفاده می کند، بنابراین تمامی محدودیتهای این نظریه را نیز دارا است. (در نظریه اقتصاد پایه هر بخش فقط در درون خود مورد استفاده قرار می گیرد و وابستگیهای بین بخشی دیده نمی شود)

۳. وقتی مقادیر عددی پارامترهای α و β و E و b برای آینده نگری استفاده می شوند، فرض بر این است که شرایط و روند موجود در آینده نیز ادامه خواهد داشت (البته ضرایب خیلی تغییر نمی کنند).

۴. صرفه جوییهای مقیاس، چون رشد زونها خطی در نظر گرفته می شود، در رشد مراکز خدماتی نادیده گرفته می شوند (مثلاً مراکز خرید بزرگتر ممکن است بالنسبه جمعیت بیشتری را جذب کنند).

۵. تعیین شاغلین پایه در خارج از مدل صورت می گیرد.

نقاط قوت مدل:

داده‌های مورد نیاز مدل در سطح قابل قبولی هستند. از نظر پیچیدگی و همچنین هزینه‌ها مدل قابل مدیریت و انجام است.

انعطاف بیشتری دارد و بسیاری از محدودیت‌های مدل قابل حل و رفع است.

بیشترین کاربرد در زیر مدل‌های لاری: MEPLAN

مدل‌های خطی:

تحلیل رگراسیونی: Regression Analysis

هرگاه بخواهیم بین دو یا چند متغیر رابطه کمی برقرار کنیم یکی از مکانیزم‌ها تحلیل رگراسیونی است.

رگراسیون ساده:

۱. متغیر وابسته (dependent variable)

۲. متغیر مستقل (independent variable)

رگراسیون چندمتغیره ($2 <$):

یک متغیر وابسته

چند متغیر مستقل

وقتی داده‌ها خطی نباشند، اکثراً می‌توان آن‌ها را تبدیل به خطی کرد.

$y_c = \alpha + \beta X$

y_c : متغیر وابسته

α : عرض از مبدأ

β : شیب خط

X : متغیر مستقل

اول باید بدانیم ماهیت داده ها چیست؟ (سرشماری یا نمونه گیری)

دوم: کدام متغیر وابسته و کدام مستقل است؟

اگر داده‌ها منتج از سرشماری باشد رابطه خطی همان است که ذکر شد. ولی اگر منتج از نمونه‌گیری باشند (اغلب موارد) رابطه به صورت زیر است:

$y = a + bX + \varepsilon$

a : تخمینی از α

b : تخمینی از β

ε : خطا (که معمولاً در نظر گرفته نمی‌شود.)

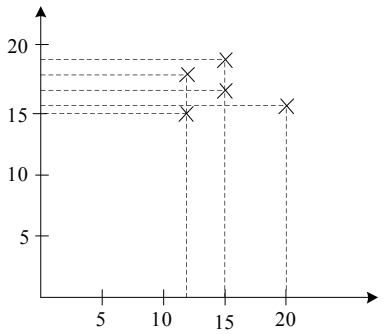
تعیین اینکه کدام متغیر وابسته است و کدام مستقل به اهداف مطالعه بستگی دارد.

مثلاً اگر بخواهیم معدل دانشجویان رشته شهرسازی را از تعداد ساعاتی که آن‌ها صرف مطالعه می‌کنند حدس بزنیم معدل وابسته به تعداد ساعات مطالعه است.

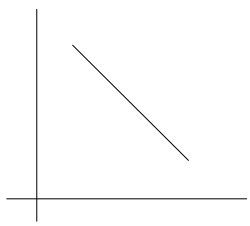
مراحل تحلیل رگرسیونی ساده خطی:

۱. ترسیم دیاگرام پراکندگی یا پراکنش (Scatter plot/scatter diagram)

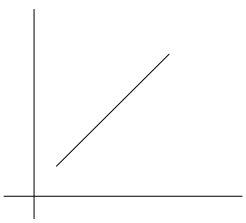
ساعات	معدل
12	18
15	19
15	17
20	16
12	15



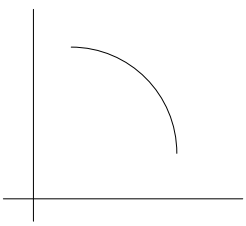
می‌خواهیم بدانیم آیا بین X و y اصلاً رابطه‌ای وجود دارد؟ اگر وجود دارد خطی است یا غیر خطی؟ و اگر خطی است مستقیم است یا معکوس؟ (+ یا −)



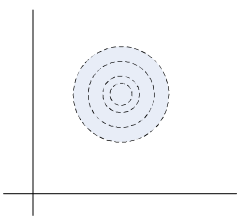
رابطه خطی معکوس



رابطه خطی



رابطه غیرخطی که می‌تواند تبدیل به خطی شود



رابطه وجود ندارد

رابطه مستقیم یا + : با افزایش یکی، دیگری هم افزایش پیدا می‌کند.

۲. به دست آوردن معادله رگرسیونی

$y_c = a + bX$

$$a = \frac{\sum y}{n} - b\left(\frac{\sum X}{n}\right)$$

$$b=\frac{n\sum XY-\left(\sum X\right)\left(\sum y\right)}{n\left(\sum X^2\right)-\left(\sum X\right)^2}$$

برای به دست آوردن a و b چند روش وجود دارد:

حداقل مربعات Least square Method

که روابط بالا از آن به دست آمده‌اند.

صرف به دست آمدن این رابطه معنی آن نیست که از نظر آماری رابطه معنی‌داری بین آنهاست و مراحل دیگری باید طی شود.

۳. به دست آوردن خطای معیار تخمین: Standard Error of Estimate (SEE)

SEE: به طور متوسط مشاهدات ما چقدر از خط رگرسیون که به دست آوردیم فاصله دارند. هر چه خطا کمتر باشد معادله بهتر است مشروط بر اینکه معیارهای دیگر برابر باشد.

میانگین $SEE = y - y_c$

منفی‌ها مثبت‌ها را خنثی نکند

$$Sy_x=\sqrt{\frac{\sum\left(y-y_c\right)^2}{n-2}}$$

یا

$$Sy_x=\sqrt{\frac{\sum y^2-a\sum y-b\sum XY}{n-2}}$$

در تحلیل رگرسیونی:

n : تعداد داده‌ها

m : تعداد متغیرهای مستقل

$df = n - m - 1$

استفاده از خطای معیار تخمین:

۱. آزمون فرضیه راجع به شیب خط (رابطه معنی دار است یا نه)

۲. انتخاب مدل بهتر بین دو مدل

۴. آزمون فرضیه در مورد β (شیب خط)

اگر $\beta = 0$ آنگاه خط موازی x است، یعنی هیچگاه آن را قطع نمی‌کند. پس هیچگاه رابطه‌ای بین x و y وجود ندارد. پس وقتی اطمینان داریم بین x و y رابطه وجود دارد که شیب خط صفر نباشد.

1) $\begin{cases} H_0 : \beta = 0 \\ H_a : \beta \neq 0 \end{cases}$ آزمون دو دنباله‌ای

2) $\begin{cases} H_0 : \beta = 0 \\ H_a : \beta > 0 \end{cases}$ اگر رابطه + باشد

3) $\begin{cases} H_0 : \beta = 0 \\ H_a : \beta < 0 \end{cases}$ اگر رابطه معکوس باشد

دقت و احتمال رد فرضیه 0 در اولی بیشتر است به اندازه 0.25 بنابراین همیشه به سراغ آن نمی‌رویم.

اگر مشاهده نزدیک به 3 است به سراغ ۲ یا ۳ برویم بهتر است.

۲. تعیین سطح اطمینان

۳. انتخاب آزمون آماری مناسب
- t : دو متغیر داریم
- f : بیش از دو متغیر داریم.

مقدار عددی β در فرضیه صفر که معمولاً صفر در نظر گرفته می‌شود.

$$t_c = \frac{b - \beta H_0}{S_b}$$

S_b: خطای معیار b

S_{yx}: خطای معیار تخمین

$$S_b = \frac{S_{yx}}{\sqrt{\sum X^2 - \frac{\left(\sum X\right)^2}{n}}}$$

یک t نیز از جدول استخراج می‌کنیم.

$$df = n - m - 1$$

۵. محاسبه ضریب همبستگی:

Coefficient of correlation

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right]\left[n\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

ضریب تعیین R^2

Coefficient of determination

درصد تغییرات متغیر وابسته که توسط متغیر مستقل انتخاب شده در مدل توضیح داده می‌شود. وقتی مفهوم پیدا می‌کند که رابطه معنی‌دار باشد. R^2 قدرت و دقت مدل را نشان می‌دهد. هر چه بزرگتر باشد، مدل مطلوب‌تر است مشروط بر اینکه رابطه معنی‌دار باشد.

برای اکثر پدیده‌ها خصوصاً در شهرسازی به بیش از یک متغیر نیاز داریم.

تحلیل رگرسیونی چند متغیره:

۱. ترسیم دیاگرام پراکندگی یا پراکنش: تک‌تک متغیرهای مستقل را با متغیر وابسته دو به دو در دیاگرام ترسیم می کنیم و سؤالات را نیز برای هر کدام مطرح می‌کنیم: تشخیص چشمی: آیا ارتباط هست یا نه؟ خطی یا غیرخطی؟ مثبت یا منفی؟

مثلاً برای ۴ متغیر مستقل ۴ دیاگرام پراکندگی ترسیم می‌کنیم. محاسبات و تفسیر در ارتباطات خطی بسیار ساده‌تر است.

۲. به دست آوردن معادله رگرسیونی

$$y_c = a + b_1X_1 + b_2X_2 + \ldots + b_mX_m$$

باز هم با روش حداقل مربعات a و b_i ها را به دست می‌آوریم.

$$\sum X_1 y = b_1 \sum X_1^2 + b_2 \sum X_1X_2 \qquad (1)$$

$$\sum X_2 y = b_1 \sum X_1X_2 + b_2 \sum X_2^2 \qquad (2)$$

دو معادله دو مجهول: b_1 و b_2 و سپس a را به دست می‌آوریم.

Y	X_1	X_2	Y^2	X_1^2	X_2^2	YX_1	YX_2	X_1X_2
$\sum Y$	$\sum X_1$	$\sum X_2$	$\sum Y^2$	$\sum X_1^2$	$\sum X_2^2$	$\sum YX_1$	$\sum YX_2$	$\sum X_1X_2$

$$\sum X_1 y = \sum X_1 Y - n\overline{X}_1\overline{Y}$$

$$\sum y^2 = \sum Y^2 - n\overline{Y}^2$$

$$\sum x_1^2 = \sum X_1^2 - nX_1^2$$

$$\sum x_2 y = \sum X_2 Y - n\overline{X}_2\overline{Y}$$

$$\sum x_1 x_2 = \sum X_1 X_2 - n\overline{X}_1\overline{X}_2$$

$$\sum x_2^2 = \sum X_2^2 - n\overline{X}_2^2$$

$$a = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2$$

۳. به دست آوردن خطای معیار تخمین

۴. آزمون فرضیه در مورد β

می‌توان کماکان با آزمون t انجام داد برای تک تک متغیرهای مستقل یا یکبار برای کل مدل آزمون f را انجام داد.

اگر رابطه رگرسیونی معنی‌دار باشد خیلی خوب است و با یک آزمون می‌توان فهمید. اگر معنی‌دار نباشد نمی‌دانیم که علت عدم ارتباط کدام متغیر مستقل است و باید تک تک برای هر کدام آزمون t انجام دهیم. هر متغیر که رابطه معنی‌دار نداشت آن را از معادله بیرون می‌آوریم و تمام مراحل را از اول با سه متغیر انجام می‌دهیم نه اینکه فقط آن را خارج می‌کنیم.

محاسبه R^2:

$$R^2 = \frac{b_1\sum X_1 y + b_2\sum X_2 y}{\sum y^2}$$

درصد تغییرات متغیر وابسته که توسط متغیرهای مستقل انتخاب شده در مدل توضیح داده می‌شود.

m : تعداد متغیرهای مستقل انتخاب شده در مدل $df = n - m - 1$

در تحلیل رگرسیونی چندمتغیره مواظب دو پدیده باشیم:

۱. پدیده هم‌خطی Multi-Collinearity

وقتی در یک تحلیل رگراسیونی چند متغیره با چند متغیر مستقل سر وکار داشته باشیم اگر بین متغیرهای مستقل همبستگی شدیدی وجود داشته باشد، ($|r| \geq 0.6$) بنابراین این متغیرهای مستقل اطلاعات همانندی را در اختیار متغیر وابسته قرار می‌دهند. پس یکی را انتخاب کرده و بقیه را حذف می‌کنیم. هر متغیر مستقلی که همبستگی بیشتری با y داشته باشد مطلوب ماست. پس یکی از راههای انتخاب، میزان همبستگی متغیرها با متغیر وابسته است. دیگری موضوع مورد بررسی است.

مثلاً درصد شهرنشینی با همه چیز همبستگی بالا دارد. بنابراین اگر موضوع خاصتری داشتیم مثلاً جذب جمعیت در مکانهای صنعتی، موضوع عام را کنار می‌گذاریم.

برای تشخیص پدیده هم خطی از ماتریس همبستگی (Correlation Matrix) استفاده می‌کنیم. در این ماتریس همبستگی بین مستقل‌ها و وابسته‌ها مشخص است. به متغیرهای مستقل نگاه کرده، هر جا کمتر باشد independent است.

	y	x_1	x_2	x_3	x_4
y	1.00	0.84	0.64		0.70
x_1		1.00	0.64		0.15
x_2			1.00	−0.71	0.17
x_3				1.00	−0.19
x_4					1.00

x_1 با x_2 همبستگی شدید دارند و x_2 با x_3

پس x_1 با x_2 همزمان نمی‌توانند در یک رابطه باشند

و x_2 با x_3 همزمان نمی‌توانند در یک رابطه باشند

اگر با توجه به هدف هم بتوان تشخیص داد می‌توان هر بار یکی را استفاده کرد.

همواره باید گزینه‌ای عمل کنیم.

۲. خودهمبستگی (یا همبستگی رشته‌ای یا همبستگی سلسله‌ای) Auto- Correlation

عمدتاً وقتی اتفاق می‌افتد که داده‌ها متعلق به سریهای زمانی است. یعنی در بعضی موارد متغیر مستقل انتخاب شده، در مدل خودش را مهمتر از آنچه که در واقعیت است نشان می‌دهد.

نحوه تشخیص: آزمون دوربین ـ واتسون Durbin – Watson

محاسبات این آزمون روی error terms است.

$$e = y - y_c$$

$$d_c = D.W. = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

1) if $d_c < d_{L,\alpha/2}$ $\Rightarrow$ H_0 رد

2) if $d_c > d_{u,\alpha/2}$ $\Rightarrow$ H_0 رد نمی‌شود

3) if $d_{L,\alpha/2} \le d_c \le d_{u,\alpha/2}$ $\Rightarrow$ آزمون بی‌نتیجه است

در این صورت می‌توان α را تغییر داد. ولی در عمل تعداد داده‌ها را بیشتر می‌کنند. اگر دقت خیلی مهم باشد، متغیرهای مستقل را جابجا می‌کنیم.

قاعده سرانگشتی: دوربین واتسون بین ۴-۰ است. پس هر چه به مرکز نزدیکتر باشد نگرانی کمتر است. یعنی بین $2.5 - 1.5$. در غیر اینصورت به جدول نیاز داریم:

$$Max(t) = 3$$

مثلاً

$$y_c = 3.14 + 1.6x_1 + \cancel{1.4x_2} + 2.6x_3$$

$$3.46 \quad -1.2 \quad 5.96$$

$\downarrow$

حذف: با مدل دیگری معادله را به دست می‌آوریم.

بدون چارچوب نظری نمی‌توان گفت با هم ارتباط کمی معنی‌داری دارند. مثلاً تعداد دندانهای اسب با دندانه‌های یک شانه می‌تـوان ارتبـاط کمی با درصد اطمینان بالایی داشته باشد ولی هیچ مفهومی ندارند.

می‌خواهیم مدلی را بین x_1, x_2, x_3, x_4 و y برقرار کنیم.

$$y_c = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

y می‌تواند با هر کدام از متغیرهای مستقل بررسی شود که بسیار طولانی است، پس ماتریس همبستگی را محاسبه می‌کنیم.

از ماتریس همبستگی برای رسیدن به محتملترین معادله استفاده می‌کنیم.

۱. ارتباط متغیرهای مستقل با متغیر وابسته (آنهایی که بیشترین ارتباط را با y دارند.)

هر متغیر که همبستگی بیشتری با y دارد بهتر می‌تواند پدیده را تبیین کند (قدر مطلق) بالای ۴۰٪ رابطه علّی وجود دارد.

اگر ارتباط زیر 0.4 باشد ارتباط علّی وجود ندارد.

$$R \geq |0.4|$$

۲. بین متغیرهای مستقل باید همبستگی شدید وجود نداشته باشد.

$$R \geq |0.6|$$

برای انتخاب بهترین رابطه ۵ موضوع باید چک شوند:

۱. آزمون β برای معنی‌دار بودن رابطه رگراسیونی
۲. محاسبه مقدار R^2 (ضریب تعیین)
۳. مناسب بودن مقدار عددی و علامت b ها
۴. مناسب بودن مقدار عددی و علامت a
۵. مقایسه خطای معیار تخمین

(اگر همه پارامترها برابر باشند مدلی بهتر است که خطای معیار تخمینش کمتر باشد)