

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

آمار و احتمال برای مهندسی مدرن

جلد دوم

ترجمه:

لارنس ال. لاپین

تألیف:

مهندس مهدی تیموری یانسری

ویرایش فنی:

دکتر محسن رضائیان

عضو هیأت علمی دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

زمستان ۱۳۹۵

سرشناسه

لایپین، لارنس ال

Lapin, Lawrence L.

عنوان و نام پدیدآور

آمار و احتمال برای مهندسی مدرن / تألیف لارنس ال. لاپین؛ ترجمه مهدی

تیموری یانسنری؛ ویرایش فنی محسن رضائیان

تهران: دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)، ۱۳۸۶

مشخصات نشر

ج: ج: مصور

مشخصات ظاهری

دوره: 4-329-463-978

شابک

ج ۲: 7-328-463-978

وضعیت فهرست‌نویسی

فیبا

عنوان اصلی: Probability and Statistics for Modern Engineering

یادداشت

مهندسی - - روش‌های آماری

موضوع

احتمالات

موضوع

تیموری نیاسری، مهدی، ۱۳۵۷، مترجم

شناسه افزوده

دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

شناسه افزوده

TA۳۴۰/J۲۱۸ ۱۳۸۶:

رده‌بندی کنگره

۵۱۹/۲۰۲۴۶۲:

رده‌بندی دیویی

۱۰۸۹۹۹۴:

شماره کتابشناسی ملی

این کتاب در جلسه مورخ ۱۳۸۵/۲/۱۲ شورای چاپ و نشر دانشگاه صنعتی امیرکبیر به تصویب رسیده است.



انتشارات دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

عنوان کتاب

آمار و احتمال برای مهندسی مدرن (جلد دوم)

تألیف

لارنس ال. لاپین

ترجمه

مهندس مهدی تیموری یانسنری

ویرایش فنی

دکتر محسن رضائیان

ناشر

انتشارات دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

لیتوگرافی، چاپ و صحافی

انتشارات دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

چاپ دوم

زمستان ۱۳۹۵

تیراژ

۲۰۰ نسخه

قیمت دوره دوجلدی

۴۰۰۰۰ تومان

شابک

ISBN : 978-964-463-328-7

شابک

۹۷۸-۹۶۴-۴۶۳-۳۲۸-۷:

آدرس مرکز پخش: خیابان ولیعصر، روبروی خیابان بزرگمهر، فروشگاه کتاب مرکز نشر

دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران) - تلفن: ۶۶۴۹۸۸۶۸

وبسایت: <http://publication.aut.ac.ir>

حق چاپ برای ناشر محفوظ است.

به نام خداوند بخشنده و مهربان

تقدیم به

پدرم و مادرم

که برایم زحمات فراوانی کشیده‌اند. همچنین تقدیم به

سرکار خانم دکتر ماه‌بانو تاتا

استاد ماشاءالله ماشین چی

جناب دکتر سعید رضاخواه

جناب دکتر ناصر بروجردیان

اساتید نیک‌اندیش و عالی‌مقام علم آمار و ریاضی که درس اخلاق، آمار و ریاضی
را از آنان آموخته‌ام.

فهرست مطالب

فصل نهم	رگرسیون و تحلیل همبستگی	۵۲۳
۱-۹	رگرسیون خطی با روش مبتنی بر حداقل مربعات	۵۲۵
۱-۱-۹	روش حداقل مربعات	۵۲۶
۲-۱-۹	منطق و مفهوم روش حداقل مربعات	۵۲۹
۳-۱-۹	اندازه‌گیری تغییر پذیری نتایج	۵۳۱
۴-۱-۹	تحلیل رگرسیون به کمک رایانه	۵۳۳
۲-۹	فرضیه‌ها و خواص تحلیل رگرسیون خطی	۵۳۴
۱-۲-۹	مفروضات و خواص تحلیل رگرسیون خطی	۵۳۵
۲-۲-۹	برآورد معادله رگرسیون حقیقی	۵۳۷
۳-۹	استنباط‌های آماری به کمک خط رگرسیون	۵۳۸
۱-۳-۹	پیش‌بینی و بازه‌های اطمینان در تحلیل رگرسیون	۵۳۸
۲-۳-۹	بازه اطمینان برای میانگین شرطی	۵۳۸
۳-۳-۹	دامنه پیش‌بینی برای یک تک مشاهده Y ، به ازاء X دلخواه	۵۴۲
۴-۳-۹	خطرات برون‌یابی	۵۴۳
۵-۳-۹	استنباط راجع به ضرایب رگرسیون	۵۴۳
۴-۹	تحلیل همبستگی	۵۴۶
۱-۴-۹	توزیع نرمال دو متغیره	۵۴۷
۲-۴-۹	ضریب همبستگی نمونه‌ای	۵۴۹
۳-۴-۹	استنباط راجع به ضریب همبستگی	۵۵۲
۵-۹	ارزیابی کیفیت رگرسیون	۵۵۳
۱-۵-۹	ضریب تعیین	۵۵۳
۲-۵-۹	نیکوئی مدل: تحلیل باقی مانده‌ها	۵۵۸
۶-۹	رگرسیون غیرخطی (منحنی الخط)	۵۶۲
۱-۶-۹	تبدیل‌های لگاریتمی	۵۶۲

۵۶۶	تبدیل‌های معکوس	۲-۶-۹
۵۶۷	تأثیر تبدیل به مدل خطی	۳-۶-۹
۵۶۸	مسائل	
۵۸۳	بخش سوم	
۵۸۵	فصل دهم	
	تحلیل آماری II: ارزیابی با چند متغیر	
	رگرسیون چندگانه و همبستگی	
۵۸۶	رگرسیون خطی چندگانه با سه متغیر	۱-۱۰
۵۸۶	رگرسیون سه بُعدی	۱-۱-۱۰
۵۸۹	مفروضات رگرسیون چندگانه	۲-۱-۱۰
۵۸۹	رگرسیون چندگانه با روش حداقل مربعات	۲-۱۰
۵۹۳	مزایای رگرسیون چندگانه	۱-۲-۱۰
۵۹۵	باقی مانده‌ها و خطای استاندارد برآورد	۲-۲-۱۰
۵۹۸	مشکلات تحلیل رگرسیون چندگانه	۳-۲-۱۰
۵۹۹	رگرسیون با متغیرهای زیاد: کاربرد رایانه	۳-۱۰
۶۰۰	کاربرد رایانه در تحلیل رگرسیون چندگانه	۱-۳-۱۰
۶۰۶	رگرسیون چندگانه گام به گام (مرحله‌ای)	۲-۳-۱۰
۶۰۸	مدل رگرسیون چندگانه تعمیم یافته	۴-۱۰
۶۰۸	نمایش ماتریسی رگرسیون چندگانه	۱-۴-۱۰
۶۱۱	استنباط‌هایی راجع به تحلیل رگرسیون چندگانه	۵-۱۰
۶۱۲	بازه اطمینان و دامنه پیش‌بینی	۱-۵-۱۰
۶۱۵	استنباط‌هایی راجع به ضرایب رگرسیون	۲-۵-۱۰
۶۲۰	همبستگی چندگانه	۶-۱۰
۶۲۰	ضریب تعیین در حالت چندگانه	۱-۶-۱۰
۶۲۳	همبستگی جزئی	۲-۶-۱۰
۶۲۷	رگرسیون چند جمله‌ای	۷-۱۰
۶۲۷	رگرسیون سهموی	۱-۷-۱۰

۶۲۹	رگرسیون چند جمله‌ای با درجه بالاتر	۲-۷-۱۰
۶۳۱	رگرسیون چندگانه چند جمله‌ای	۳-۷-۱۰
۶۳۱	رگرسیون چندگانه با متغیرهای نشانگر	۸-۱۰
۶۳۲	مدل رگرسیون چندگانه پایه	۱-۸-۱۰
۶۳۵	مزایای استفاده از متغیرهای نشانگر	۲-۸-۱۰
۶۳۶	رگرسیون چندگانه تعاملی با متغیرهای نشانگر متقابل	۳-۸-۱۰
۶۴۰	مسائل	

۶۶۳	تحلیل واریانس	فصل یازدهم
۶۶۴	چارچوب تحلیل تک عاملی	۱-۱۱
۶۶۵	آزمون برابری میانگین‌ها	۱-۱-۱۱
۶۶۸	مدل تک عاملی	۲-۱-۱۱
۶۷۰	میانگین‌های نمونه‌ای	۳-۱-۱۱
۶۷۱	تحلیل واریانس تک عاملی	۲-۱۱
۶۷۱	انحراف از میانگین‌های نمونه‌ای	۱-۲-۱۱
۶۷۳	خلاصه سازی انحراف موجود در مشاهده‌های نمونه‌ای	۲-۲-۱۱
۶۷۶	مبنایی برای مقایسه: میانگین مربعات	۳-۲-۱۱
۶۷۷	جدول ANOVA و آماره F	۴-۲-۱۱
۶۷۹	مراحل آزمون فرضیه	۵-۲-۱۱
۶۸۰	تحلیل به کمک رایانه	۶-۲-۱۱
۶۸۰	تصمیم‌گیری: مقایسه میانگین راهکارها	۷-۲-۱۱
۶۸۱	روش‌های آزمون با قابلیت تشخیص بیشتر	۸-۲-۱۱
۶۸۳	خطای نوع II	۹-۲-۱۱
۶۸۴	تخطی‌های مدل زیربنایی	۱۰-۲-۱۱
۶۸۴	تحلیل مقایسه‌ای راهکارها	۳-۱۱
۶۸۵	استنباط منفرد راجع به یک میانگین یا یک تفاوت دوتایی	۱-۳-۱۱

۶۸۸	مقایسه‌های چندگانه و استنباط‌های مجموعه‌ای	۲-۳-۱۱
۶۹۰	مقایسه‌های چندگانه از میانگین‌های جامعه‌ی راهکارها	۳-۳-۱۱
۶۹۳	ارزیابی‌ها به کمک رایانه	۴-۳-۱۱
۶۹۶	برآورد میانگین منفرد برای هر جامعه‌ی راهکار	۵-۳-۱۱
۶۹۷	ترکیب دو خانواده از برآوردهای مجموعه‌ای	۶-۳-۱۱
۶۹۸	طرح آزمایشی	۴-۱۱
۷۰۱	روش غیر پارامتری: آزمون کروسکال-والیس	۵-۱۱
۷۰۴	مسائل	

۷۱۵	تحلیل واریانس چندعاملی	فصل دوازدهم
۷۱۶	تحلیل واریانس دو عاملی	۱-۱۲
۷۱۶	جامعه‌ها و میانگین‌ها	۱-۱-۱۲
۷۱۸	مدل زیربنایی	۲-۱-۱۲
۷۲۱	فرضیه‌ی صفر	۳-۱-۱۲
۷۲۳	مشاهده‌های نمونه‌ای در آزمایش‌های دو عاملی	۴-۱-۱۲
۷۲۵	چارچوب تحلیلی	۵-۱-۱۲
۷۲۶	جدول ANOVA دو عاملی	۶-۱-۱۲
۷۲۹	چه باید کرد؟	۷-۱-۱۲
۷۳۰	تحلیل به کمک رایانه	۸-۱-۱۲
۷۳۲	حالت خاص-یک مشاهده در هر سلول	۹-۱-۱۲
۷۳۵	تحلیل اثرات عاملی	۲-۱۲
۷۳۶	حالت بدون تداخل اثر: مقایسه‌های دوتایی چندگانه برای میانگین‌های سطوح عامل	۱-۲-۱۲
۷۴۰	حالت تداخل اثر عوامل یا تأثیر متقابل معنی‌دار: تقابل‌های چندگانه بین میانگین‌های راهکارها	۲-۲-۱۲
۷۴۶	برآورد پارامترهای اثرات	۳-۲-۱۲

۷۴۶	طرح بلوکی تصادفی	۳-۱۲
۷۴۸	چارچوب تحلیلی و مدل نظری	۱-۳-۱۲
۷۵۲	تحلیل به کمک رایانه	۲-۳-۱۲
۷۵۳	مقایسه های چندگانه میانگین های سطوح عوامل	۳-۳-۱۲
۷۵۵	طرح بلوکی تصادفی مربع لاتین	۴-۱۲
۷۵۹	مسائل	

۷۷۵ فصل سیزدهم

آزمون استقلال، تصادفی بودن و نیکوئی برازش

۷۷۷	آزمون چی - اسکوئر برای استقلال	۱-۱۳
۷۷۷	استقلال بین متغیرهای کیفی	۱-۱-۱۳
۷۷۹	آزمون استقلال	۲-۱-۱۳
۷۸۳	مراحل آزمون فرضیه	۳-۱-۱۳
۷۸۴	ملاحظات خطای نوع II	۴-۱-۱۳
۷۸۵	کمترین فراوانی انتظاری مورد نیاز	۵-۱-۱۳
۷۸۵	کاربرد رایانه	۶-۱-۱۳
۷۸۶	آزمون تساوی چندین نسبت	۷-۱-۱۳
۷۸۸	آزمون تصادفی بودن	۲-۱۳
۷۸۹	آزمون تعداد گردش ها	۱-۲-۱۳
۷۹۴	آزمون های دیگر برای تصادفی بودن	۲-۲-۱۳
۷۹۵	آزمون نیکوئی برازش	۳-۱۳
۷۹۶	اهمیت معلوم بودن توزیع	۱-۳-۱۳
۸۰۱	مراحل آزمون فرضیه	۲-۳-۱۳
۸۰۲	آزمون در حالتی که پارامترهای جامعه معلوم باشند	۳-۳-۱۳
۸۰۴	توصیه پایانی	۴-۳-۱۳
۸۰۵	آزمون نیکوئی برازش کلموگرف - اسمیرنف	۴-۱۳

۸۰۹	یافتن مقدار بحرانی	۱-۴-۱۳
۸۱۱	مقایسه دو آزمون	۲-۴-۱۳
۸۱۳	مسائل	
۸۲۹	ضمیمه	
۸۲۹	پاسخ به برخی مسائل	
۸۳۷	جدول‌های آماری	
۸۶۱	منابع	
۸۶۳	نمایه	
۸۶۷	واژه‌نامه انگلیسی به فارسی	

رگرسیون و تحلیل همبستگی

تاکنون، کاربردهای نمونه‌برداری راجع به فقط یک جامعه با مشاهده‌هایی مانند X بوده است. در این فصل با در نظر گرفتن جامعه دیگری مانند Y ، هم‌زمان ۲ جامعه مورد مطالعه قرار خواهند گرفت. هدف این بخش، پیش‌بینی مقادیر یک جامعه بر اساس مشاهده‌های به دست آمده از جامعه دیگر است. مشابه این پیش‌بینی‌ها در مهندسی اهمیت زیادی دارند به نحوی که، اگر نتوان Y را مستقیماً مشاهده کرد، باید مقدار آن را بر اساس مقدار معلوم X پیش‌بینی کرد. بر مبنای یک روش آماری می‌توان تابعی مانند $Y=f(X)$ را پیشنهاد کرد که تحت این تابع، متغیر مستقل^۱ X ، به متغیر وابسته^۲ Y ، تبدیل می‌شود. برای مثال، شاید بتوان زمان نامعلوم Y ، در انجام یک فرآیند شیمیایی را بر-حسب چگالی محلول اولیه X ، بیان کرد.

با اندازه‌گیری و زمان‌سنجی، نمونه‌ای از بسته‌های ماده شیمیایی این فرآیند گردد-آوری شده و مشاهده می‌شود که رابطه‌ای مانند رابطه در پی آمده

۱-Independent Variable

۲-Dependent Variable

$$Y = 9/5 + 0/45 X$$

بین دو کمیت برقرار است. این فرآیند، تحلیل رگرسیون نام دارد. کلمه **رگرسیون** اولین بار توسط فرنسیس گالتون^۱ ارائه شده بود. او با استفاده از قد پدران، توانست قد پسران را پیش‌بینی کند. گالتون دریافت که فرزندان افراد بلند قد، به‌طور متوسط از والدین خود کوتاه‌تر خواهند بود در حالی‌که، فرزندان افراد کوتاه قد به‌طور متوسط از خودشان بلندتر هستند. طبق گزارش او، قد فرزندان دوم به میانگین جامعه کلی گرایش دارد (اما توزیع فراوانی جامعه کلی قد، از تولدی به تولد دیگر ثابت باقی خواهد ماند). آماردانان برای نمایش پیش‌بینی گروه کلی یک متغیر بر اساس یک یا چند متغیر دیگر از واژه **رگرسیون**^۲ استفاده می‌کنند.

مهندسان هنگامی با تحلیل رگرسیون آشنا می‌شوند که بخواهند یک منحنی را بر یک سری از داده‌ها انطباق و برازش^۳ دهند. اما کارایی و کاربرد رگرسیون فراتر از یک عمل مقدماتی مانند «برازش منحنی» است. زیرا، رگرسیون روش توان‌مندی است که به‌طور دقیق خطاهای پیش‌بینی را کنترل می‌کند. یک روش ساده و مناسب به‌منظور دستیابی به منحنی خوب برای برازش، روش حداقل مربعات نام دارد. این روش، تکنیک اصلی رگرسیون جهت کمینه کردن مجموعه انحراف‌های مشاهده‌ها در یک مسئله چند متغیره از سطح یا **رویه رگرسیون**^۴ می‌باشد. بنابراین، یک تابع ریاضی یا همان معادله رگرسیون، خط، منحنی و یا سطح موردنظر را کاملاً مشخص می‌سازد. پس، دو آماردان می‌توانند با استفاده از مشاهده‌های یکسان، نتایج یکسانی را به‌دست بیاورند. تحلیل رگرسیون، به مدد تئوری احتمال، از روش‌های استنباط آماری که تاکنون در این کتاب ارائه شده است استفاده و آن‌ها را توسعه می‌دهد. بنابراین، سطح یا **رویه رگرسیون** دقیقاً یک نقطه آغاز یا شروع است. به‌منظور تأیید معادله رگرسیون ارائه شده، می‌توان از مفهومی چون بازه اطمینان و یا روش‌های آزمون فرضیه، استفاده کرده و کیفیت پیش‌بینی‌های ارائه شده را سنجید. همچنین، می‌توان مفهوم احتمال در بحث

۱- Francis Galton

۲-Regression

۳-Curve Fitting

۴-Regression Surface

کوواریانس را به روش همتای رگرسیون یعنی **تحلیل همبستگی**^۱ گسترش داد به نحوی- که، قوت ارتباط بین X و Y را بتوان مورد سنجش قرار داد.

۹-۱ رگرسیون خطی با روش مبتنی بر حداقل مربعات

کاربرد ابتدایی و اساسی رگرسیون، مبتنی بر **رگرسیون خطی**^۲ می باشد. در این حالت، متغیر Y به وسیله یک خط مستقیم مانند $Y=a+bX$ با X در ارتباط است. بعد از گردآوری یک مجموعه از زوج مشاهده‌ها مانند (X,Y) ، محقق به طور منطقی رابطه بین این زوج مشاهده‌ها را با به کارگیری روش حداقل مربعات می یابد. برای ارائه بهترین خط، کافی است که بهترین شیب خط b ، و عرض از مبدا a را به دست آورد. این رابطه از یک تعداد مشاهده‌های محدود به دست می آید بنابراین، از آن با عنوان **معادله رگرسیون برآوردی**^۳ یاد می کنند. سپس، به منظور پیش بینی Y که بهترین پیش بینی متناظر با یک مقدار خاص از متغیر مستقل X است، از خط رگرسیون استفاده می شود. مقادیر شیب و عرض از مبدا به دست آمده، فقط برآوردی از پارامترهای خط حقیقی هستند. چون مقادیر a و b از نمونه‌ای به نمونه دیگر متفاوت هستند، آن‌ها را با نام **ضرایب رگرسیون برآوردی**^۴ می شناسند.

مثال: نمایش اطلاعات

شکل ۹-۱ نشان می دهد که چگونه می توان برای پیش بینی زمان انجام کار Y (مجهول)، با استفاده از تعداد ورودی‌ها X ، معادله یک خط مناسبی را ارائه کرد. نقاط ترسیم شده در شکل، **مشاهده‌های خام** به دست آمده از یک نمونه ۲۵ تایی، مطابق جدول ۹-۱ مربوط به تعداد ورودی‌ها و زمان انجام کار مربوطه می باشد. این نمایش گرافیکی را **نمودار پراکنش**^۵ می نامند.

۱-Correlation Analysis

۲-Linear Regression

۳-Estimated Regression Equation

۴-Estimated Regression Coefficients

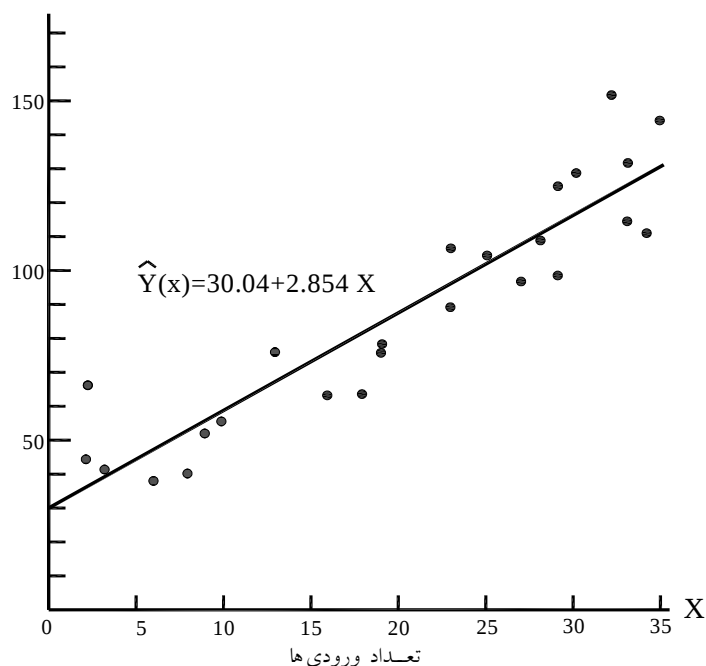
۵-Scatter Diagram

برای تعیین خط رگرسیون برازش شده به این مشاهده‌ها، از روش حداقل مربعات استفاده می‌شود.

۹-۱-۱ روش حداقل مربعات^۱

روش حداقل مربعات، معادله خطی را به دست می‌دهد که مجموع انحراف مشاهده‌های بالا و پایین نسبت به آن خط کمینه باشد. در واقع این روش، بر اساس کمینه کردن مجموع مربع انحراف‌ها در راستای محور عرض نسبت به یک خط خاص مورد تعیین، استوار است. لذا، مجموع انحراف مشاهده‌ها از این خط، نسبت به هر خط مستقیم دیگری کمتر خواهد بود. به تدریج خواهید یافت که این روش علاوه بر سادگی دارای خواص آماری مطلوب دیگری نیز می‌باشد.

زمان انجام کار(ثانیه)



شکل ۹-۱: نمودار پراکنش و خط رگرسیون برای زمان انجام کار

^۱–Method of Least Squares