

بسم الله الرحمن الرحيم

مدل‌سازی و روش‌های انتخاب مدل در آمار و علم داده

(با استفاده از نرم افزارهای SPSS، R و Python)

تألیف:

دکتر سید کامران قریشی
دانشیار گروه آمار

سرشناسه	- ۱۳۵۰ : قریشی، سید کامران،
عنوان و نام	Ghoreishi, Seyed Kamran
پدیدآور	: مدل سازی و روش های انتخاب مدل در آمار و علم داده (با استفاده از نرم افزار، Python و R و SPSS) / تالیف سید کامران قریشی؛ ویراستار عطار دسادات مصطفوی.
مشخصات نشر	: قم: دانشگاه قم، ۱۴۰۳. مشخصات ظاهری: ۲۴۰ ص.
شابک	۹۷۸-۶۲۲-۵۵۱۵-۲۷-۷
موضوع	: آمار ریاضی.
	Mathematical statistics
	تجزیه و تحلیل تک متغیره.
	Univariate analysis
ردی بندی کنگره	ردی بندی دیوبی: ۵۱۹/۵ شماره کتابشناسی ملی: ۹۵۹۰۹۲۵ QA276



انتشارات دانشگاه قم

عنوان: مدل سازی و روش های انتخاب مدل در آمار و علم داده
(با استفاده از نرم افزار Python، R، SPSS و

مؤلف: سید کامران قریشی
ویراستار: دکتر عطار دسادات مصطفوی
چاپ و صحافی: هوشنگی

ناظر فنی: علیرضا معظمی
طراح جلد: احمد رضا حیدری

نوبت و سال چاپ: اول، بهار ۱۴۰۳

شماره کان: ۵۰۰

بهاء: ۱۵۰۰۰۰ تومان

شابک: ۹۷۸-۶۲۲-۵۵۱۵-۲۷-۷

آدرس الکترونیکی: Publication@Qom.ac.ir

کلیه حقوق مادی و معنوی برای ناشر محفوظ است.

قم، بلوار الغدیر، دانشگاه قم، اداره چاپ و انتشارات دانشگاه

تلفن: ۰۲۵-۳۲۱۰۳۳۷۷ - ۰۲۵ - ۳۲۱۰۳۳۴۴

فهرست مطالب

فهرست مطالب

۳۸	استقلال در جداول پیشایندی دو بعدی	۱.۵.۲
۴۰	پیوند در جداول پیشایندی دو بعدی	۲.۵.۲
۴۶	تمرین‌ها	۶.۲
۳ مدل‌های خطی ساده		
۴۹	انواع مدل‌های خطی ساده	۱.۳
۵۰	مدل رگرسیون ساده	۱.۱.۳
۵۲	مدل سری زمانی اتورگرسیو مرتبه اول	۲.۱.۳
۵۳	مدل‌های لوژستیک ساده	۳.۱.۳
۵۴	مدل لگ-خطی ساده	۴.۱.۳
۵۵	مدل طرح آزمایش تک عاملی	۵.۱.۳
۵۸	برآورد پارامترهای مدل خطی ساده	۲.۳
۵۸	برآورد نقطه‌ای پارامترها	۱.۲.۳
۶۶	برآورد بازه‌ای پارامترها	۲.۲.۳
۷۶	آزمون فرض پارامترهای مدل خطی ساده	۳.۳
۸۰	مباحث تکمیلی	۴.۳
۸۴	تمرین‌ها	۵.۳
۴ مدل‌های خطی چندگانه		
۸۵	مقدمه	۱.۴
۹۲	برآورد نقطه‌ای پارامترهای مدل	۲.۴
۹۲	مدل رگرسیون چندگانه	۱.۲.۴
۹۵	مدل اتورگرسیو مرتبه p	۲.۲.۴
۹۷	مدل رگرسیون لوژستیک چندگانه	۳.۲.۴
۱۰۰	مدل رگرسیون لگ-خطی چندگانه	۴.۲.۴
۱۰۱	مدل طرح آزمایش دو عاملی	۵.۲.۴

فهرست مطالب

پ

۱۰۳	برآورد بازه‌ای پارامترهای مدل	۳.۴
۱۰۴	مدل رگرسیون چندگانه	۱.۳.۴
۱۰۶	مدل اتوگرگرسیو مرتبه p	۲.۳.۴
۱۰۷	مدل رگرسیون لوزیتیک چندگانه	۳.۳.۴
۱۰۸	مدل رگرسیون لگ-خطی چندگانه	۴.۳.۴
۱۱۰	مدل طرح آزمایش دو عاملی	۵.۳.۴
۱۱۲	آزمون فرض پارامترهای مدل	۴.۴
۱۱۵	مباحث تكمیلی	۵.۴
۱۲۱	تمرین‌ها	۶.۴

۱۲۳

۵ تحلیل مانده‌ها در مدل‌های خطی

۱۲۳	مانده‌های مدل رگرسیون خطی	۱.۵
۱۲۴	روش‌های استاندارد کردن مانده‌های رگرسیونی	۱.۱.۵
۱۲۷	انواع نمودار مانده‌ها	۲.۱.۵
۱۳۰	مانده‌های مدل اتو رگرسیو	۲.۵
۱۳۱	مانده‌های مدل رگرسیون لوزیتیک	۳.۵
۱۳۵	مانده‌های مدل لگ-خطی	۴.۵
۱۳۷	مانده‌های مدل طرح آزمایش دو عاملی	۵.۵
۱۴۰	نکات تکمیلی	۶.۵
۱۴۰	آزمون نرمال بودن مانده‌ها	۱.۶.۵
۱۴۵	آزمون ثابت بودن واریانس مانده‌ها	۲.۶.۵
۱۴۸	آزمون تصادفی بودن مانده‌ها	۳.۶.۵
۱۵۲	تمرین‌ها	۷.۵

۱۵۵

۶ معیارهای انتخاب بهترین مدل

۱۵۵	مقدمه	۱.۶
---------------	-----------------	-----

فهرست مطالب

۲.۶	معیارهای انتخاب مدل در رگرسیون خطی	۱۵۹
۱.۲.۶	ضریب تعیین	۱۵۹
۲.۲.۶	ضریب تعیین تعدل شده	۱۶۰
۳.۲.۶	معیار C_p -مالوس	۱۶۱
۴.۲.۶	معیار PRESS	۱۶۱
۵.۲.۶	معیار فاصله‌های Cook	۱۶۲
۶.۲.۶	معیار اطلاع آکائیکه	۱۶۲
۷.۲.۶	معیار اطلاع بیزی	۱۶۳
۳.۶	معیارهای انتخاب مدل در سری زمانی	۱۶۴
۱.۳.۶	معیار میانگین مطلق درصد خطا	۱۶۵
۲.۳.۶	معیار ریشه دوم میانگین توان دوم خطا	۱۶۵
۳.۳.۶	معیار میانگین مطلق خطا	۱۶۵
۴.۳.۶	معیار اطلاع آکائیکه	۱۶۶
۵.۳.۶	معیار اطلاع بیزی	۱۶۶
۴.۶	معیارهای انتخاب مدل در رگرسیون لوژستیک	۱۶۶
۱.۴.۶	معیار خی دوی پیرسون	۱۶۷
۲.۴.۶	معیار انحراف	۱۶۷
۳.۴.۶	معیار اطلاع آکائیکه	۱۶۸
۴.۴.۶	معیار اطلاع بیزی	۱۶۸
۵.۶	معیارهای انتخاب مدل در رگرسیون لگ-خطی	۱۷۰
۱.۵.۶	معیار خی دوی پیرسون	۱۷۰
۲.۵.۶	معیار آماره خی دوی نسبت درستنمایی	۱۷۱
۳.۵.۶	معیار اطلاع آکائیکه	۱۷۱
۴.۵.۶	معیار اطلاع بیزی	۱۷۲
۶.۶	معیارهای انتخاب مدل در مدل‌های طرح آزمایش	۱۷۳

فهرست مطالب

ث

۱۷۳	ضریب تعیین	۱.۶.۶
۱۷۴	ضریب تعیین تعدل شده	۲.۶.۶
۱۷۵	معیار اطلاع آکائیکه	۳.۶.۶
۱۷۵	معیار اطلاع بیزی	۴.۶.۶
۱۷۷	تمرین ها	۷.۶
۱۷۹	۷ مدل های منظم شده	
۱۷۹	مقدمه	۱.۷
۱۷۹	غربالگری	۱.۱.۷
۱۸۰	الگوریتم حریص	۲.۱.۷
۱۸۱	منظم سازی	۳.۱.۷
۱۸۳	مدل رگرسیون خطی منظم شده	۲.۷
۱۸۳	رگرسیون ستیغی	۱.۲.۷
۱۸۵	رگرسیون لاسو	۲.۲.۷
۱۸۵	رگرسیون شبکه ارتجاعی	۳.۲.۷
۱۸۹	مدل اتو رگرسیو منظم شده	۳.۷
۱۹۱	مدل رگرسیون لوژستیک منظم شده	۴.۷
۱۹۲	مدل رگرسیون لگ-خطی منظم شده	۵.۷
۱۹۳	مدل طرح آزمایش منظم شده	۶.۷
۱۹۷	تمرین ها	۷.۷
۱۹۹	۸ روش های محاسباتی در مدل های خطی	
۱۹۹	محاسبات مربوط به ضرایب همبستگی	۱.۸
۱۹۹	با نرم افزار SPSS	۱.۱.۸
۲۰۰	با نرم افزار R	۲.۱.۸
۲۰۱	با نرم افزار پایتون	۳.۱.۸

فهرست مطالب

۲۰۲	محاسبات مربوط به مدل رگرسیون	۲.۸
۲۰۲	با نرم افزار SPSS	۱.۲.۸
۲۰۳	با نرم افزار R	۲.۲.۸
۲۰۵	با نرم افزار پایتون	۳.۲.۸
۲۰۵	محاسبات مربوط به مدل اتو رگرسیو	۳.۸
۲۰۵	با نرم افزار SPSS	۱.۳.۸
۲۰۶	با نرم افزار R	۲.۳.۸
۲۰۷	با نرم افزار پایتون	۳.۳.۸
۲۰۸	محاسبات مربوط به مدل رگرسیون لوژستیک	۴.۸
۲۰۸	با نرم افزار SPSS	۱.۴.۸
۲۰۸	با نرم افزار R	۲.۴.۸
۲۱۰	با نرم افزار پایتون	۳.۴.۸
۲۱۱	محاسبات مربوط به مدل رگرسیون لگ-خطی	۵.۸
۲۱۱	با نرم افزار SPSS	۱.۵.۸
۲۱۲	با نرم افزار R	۲.۵.۸
۲۱۳	با نرم افزار پایتون	۳.۵.۸
۲۱۴	محاسبات مربوط به مدل طرح آزمایش	۶.۸
۲۱۴	با نرم افزار SPSS	۱.۶.۸
۲۱۴	با نرم افزار R	۲.۶.۸
۲۱۶	با نرم افزار پایتون	۳.۶.۸
۲۱۷	محاسبات مربوط به مدل های منظم شده	۷.۸
۲۱۷	رگرسیون منظم شده	۱.۷.۸
۲۱۹	اتو رگرسیو منظم شده	۲.۷.۸
۲۲۰	رگرسیون لوژستیک منظم شده	۳.۷.۸
۲۲۱	رگرسیون لگ-خطی منظم شده	۴.۷.۸

فهرست مطالب

ج

۲۲۲ طرح آزمایش منظم شده ۵.۷.۸

ح

فهرست مطالب

پیشگفتار

کتاب حاضر جهت آشنایی دانشجویان و محققان در زمینه مدل‌سازی آماری به رشته تحریر در آمده است. به دلیل آنکه در عمل داده‌های گوناگون با مدل‌های متنوع تبیین و تفسیر می‌شوند لذا سعی کرده‌ایم تا از دامنه نسبتاً وسیعی از مدل‌ها برای پیشبرد بحث استفاده نماییم.

هدف اصلی از نگارش این کتاب استفاده از روش ماکسیمم درستنامی، به عنوان رویه یکسان، برای برآورد پارامترها و انحراف معیار متناظر در انواع مدل‌های آماری است.

به جهت اهمیت موضوع همبستگی در بحث مدل‌سازی، در فصل دوم به طور مختصر انواع ضرایب همبستگی و خواص آنها بحث شده است. فصل سوم به تبیین مدل‌های ساده برای تحلیل داده‌های متنوع از جنس پیوسته و گسسته (با ویژگی وابسته و مستقل) اختصاص دارد. برای رفع هر گونه ابهام تأکید می‌شود که منظور از مدل‌های خطی ساده در اینجا شامل کلیه مدل‌هایی است که در آن‌ها تنها یک متغیر تبیینی، یک عامل یا یک گام قبل یک متغیر به صورت خطی بر متغیر پاسخ تأثیرگذار است.

در فصل چهارم مدل‌های مورد بحث به حالت چندگانه، برای تحلیل داده‌های با متغیرهای بیشتر، توسعه یافته‌اند. منظور از مدل‌های خطی چندگانه در این فصل شامل کلیه مدل‌های آماری است که در آن‌ها چند متغیر تبیینی، چند عامل یا چند گام قبل یک متغیر به صورت خطی بر متغیر پاسخ تأثیرگذارند.

در فصل پنجم تحلیل مانده‌های حاصل از به کارگیری انواع مدل‌های آماری بحث و معرفی شده‌اند. فصل ششم به معرفی معیارهای انتخاب بهترین مدل اختصاص دارد. در این فصل جزئیات عمومی و اختصاصی معیارهای انتخاب بهترین مدل، به تفکیک نوع مدل، تشریح شده است. به دلیل اهمیت موضوع منظم‌سازی در مدل‌سازی، در فصل هفتم به این مبحث پرداخته شده است.

به منظور افزایش توانمندی عملی خوانندگان، در به کارگیری مباحث این کتاب، در فصل هشتم کدها و روش‌های محاسباتی بر اساس سه نرم افزار SPSS، R و Python ارائه شده است.

امید است مطالب این کتاب کمک مؤثری بر افزایش توانایی محققین و دانشجویان در

زمینه مدل‌سازی آماری باشد.

هرچند در نوشتمن این کتاب سعی وافر شده است که از لفاظی و کلی‌گویی خودداری شود لیکن کتاب خالی از اشکال نبوده و نویسنده با آغوش باز از نظرات و پیشنهادات کارشناسان و اساتید گرامی برای اصلاح کتاب استقبال می‌کند.

در پایان لازم می‌دانم از خانواده عزیزم که با صبر و حوصله با من همراه بودند تشکر و قدردانی کنم. وَآخِرُ دَعْوَانَا أَنِ الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ.

سید کامران قریشی، پاییز ۱۴۰۲