

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

مبانی و مدل بندی

بیزی داده‌ها



# مبانی و مدل‌بندی بیزی داده‌ها

با استفاده از برنامه‌نویسی BUGS و نرم‌افزار R

مجتبی گنجعلی

دانشگاه شهید بهشتی

تابان باغفلکی

دانشگاه تربیت مدرس



۶۷۱

مرکز چاپ و انتشارات دانشگاه شهید بهشتی

میانی و مدل‌بندی بیزی داده‌ها (با استفاده از برنامه‌نویسی BUGS و نرم‌افزار R)  
دکتر مجتبی گنجعلی - دکتر تابان باغفلکی

ویراستار: دکتر محمدقاسم وحیدی اصل  
حروف‌نگار و صفحه‌آرا: سمیرا دهقان  
ناظر چاپ: صفر ممیزاد  
چاپ اول: ۱۳۹۶  
شمارگان: ۵۰۰  
قیمت: ۳۰۰.۰۰۰ ریال

کلیه حقوق برای دانشگاه شهید بهشتی محفوظ است.

سرشناسه:	گنجعلی، مجتبی، ۱۳۴۳-
عنوان و نام پدیدآور:	میانی و مدل‌بندی بیزی داده‌ها: با استفاده از برنامه‌نویسی BUGS و نرم‌افزار R / مجتبی گنجعلی، تابان باغفلکی.
مشخصات نشر:	تهران: دانشگاه شهید بهشتی، ۱۳۹۶.
مشخصات ظاهری:	هجده، ۴۰۲ ص.
شابک:	۹۷۸ ۹۶۴ ۴۵۷ ۴۰۵ ۴
فروست:	مرکز چاپ و انتشارات شهید بهشتی، ۶۷۱.
وضعیت فهرست‌نویسی:	فیا
موضوع:	نظریه تصمیم‌گیری آماری بیزی؛ Bayesian statistical decision theory؛ وین‌باگ؛ WinBUGS؛ آر (زبان برنامه‌نویسی کامپیوتر)؛ Statistics (Computer program language)؛ R؛ آمار -- داده‌پردازی؛ -- Data processing
شناسه افزوده:	باغفلکی، تابان، ۱۳۶۴-
شناسه افزوده:	دانشگاه شهید بهشتی
رده‌بندی کنگره:	۱۳۹۶ م۲ گ/۵/۵/۲۷۹/۵ QA۲۷۹
رده‌بندی دیویی:	۵۱۹/۵۴۲
شماره کتابشناسی ملی:	۵۰۰۴۱۹۹

کد ناشر ۱۰۰۱۷۳۴

www.pub.sbu.ac.ir  
unipress@mail.sbu.ac.ir

تقدیم به فرزندان عزیزم علیرضا و فاطمه  
مجتبی گنجعلی

تقدیم به پدر، مادر و خواهر عزیزم  
تابان باغفلکی



## فهرست مطالب

پیشگفتار ..... سیزده

### بخش اول: مبانی آمار بیزی

فصل ۱ مقدمه‌ای بر استنباط بیزی .....	۳
۱.۱. مقدمه .....	۳
۲.۱. روش‌های استنباط کلاسیک .....	۸
۱.۲.۱. برآوردگر نقطه‌ای .....	۸
۲.۲.۱. برآورد بازه‌ای و آزمون فرض .....	۱۰
۳.۱. روش استنباط بیزی .....	۱۰
۱.۳.۱. توزیع پیشینی .....	۱۰
۲.۳.۱. توزیع پسینی .....	۱۲
۴.۱. مقایسه روش‌های استنباط بیزی و کلاسیک .....	۱۴
۱.۴.۱. استنباط در نسبت رخداد پیشامد مطلوب .....	۱۴
۲.۴.۱. مشکل کوچک‌بودن اندازه نمونه .....	۱۸
۳.۴.۱. ناکامی برآورد کلاسیک در یافتن برآورد ماکسیمم درست‌نمایی .....	۲۰
۴.۴.۱. برآورد نقطه‌ای در آمار بیزی و استنباط کلاسیک از لحاظ تفاوت در گزینش اولیه .....	۲۱
۵.۴.۱. استنباط بیزی برای یک بردار پارامتر دوبعدی .....	۲۴
۶.۴.۱. تفاوت استنباط بیزی و کلاسیک در حضور پارامتر مزاحم .....	۲۶
۷.۴.۱. مقایسه روش‌های کلاسیک و فراوانی‌گرا در مدل‌های آمیخته خطی .....	۲۸
تمرین‌های فصل اول .....	۳۰
فصل ۲ روش‌های گزینش توزیع‌های پیشینی در استنباط بیزی .....	۳۳
۱.۲. مقدمه .....	۳۳
۲.۲. توزیع پیشینی بر مبنای احتمال ذهنی .....	۳۳
۱.۲.۲. تعریف احتمال ذهنی .....	۳۳
۲.۲.۲. روش بافت‌نگاشت برای تعیین توزیع پیشینی .....	۳۴
۳.۲.۲. روش محتمل‌نمایی نسبی .....	۳۶
۴.۲.۲. روش انطباق یک شکل تابعی .....	۳۷
۵.۲.۲. تعیین تابع توزیع احتمال .....	۳۸
۳.۲. چگالی‌های پیشینی ناآگاهی‌بخش .....	۳۹
۱.۳.۲. پیشینی‌های ناآگاهی‌بخش برای پارامترهای مکان و پارامترهای مقیاس .....	۳۹

۴۱	..... پیشینی جفریز	۲.۳.۲
۴۴	..... پیشینی مزدوج	۴.۲
۴۵	..... پیشینی ماکسیمم آنتروپی	۵.۲
۴۸	..... پیشینی ماکسیمم درستنمایی نوع II	۶.۲
۵۰	..... تمرین‌های فصل دوم	
۵۳	..... فصل ۳ روش‌های شبیه‌سازی مونت کارلویی	
۵۳	..... ۱.۳ مقدمه	
۵۴	..... ۲.۳ تبدیل انتگرال احتمال	
۵۷	..... ۳.۳ انتگرال گیری مونت کارلویی	
۶۴	..... ۴.۳ روش نمونه‌گیری رد- پذیرش	
۷۴	..... ۵.۳ روش نمونه‌گیری از نقاط مهم	
۷۶	..... ۶.۳ روش نمونه‌گیری رد- سازوار	
۸۴	..... ۷.۳ روش بازنمونه‌گیری از نمونه‌گیری از نقاط مهم	
۸۶	..... تمرین‌های فصل سوم	
۸۹	..... فصل ۴ شبیه‌سازی به روش مونت کارلویی زنجیر مارکوفی	
۸۹	..... ۱.۴ مقدمه	
۹۰	..... ۲.۴ مقدمه‌ای بر فرایندهای تصادفی	
۹۰	..... ۱.۲.۴ تعریف فرایند تصادفی	
۹۰	..... ۲.۲.۴ مدل ریاضی برای یک فرایند تصادفی	
۹۱	..... ۳.۴ زنجیرهای مارکوف	
۹۲	..... ۱.۳.۴ زنجیرهای مارکوف زمان‌ناوردا با فضای وضعیت متناهی	
۹۳	..... ۲.۳.۴ ماتریس‌های احتمال تغییر وضعیت از مراتب بالاتر	
۹۴	..... ۳.۳.۴ توزیع احتمال اشغال	
۹۵	..... ۴.۳.۴ توزیع درازمدت زنجیر و معادله‌های وضعیت پایا	
۹۵	..... ۴.۴ رده‌بندی وضعیت‌های یک زنجیر مارکوف	
۹۵	..... ۱.۴.۴ زمان‌های گذر و احتمال‌های نخستین بازگشت	
۹۶	..... ۲.۴.۴ میانگین زمان نخستین گذر	
۹۷	..... ۳.۴.۴ رده‌بندی وضعیت‌های زنجیر مارکوف	
۹۸	..... ۴.۴.۴ تابع مولد احتمال‌های نخستین بازگشت	
۹۹	..... ۵.۴.۴ زنجیرهای مارکوف نافروکاستنی	
۱۰۰	..... ۵.۴ نمونه‌گیری از یک زنجیر مارکوف	
۱۰۳	..... ۱.۵.۴ زنجیرهای مانای زمان‌برگشت‌پذیر و تعادل تفصیلی	



۱۰۴	..... ۲.۵.۴. زنجیرهای مارکوف زمان برگشت پذیر
۱۰۵	..... ۶.۴. الگوریتم متروپولیس
۱۰۸	..... ۷.۴. زنجیرهای مارکوف با فضای وضعیت پیوسته
۱۰۹	..... ۸.۴. الگوریتم متروپولیس- هستینگس و ساختار توزیع پسینی
۱۱۱	..... ۱.۸.۴. یافتن یک زنجیر مارکوف که توزیع پسینی را به عنوان توزیع درازمدت خود دارد
۱۱۳	..... ۹.۴. الگوریتم متروپولیس- هستینگس برای بردارهای یک بعدی
۱۱۸	..... ۱.۹.۴. الگوریتم متروپولیس- هستینگس مستقل
۱۲۱	..... ۲.۹.۴. الگوریتم متروپولیس- هستینگس قدم زدن تصادفی
۱۲۵	..... ۱۰.۴. الگوریتم متروپولیس- هستینگس برای برداری از پارامترها
۱۳۷	..... ۱۱.۴. الگوریتم متروپولیس- هستینگس بلوک به بلوک
۱۴۰	..... ۱۲.۴. نمونه گیری گیبز
۱۴۳	..... تمرین های فصل چهارم
۱۴۵	..... فصل ۵ آزمون فرض های آماری، بازه های باورمند و توزیع پیشگو از دیدگاه بیزی
۱۴۵	..... ۱.۵. مقدمه
۱۴۵	..... ۲.۵. آزمون بیزی فرض ها
۱۵۲	..... ۳.۵. عامل بیزی
۱۵۶	..... ۴.۵. $p$ - مقدار بیزی
۱۶۱	..... ۵.۵. ناحیه های باورمند یا بازه های اطمینان بیزی
۱۶۴	..... ۶.۵. توزیع پسینی پیشگوی مقدار جدید
۱۶۷	..... ۷.۵. استنباط پیشگو بر مبنای تشخیص حذف موردی
۱۶۹	..... تمرین های فصل پنجم
۱۷۱	..... فصل ۶ تحلیل بیزی سلسله مراتبی و بیزی تجربی
۱۷۱	..... ۱.۶. مقدمه
۱۷۱	..... ۲.۶. مدل بیزی تجربی
۱۷۳	..... ۳.۶. برآورد نقطه ای بیزی تجربی ناپارامتری
۱۷۳	..... ۱.۳.۶. روش بیزی تجربی ناپارامتری ساده
۱۷۶	..... ۴.۶. برآورد نقطه ای بیزی تجربی پارامتری
۱۸۰	..... ۵.۶. مدل بیزی سلسله مراتبی
۱۸۱	..... ۱.۵.۶. یک مثال انگیزه بخش
۱۸۲	..... ۲.۵.۶. چرا از مدل سلسله مراتبی استفاده می شود؟
۱۸۴	..... تمرین های فصل ۶

## بخش دوم: مدل‌بندی بیزی داده‌ها

فصل ۷ تحلیل بیزی مدل‌های خطی با فرض توزیعی نرمال.....	۱۸۹
۱.۷. مقدمه.....	۱۸۹
۲.۷. رگرسیون خطی ساده بیزی.....	۱۹۰
۱.۲.۷. رگرسیون خطی ساده از دیدگاه کلاسیک.....	۱۹۰
۲.۲.۷. رگرسیون خطی ساده از دیدگاه بیزی.....	۱۹۲
۳.۷. مدل‌های نرمال چندمتغیره.....	۲۰۳
۱.۳.۷. ساختار ماتریس کوواریانس.....	۲۰۳
۴.۷. رگرسیون نرمال چندمتغیره.....	۲۱۳
۵.۷. داده‌های گم‌شده و نمونه‌های پسینی.....	۲۱۶
۶.۷. تحلیل واریانس یک‌طرفه.....	۲۲۲
۱.۶.۷. تحلیل واریانس یک‌طرفه بیزی.....	۲۲۲
۷.۷. تحلیل واریانس دو طرفه.....	۲۲۷
تمرین‌های فصل ۷.....	۲۳۲
فصل ۸ تحلیل بیزی مدل‌های آمیخته خطی.....	۲۳۷
۱.۸. مقدمه.....	۲۳۷
۲.۸. مدل‌های آمیخته خطی.....	۲۳۷
۳.۸. مشخص‌سازی مدل.....	۲۴۰
۱.۳.۸. ساختار بیزی مدل‌های اثرهای تصادفی.....	۲۴۳
۲.۳.۸. مثال‌هایی از استفاده از مدل‌های آمیخته خطی.....	۲۴۴
۳.۳.۸. مدل‌های آمیخته خطی با عرض از مبدأ تصادفی برای مدل‌بندی داده‌های طولی.....	۲۴۷
تمرین‌های فصل ۸.....	۲۵۵
فصل ۹ تحلیل بیزی مدل‌های خطی و آمیخته خطی تعمیم‌یافته.....	۲۵۷
۱.۹. مقدمه.....	۲۵۷
۲.۹. مروری بر مدل‌های خطی تعمیم‌یافته از دیدگاه کلاسیک.....	۲۵۸
۳.۹. مدل‌های خطی تعمیم‌یافته از دیدگاه بیزی.....	۲۶۰
۱.۳.۹. تحلیل بیزی داده‌های دودویی با استفاده از مدل رگرسیونی لوژستیک.....	۲۶۰
۲.۳.۹. تحلیل بیزی داده‌های شمارشی با فرض توزیعی پواسون.....	۲۶۶
۴.۹. مدل‌های آمیخته خطی تعمیم‌یافته از دیدگاه بیزی.....	۲۶۹
۱.۴.۹. مدل رگرسیونی لوژستیک با اثرهای تصادفی از دیدگاه بیزی.....	۲۷۱
۲.۴.۹. مدل رگرسیونی پواسون با اثرهای تصادفی از دیدگاه بیزی.....	۲۷۲
تمرین‌های فصل ۹.....	۲۷۶

فصل ۱۰ تحلیل بیزی مدل‌های زمان تا رخداد یک پیشامد .....	۲۸۱
۱.۱.۰ مقدمه .....	۲۸۱
۲.۱.۰ داده‌های سانسوریده .....	۲۸۲
۱.۲.۱.۰ سانسور نوع اول .....	۲۸۲
۲.۲.۱.۰ سانسور نوع دوم .....	۲۸۳
۳.۲.۱.۰ سانسور نوع سوم .....	۲۸۳
۳.۱.۰ تابع بقا و تابع خطر .....	۲۸۴
۴.۱.۰ تأثیر متغیرهای تبیینی در مطالعه‌های بقا .....	۲۸۶
۱.۴.۱.۰ مدل خطرهای متناسب .....	۲۸۶
۲.۴.۱.۰ مدل زمان شکست شتابیده .....	۲۸۸
۵.۱.۰ توصیف داده‌ها .....	۲۸۹
۶.۱.۰ مدل‌های مورد استفاده برای تحلیل داده‌های طول مدت بیکاری .....	۲۹۲
۷.۱.۰ کاربرد .....	۲۹۵
۱.۷.۱.۰ تحلیل داده‌های طول مدت بیکاری در ایران .....	۲۹۵
۲.۷.۱.۰ تحلیل حساسیت .....	۳۰۱
۸.۱.۰ بررسی ناهمگنی در مدل‌های طول مدت بیکاری با استفاده از مدل اثرهای تصادفی .....	۳۰۳
۱.۸.۱.۰ روش برآورد بیزی اثرهای تصادفی برای کنترل تغییرات گروه مشخص .....	۳۰۴
۲.۸.۱.۰ مدل زمان شکست شتابیده تحت فرض توزیعی لگ‌نرمال و لگ‌لوژستیک .....	۳۰۵
۳.۸.۱.۰ مدل زمان شکست شتابیده با فرض توزیعی وایبول .....	۳۰۶
۹.۱.۰ اندازه‌گیری ناهمگنی مشاهده‌نشده در داده‌های طول مدت بیکاری .....	۳۰۷
۱.۹.۱.۰ تحلیل داده‌های طول مدت بیکاری با استفاده از مدل‌هایی برای بیان ناهمگنی مشاهده‌نشده .....	۳۰۷
تمرین‌های فصل ۱۰ .....	۳۱۲
فصل ۱۱ تحلیل بیزی توأم اندازه‌های طولی و برآمدهای زمان تا رخداد پیشامد .....	۳۱۵
۱.۱.۱ مقدمه .....	۳۱۵
۲.۱.۱ توزیع متغیر نرمال تقسیم‌شده بر یک متغیر تصادفی با تکیه‌گاه مقادیر مثبت .....	۳۱۸
۳.۱.۱ مدل‌ها و نمادها .....	۳۱۹
۴.۱.۱ مدل‌بندی توأم اندازه‌های طولی و داده‌های بقا .....	۳۲۰
۵.۱.۱ تحلیل داده‌ها .....	۳۲۵
تمرین‌های فصل ۱۱ .....	۳۳۶
منابع .....	۳۳۷

پیوست الف آشنایی با عملیات بیزی با استفاده از نرم افزار R	۳۴۹
الف.۱. مقدمه	۳۴۹
الف.۲. استنباط بیزی برای نسبت	۳۴۹
الف.۳. پیشگوی پسینی در توزیع دوجمله‌ای	۳۵۵
الف.۴. برآورد بیزی واریانس توزیع نرمال با فرض معلوم بودن میانگین	۳۵۸
الف.۵. استنباط بیزی در توزیع نرمال (هر دو پارامتر نامعلوم)	۳۵۸
الف.۶. استنباط بیزی در مدل لوژستیک	۳۶۰
الف.۷. محاسبه مد پسینی با استفاده از نرم افزار R	۳۶۲
الف.۸. نمونه‌گیری رد- پذیرش	۳۶۶
الف.۹. روش نمونه‌گیری از نقاط مهم	۳۶۷
الف.۱۰. الگوریتم متروپولیس- هستینگس با استفاده از R	۳۶۸
الف.۱۱. نمونه‌گیری گیبز با استفاده از R	۳۷۱
<b>پیوست ب آشنایی با نرم افزار WinBUGS</b>	<b>۳۷۳</b>
ب.۱. مقدمه	۳۷۳
ب.۲. نرم افزارهای WinBUGS و OpenBUGS	۳۷۳
ب.۳. منوهای WinBUGS	۳۷۴
ب.۴. انواع متغیرها و پارامترها در WinBUGS	۳۷۵
ب.۵. ساختار ورود داده‌ها در نرم افزار WinBUGS	۳۷۵
ب.۶. تابع‌های معرفی شده در WinBUGS	۳۷۶
ب.۷. تشریح یک مدل ساده با استفاده از WinBUGS	۳۷۸
ب.۸. مشخص سازی داده‌ها	۳۷۹
ب.۹. مقادیر اولیه در WinBUGS	۳۸۱
ب.۱۰. اجرای برنامه در WinBUGS	۳۸۲
ب.۱۱. بررسی همگرایی زنجیر مارکوف	۳۸۷
واژه‌نامه فارسی- انگلیسی	۳۸۹
واژه‌نامه انگلیسی- فارسی	۳۹۱
نام‌نامه	۳۹۳
نمایه	۳۹۷

## پیشگفتار

پیشرفت قابل توجه در نظریه و کاربرد روش‌های استنباط بیزی در تحلیل آماری داده‌ها باعث افزایش علاقه پژوهشگران آماری به یادگیری روش‌های بیزی شده است. در انجام استنباط آماری با استفاده از روش‌های بیزی گاهی لازم است از روش‌های محاسباتی بر مبنای نمونه‌گیری استفاده شود. از آنجاکه امروزه نرم‌افزارهای زیادی با قابلیت نمونه‌گیری از توزیع‌های آماری در دسترس عموم قرار گرفته‌اند، بسیاری از مسائل پیچیده در آمار کلاسیک با روش‌های بیزی به‌سادگی حل می‌شوند. با توجه به نبود کتاب‌های استنباط بیزی به زبان فارسی در کشور، هدف از ارائه این کتاب فراهم آوردن یک نوشتار آموزشی درباره آمار بیزی است. این کتاب می‌تواند برای پژوهشگران، دانشجویان آمار و دیگر افرادی که به تحلیل آمار کاربردی با استفاده از روش‌های بیزی علاقه‌مند هستند، مفید باشد.

برای مطالعه این کتاب، داشتن اطلاعات پایه‌ای از نظریه احتمال و آمار و همچنین اطلاعاتی در مورد استفاده از نرم‌افزار R مورد نیاز است. در این کتاب که شامل دو بخش (۱) مبانی آمار بیزی و (۲) مدل‌بندی بیزی داده‌ها است، ضمن بیان مفاهیم آمار بیزی، به دلیل کاربردی بودن این مفاهیم، برنامه‌های کامپیوتری لازم نیز ارائه شده‌اند. این نکته کتاب حاضر را از کتاب‌های موجود داخلی و خارجی متمایز می‌کند. ابزارهای کامپیوتری که در این راستا از آن‌ها بهره زیادی برده شده است، نرم‌افزارهای R و برنامه‌نویسی به زبان BUGS (با استفاده از نرم‌افزار WinBUGS و OpenBUGS) برای تحلیل بیزی داده‌ها است. استفاده از زبان برنامه‌نویسی BUGS، امکان معرفی روش‌های ساده و انعطاف‌پذیر برای مدل‌بندی داده‌های حتی پیچیده را در بخش دوم کتاب فراهم می‌کند. نکته قابل توجه این است که در این وضعیت مشخص‌سازی چگالی‌های شرطی کامل که در برخی مباحث بیزی به کار می‌روند، ضروری نیست. علاوه بر برنامه‌های ارائه شده در فصل‌های گوناگون، مروری بر نحوه استفاده از زبان برنامه‌نویسی BUGS و تحلیل بیزی با استفاده از نرم‌افزار R در قالب دو پیوست داده شده‌اند که خوانندگان را در اجرای کاربردی عملیات بیزی یاری می‌دهند. از دیگر مزایای استفاده از این دو نرم‌افزار امکان فراخوانی نرم‌افزار OpenBUGS یا WinBUGS در R است، که با استفاده از بسته نرم‌افزاری R2OpenBUGS یا R2WinBUGS انجام می‌شود.

دنيس ليندلى، آماردان بيزى مشهور، براى استنباط بيزى بر دو قانون بنيادى (۱) پيروى از قوانين احتمال و (۲) مدل بندى عدم حتميت با استفاده از اصول احتمال تأكيد كرد. بنا بر اين، نياز به دانستن مفاهيم احتمال براى تحليل بيزى داده ها ضرورى است. آهاگان و فورستر (۲۰۰۴) چهار دليل زير را براى ضرورت استفاده از استنباط بيزى به جاى روش هاى كلاسيك بيان مى كنند:

۱. استنباط بيزى داراى يك پايه فلسفى محكم بر مبنای قوانين احتمال و از همه مهم تر قضيه بيز است.

۲. استنباط بيزى انعطاف پذير است. چارچوب استنباط بيزى شامل ايجاد تابع درست نمايى، تعريف توزيع پيشينى و در نهايت محاسبه توزيع پسينى، با شرطى كردن بر داده هاى مشاهده شده و با استفاده از قضيه بيز است. اين چارچوب قالب انعطاف پذيرى را براى استنباط در مدل هاى پيچيده فراهم مى كند.

۳. تفسير نتايج در استنباط هاى كلاسيك واضح نيست. به طور مثال، تفسير دقيق برآوردهاى بازه هاى و  $p$ -مقدار مشكل است. اغلب استفاده كنندگان از بازه هاى اطمينان و  $p$ -مقدار معنى اين كميت ها را نمى دانند و به طور تلويحى از مفهوم بيزى آن ها براى تفسير نتايج خود استفاده مى كنند. در مقابل فهم و تفسير هم تاي بيزى اين كميت ها بسيار ساده است. بازه هاى باورمند بيزى دقيقاً همان چيزهائى هستند كه اغلب مردم در مورد يك بازه فكر مى كنند و يك آزمون فرض بيزى به سادگى احتمال پسينى فرض ها را براى پذيرش يا رد يك فرض مقايسه مى كند.

۴. گاهى شيوه گزينش توزيع هاى پيشينى به عنوان يك نقطه ضعف استنباط بيزى قلمداد مى شود. از طرفى قابليت در نظر گرفتن توزيع هاى پيشينى به اين معنى است كه اطلاعات بيشترى را مى توان در استنباط لحاظ كرد (مثال ۲.۱ از فصل ۱ در مورد موسيقى دان و مرد مست را ببينيد).

در بخش اول اين كتاب توجه خاصى به مفاهيم پايه هاى آمار بيزى به عمل آمده است. علاوه بر اين، ضمن بيان مباحث اساسى آمار بيزى شامل روش هاى گزينش توزيع هاى پيشينى در استنباط بيزى، روش هاى شبیه سازی مونت كارلويى، شبیه سازی به روش مونت كارلويى زنجير ماركوفى، آزمون فرض آمارى از ديده گاه بيزى و تحليل بيزى تجربى و سلسله مراتبى، در قالب فصل هاى جداگانه، مدل هاى خطى نيز با فرض توزيعى نرمال از ديده گاه بيزى، مدل هاى آميخته خطى، تحليل بيزى مدل هاى خطى تعميم يافته، تحليل مدل هاى زمان تا رخداد پيشامد

و مدل‌بندی توأم اندازه‌های طولی و داده‌های بقا از دیدگاه بیزی در بخش دوم این کتاب گنجانده شده است. فصل اول، شامل مروری اجمالی بر آمار کلاسیک (فراوانی‌گرا) و بیزی، و مقایسه روش‌های بیزی و کلاسیک است. در ادامه مطلب، ابتدا روش‌های گزینش توزیع‌های پیشینی در استنباط بیزی در فصل دوم بحث می‌شوند، سپس در فصل سوم به معرفی روش‌های مونت کارلویی برای شبیه‌سازی نمونه از توزیع پسینی می‌پردازیم. این فصل شامل روش انتگرال‌گیری مونت کارلویی، روش نمونه‌گیری رد-پذیرش، روش نمونه‌گیری از نقاط مهم و روش بازنمونه‌گیری نمونه‌گیری از نقاط مهم است. در فصل چهارم به روش‌های مونت کارلویی زنجیر مارکوفی برای شبیه‌سازی نمونه از برداری از پارامترها پرداخته می‌شود. این فصل بحث مفصلی در استفاده از الگوریتم متروپولیس-هستینگس را شامل می‌شود. در فصل پنجم به آزمون فرض به روش بیزی پرداخته شده که علاوه بر بحث در چگونگی آزمون بیزی فرض‌ها، عامل بیزی،  $p$ -مقدار بیزی، بازه‌های باورمند بیزی، چگال‌ترین بازه پسینی و توزیع پسینی پیشگو در آن گنجانده شده است. مدل‌های بیزی سلسله‌مراتبی و بیزی تجربی در فصل ششم گنجانده شده است.

در بخش دوم کتاب، مدل‌های خطی با فرض توزیعی نرمال از دیدگاه بیزی عنوان فصل هفتم است. در این فصل ساختارهای مختلف ماتریس کوواریانس مانند ساختار اتورگرسیو، میانگین متحرک و ساختارهای دیگر بحث شده است. علاوه بر این چگونگی استفاده از مدل‌های حاشیه‌ای برای مدل‌بندی داده‌های طولی نیز مورد بحث قرار گرفته است. همچنین تحلیل واریانس یک‌طرفه در این فصل بحث شده است. در فصل هشتم به بحث در تحلیل بیزی مدل‌های آمیخته خطی پرداخته شده و از مدل‌های آمیخته خطی برای مدل‌بندی داده‌های گروه‌بندی شده و داده‌های طولی استفاده شده است. در فصل نهم مدل‌های خطی تعمیم‌یافته از دیدگاه بیزی بحث شده است که در این فصل به مدل‌بندی داده‌های دودویی، دو جمله‌ای، و داده‌های شمارشی در هر دو حالت مقطعی و طولی از دیدگاه بیزی پرداخته شده است. فصل دهم بعد از مقدمه‌ای بر تحلیل بقا، به روش‌های تحلیل بیزی داده‌های بقا می‌پردازد که حاوی برخی از پژوهش‌ها و برنامه‌های نوشته‌شده توسط نویسندگان در این راستا است. بالاخره در فصل ۱۱ به مدل‌بندی توأم داده‌های طولی و داده‌های بقا از دیدگاه بیزی پرداخته می‌شود. همچنین، همان‌طور که گفته شد، آشنایی با عملیات بیزی با استفاده از نرم‌افزار R و آشنایی با زبان برنامه‌نویسی BUGS در دو پیوست ارائه شده‌اند.

این کتاب برای همه علاقه‌مندان به یادگیری تحلیل بیزی داده‌ها قابل استفاده است. با وجود

این فصل‌های ذکرشده در زیر برای تدریس در دوره‌های مختلف تحصیلی پیشنهاد می‌شوند. برای مثال در دوره کارشناسی آمار تدریس فصل‌های ۱، ۲، ۳ و ۵ بخش اول کتاب، به شرط آنکه دانشجوی درس آمار ریاضی ۲ را گذرانده باشد، توصیه می‌شود.

#### بخش اول

فصل‌های ۱ تا ۳	فصل ۴	فصل ۵	فصل ۶
<ul style="list-style-type: none"> <li>• کارشناسی آمار</li> <li>• کارشناسی ارشد آمار (همه گرایش‌ها)</li> <li>• دکتری آمار و آمار زیستی</li> <li>• کارشناسی ارشد و دکتری همه‌گیرشناسی</li> </ul>	<ul style="list-style-type: none"> <li>• کارشناسی ارشد آمار (گرایش آمار ریاضی)</li> <li>• دکتری آمار و آمار زیستی</li> </ul>	<ul style="list-style-type: none"> <li>• کارشناسی آمار</li> <li>• کارشناسی ارشد آمار (همه گرایش‌ها)</li> <li>• دکتری آمار و آمار زیستی</li> <li>• کارشناسی ارشد و دکتری همه‌گیرشناسی</li> </ul>	<ul style="list-style-type: none"> <li>• کارشناسی ارشد آمار (همه گرایش‌ها)</li> <li>• دکتری آمار و آمار زیستی</li> <li>• کارشناسی ارشد و دکتری همه‌گیرشناسی</li> </ul>

#### بخش دوم

فصل ۷ و ۸	فصل ۹	فصل ۱۰	فصل ۱۱
<ul style="list-style-type: none"> <li>• کارشناسی ارشد آمار (همه گرایش‌ها)</li> <li>• دکتری آمار و آمار زیستی</li> <li>• کارشناسی ارشد و دکتری همه‌گیرشناسی</li> </ul>	<ul style="list-style-type: none"> <li>• دکتری آمار و آمار زیستی</li> <li>• دکتری همه‌گیرشناسی</li> </ul>	<ul style="list-style-type: none"> <li>• کارشناسی ارشد آمار (همه گرایش‌ها)</li> <li>• دکتری آمار و آمار زیستی</li> <li>• کارشناسی ارشد و دکتری همه‌گیرشناسی</li> </ul>	<ul style="list-style-type: none"> <li>• کارشناسی ارشد آمار (همه گرایش‌ها)</li> <li>• دکتری آمار و آمار زیستی</li> <li>• کارشناسی ارشد و دکتری همه‌گیرشناسی</li> </ul>

از ویراستار محترم این کتاب آقای دکتر محمدقاسم وحیدی اصل کمال تقدیر و تشکر را داریم. همچنین از انتشارات دانشگاه شهید بهشتی به‌ویژه جناب آقای دکتر محمدرضا نبید، همکاران بزرگوار سرکار خانم‌ها آذر مه سنجری و خانم سمیرا دهقان برای آماده‌سازی و چاپ کتاب تشکر می‌کنیم.

به‌عنوان کلام آخر تقاضا می‌کنیم که ما را از انتقادات و پیشنهادهای خود در جهت ارتقای کیفی کتاب بهره‌مند سازید و همچنین سپاسگزار خواهیم بود اگر نظرها، پیشنهادهای و سؤال‌های احتمالی خود را با ما در میان بگذارید.

مجتبی گنجعلی، عضو هیئت علمی گروه آمار دانشگاه شهید بهشتی

تابان باغفلکی، عضو هیئت علمی گروه آمار دانشگاه تربیت مدرس

پاییز ۱۳۹۶



جدول توابع چگالی استفاده شده

نماد اختصاری	تابع چگالی	دامنه متغیر	پارامترها	نام توزیع
$N(\mu, \sigma^2)$	$f(x   \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$	$x \in R$	$\mu, \sigma^2$ ( $\mu \in R, \sigma^2 > 0$ )	نرمال
$\text{Exp}(\lambda)$	$f(x   \lambda) = \lambda e^{-\lambda x}$	$x > 0$	$\lambda$ ( $\lambda > 0$ )	نمایی
$\Gamma(\alpha, \beta)$	$f(x   \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$	$x > 0$	$\alpha, \beta$ ( $\alpha, \beta > 0$ )	گاما
$I\Gamma(\alpha, \beta)$	$f(x   \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} \exp\left(-\frac{\beta}{x}\right)$	$x > 0$	$\alpha, \beta$ ( $\alpha, \beta > 0$ )	گامای معکوس
$\text{Beta}(\alpha, \beta)$	$f(x   \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$	$0 < x < 1$	$\alpha, \beta$ ( $\alpha, \beta > 0$ )	بتا
$C(\mu, \tau)$	$f(x   \mu, \tau) = \frac{1}{\tau\pi \left(1 + \left(\frac{x-\mu}{\tau}\right)^2\right)}$	$x \in R$	$\mu, \tau$ ( $\mu \in R, \tau > 0$ )	کوشی
$U(a, b)$	$f(x   a, b) = \frac{1}{b-a} I_{[a,b]}(x)$	$x \in [a, b]$	$a, b$ ( $a, b \in R$ )	یکنواخت
$\text{Weibull}(\lambda, \gamma)$	$f(x   \lambda, \gamma) = \lambda\gamma x^{\gamma-1} e^{-\lambda x^\gamma}$	$x > 0$	$\lambda, \gamma$ ( $\lambda, \gamma > 0$ )	وایبول
$LN(\mu, \sigma^2)$	$f(x   \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\log(x)-\mu)^2}$	$x > 0$	$\mu, \sigma^2$ ( $\mu \in R, \sigma^2 > 0$ )	لگ نرمال
$LL(\mu, \sigma^2)$	$f(x   \mu, \sigma^2) = \frac{\exp\left(\frac{\log(x)-\mu}{\sigma}\right)}{x\sigma \left(1 + \exp\left(\frac{\log(x)-\mu}{\sigma}\right)\right)^2}$	$x > 0$	$\mu, \sigma^2$ ( $\mu \in R, \sigma^2 > 0$ )	لگ لوژستیک
$T(v)$	$f(x   v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \frac{1}{\sqrt{v\pi}} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}}$	$x \in R$	$v$ ( $v > 0$ )	t
$\text{Ber}(p)$	$f(x   p) = p^x (1-p)^{1-x}$	$x \in \{0, 1\}$	$p$ ( $0 < p < 1$ )	برنولی
$\text{Bin}(n, p)$	$f(x   n, p) = \binom{n}{x} p^x (1-p)^{n-x}$	$x \in \{0, 1, \dots, n\}$	$n, p$ ( $0 < p < 1, n = 1, 2, \dots$ )	دوجمله‌ای

ادامهٔ جدول توابع چگالی استفاده شده

نماد اختصاری	تابع چگالی	دامنهٔ متغیر	پارامترها	نام توزیع
$\text{Pos}(\lambda)$	$f(x \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$	$x \in \{0, 1, 2, \dots\}$	$\lambda (\lambda > 0)$	پواسون
$\text{Dirichlet}(\boldsymbol{\alpha})$ , $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)$	$f(\boldsymbol{\pi} \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k \pi_i^{\alpha_i - 1}$	$0 < \pi_i < 1,$ $\sum_{i=1}^k \pi_i = 1$	$\alpha_i > 0,$ $i = 1, 2, \dots, k$	دیریکله
$IW(\boldsymbol{\Psi}, \nu)$	$f(\mathbf{x} \boldsymbol{\Psi}, \nu) = \frac{ \boldsymbol{\Psi} ^{\frac{\nu}{2}}}{2^{\frac{\nu p}{2}} \Gamma_p(\frac{\nu}{2})}  \mathbf{x} ^{-\frac{\nu+p+1}{2}} e^{-\frac{1}{2} \text{trace}(\boldsymbol{\Psi} \mathbf{x}^{-1})}$	$\mathbf{x}$ ماتریسی $p \times p$ معین مثبت است.	$\nu > p - 1$ $\boldsymbol{\Psi}$ ماتریسی $p \times p$ و معین مثبت است.	ویشارت معکوس
$EV(\mu, \sigma)$	$f(x) = \frac{1}{\sigma} e^{-(x-\mu)/\sigma} e^{-e^{-(x-\mu)/\sigma}}$	$x \in R$	$\mu \in R, \sigma > 0$	مقدار کرانگین

بخش اول  
مبانی آمار بیزی



## فصل ۱

### مقدمه‌ای بر استنباط بیزی

#### ۱.۱. مقدمه

اصطلاح "آمار بیزی" از نام ریاضی‌دان و کشیش انگلیسی تامس بیز (۱۷۰۲-۱۷۶۱) به دلیل ارائه قضیه بیز (هاجز و لی‌من، ۱۹۷۰ و مین‌هال و همکاران، ۲۰۰۹) اقتباس شده است. در واقع، قضیه معروف بیز نقطه شروع آمار بیزی است. استنباط در روش‌های بیزی براساس این قضیه با ترکیب اطلاعات پیشینی غیرنمونه‌ای و داده‌های موجود صورت می‌گیرد. اطلاعات پیشینی غیرنمونه‌ای اطلاعاتی درباره کمیت مورد بررسی است که غالباً از منابع دیگر گرفته می‌شود. به‌طور مثال، تجربیات قبلی در وضعیت‌های مشابه منبع خوبی از این اطلاعات پیشینی است. اطلاعات پیشینی کمک می‌کند تا با استفاده از آن‌ها یک توزیع پیشینی، که حاوی این اطلاعات است، در نظر گرفته شود. این توزیع پیشینی را می‌توان (درست مانند آن جنبه‌های آماری که یک توزیع برای کمیت تحت بررسی در نظر می‌گیرد) با استفاده از مقادیر گذشته یک پارامتر به‌دست آورد.

فرض کنید یک شرکت پخش دارو قصد خرید یک مسکن جدید را دارد. دو مورد از مهم‌ترین عواملی که بر تصمیم‌گیری شرکت اثر می‌گذارد. عبارت‌اند از: نسبت افرادی که دارو برای آن‌ها مؤثر است ( $\theta_1$ ) و سهمی از بازار که دارو ( $\theta_2$ ) به‌دست خواهد آورد. در این مثال اگرچه هر دو پارامتر  $\theta_1$  و  $\theta_2$  نامعلوم‌اند، اما می‌توان قبل از جمع‌آوری هر داده‌ای یک منبع غنی از اطلاعات درباره این دو پارامتر را از داروهای مسکن مشابه به‌دست آورد. به‌عنوان مثالی دیگر، احتمال ابتلای خفیف به تالاسمی یا تالاسمی مینور فرد O (آن را با  $\omega$  نشان می‌دهیم) را در نظر بگیرید. به فردی که یک ژن تالاسمی را از والدین خود به ارث می‌برد، ناقل ژن می‌گویند و اصطلاحاً گویند این فرد مبتلا به تالاسمی خفیف است. اطلاع از اینکه یکی از والدین