

بلاغ الحزم

درك خطای آماری

سرآغازی برای بیولوژیست‌ها

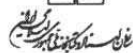
نویسنده

دکتر جیرلینسکی

مترجم

دکتر مهدی ابراهیمزاده





سرشناسه	: گیرلینسکی، مارک Gierliński, Marek
عنوان و نام پدیدآور	: درک خطای آماری، سرآغازی برای بیولوژیست‌ها/ نویسنده مارک گیرلینسکی؛ مترجم مهدی ابراهیم‌زاده.
مشخصات نشر	: تهران: اساطیر پارس؛ چاپار، ۱۳۹۸.
مشخصات ظاهری	: ۲۵۶ ص.: جدول، نمودار.
شابک	: 978-600-8890-64-5
وضعیت فهرست نویسی	: فیپا
یادداشت	: Understanding statistical error : a primer for biologists, 2016.
موضوع	: عنوان اصلی:
موضوع	: آمار پزشکی
موضوع	:
موضوع	: زیست‌سنجی -- داده‌پردازی
موضوع	:
موضوع	: تحلیل واریانس
موضوع	:
موضوع	: Analysis of variance
موضوع	:
موضوع	: Error analysis (Mathematics)
موضوع	:
شناسه افزوده	: ابراهیم‌زاده، مهدی، ۱۳۴۸ - مترجم
رده بندی کنگره	: ۳۸۵۳
رده بندی دیویی	: ۶۱۰/۷۲۷
شماره کتابشناسی ملی	: ۵۸۶۷۴۸۲



تہران: ۸۸۸۹۹۶۸۰ - ۹۱۱۹۶۰۱۴

تهران - خیابان ولی عصر (عج) - بالا

کوچه نوربخش - پلاک ۴۰ - نشر چایا

درک خطای آماری / سرآغازی برای سولوژیست

نویسنده: مارک جیرلینسکی

مترجم: دکتر مهدی ابراهیمزاده

ویراستار فنی و صفحہ آرا: رحیم کبیر صابر

طراح جلد: سرور محمدی بزنج

ناظر فنی: محمد محمدی

چاپ اول ۱۳۹۸

شمارگان ۱۰۰۰ نسخه

ISBN: 978-600-8890-64-5 شابک: ۹۷۸-۶۰۰-۸۸۹۰-۶۴-۵

حق هرگونه چاپ و انتشار محفوظ است.

قیمت: ۴۱.۰۰۰ تومان

فهرست مطالب

۱۳	مقدمه
۱۳	چرا یک دمه را می‌خوانید؟
۱۳	این کتاب درباره چیست؟
۱۴	این کتاب برای چه کسی است؟
۱۴	درباره ریاضیات
۱۵	سپاسگزاری
۱۷	فصل ۱: چرا نیاز به ارزیابی خطا داریم؟
۲۱	فصل ۲: توزیع‌های احتمال
۲۲	۱.۲ متغیرهای تصادفی
۲۲	۲.۲ توزیع احتمال چیست؟
۲۳	توزیع احتمال یک متغیر گسسته
۲۵	توزیع احتمال یک متغیر پیوسته
۲۵	توزیع احتمال تجمعی
۲۶	۳.۲ میانگین، میانه، واریانس و انحراف استاندارد
۲۷	۴.۲ توزیع نرمال
۳۰	مثال: برآورد یک مشاهده دور افتاده
۳۱	۵.۲ قضیه حد مرکزی
۳۳	۶.۲ توزیع لگ نرمال
۳۶	۷.۲ توزیع دو جمله‌ای
۳۹	۸.۲ توزیع پواسن
۴۲	مثال کلاسیک: لگدهای اسب
۴۴	زمان‌های بین دو ورود
۴۶	۹.۲ توزیع t -استیودنت
۴۸	۱۰.۲ تمرینات

فصل ۳: خطاهای اندازه‌گیری..... ۴۹

۱.۳ خطاها از کجا می‌آیند؟..... ۴۹

خطاهای سیستماتیک..... ۵۰

خطاهای تصادفی..... ۵۲

۲.۳ مدل ساده‌ای برای خطاهای تصادفی اندازه‌گیری..... ۵۲

۳.۳ تغییرپذیری ذاتی..... ۵۶

۴.۳ خطای نمونه‌گیری..... ۵۷

نمونه‌گیری در زمان..... ۵۸

۵.۳ خطای ساده اندازه‌گیری..... ۶۰

خطای خواندن..... ۶۰

خطای شمارش..... ۶۲

۶.۳ تمرینات..... ۶۵

فصل ۴: برآوردهای آماری..... ۶۷

۱.۴ جامعه و نمونه..... ۶۷

۲.۴ برآوردهای آماری چیست؟..... ۶۹

۳.۴ آریبی برآوردهای آماری..... ۷۲

۴.۴ برآوردهای آماری که در استفاده رایج‌ترند..... ۷۳

میانگین..... ۷۳

میانگین موزون..... ۷۴

میانگین هندسی..... ۷۶

میانۀ..... ۷۷

انحراف استاندارد..... ۷۸

برآوردهای ناریب انحراف استاندارد..... ۸۰

انحراف میانگین..... ۸۴

ضریب همبستگی پیرسون..... ۸۵

نسبت..... ۸۶

۵.۴ خطای استاندارد..... ۸۷

۶.۴ خطای استاندارد میانگین موزون..... ۹۲

۷.۴ خطای در خطا..... ۹۴

۸.۴ درجه آزادی..... ۹۵

۹.۴ تمرینات..... ۹۶

۹۷.....	فصل ۵: بازه‌های اطمینان
۹۸.....	۱.۵ توزیع نمونه‌گیری
۱۰۱.....	۲.۵ بازه اطمینان: واقعاً به چه معنا است؟
۱۰۲.....	۳.۵ چرا ۹۵٪؟
۱۰۴.....	۴.۵ بازه اطمینان برای میانگین
۱۰۷.....	مثال
۱۰۸.....	۵.۵ خطای استاندارد در مقابل بازه اطمینان
۱۰۸.....	در یک بازه اطمینان چه مقدار خطای استاندارد وجود دارد؟
۱۱۰.....	۱.۱ اطمینان به خطای استاندارد چقدر است؟
۱۱۰.....	۵.۵ بازه اطمینان برای میان
۱۱۳.....	تقریب ساده
۱۱۴.....	مثال
۱۱۵.....	۷.۵ بازه اطمینان برای نسبت
۱۱۸.....	معناداری همبستگی
۱۲۰.....	۸.۵ بازه اطمینان برای نسبت
۱۲۵.....	۹.۵ بازه اطمینان برای داده‌های شمارشی
۱۲۸.....	تقریب ساده
۱۲۹.....	خطاهای داده‌های شمارشی، اعداد صحیح نیستند
۱۲۹.....	۱۰.۵ بوت استرپ‌سازی
۱۳۲.....	۱۱.۵ تکرارها
۱۳۵.....	حجم نمونه برای به دست آوردن میانگین
۱۳۶.....	۱۲.۵ تمرینات
۱۳۹.....	فصل ۶: میله‌های خطا
۱۴۰.....	۱.۶ طراحی یک نمودار خوب
۱۴۱.....	عناصر یک نمودار خوب
۱۴۳.....	خطوط در نمودارها
۱۴۴.....	نقدی بر برجسب‌های نمودار
۱۴۵.....	نمودارهای لگاریتمی
۱۴۷.....	۲.۶ میله‌های خطا در نمودارها
۱۴۸.....	انواع مختلف خطا
۱۴۹.....	چگونگی رسم میله‌های خطا

۱۵۰	نمودار جعبه‌ای
۱۵۲	نمودار نواری
۱۵۷	نمودار کلوچه‌ای
۱۵۸	همپوشانی میله‌های خطا
۱۶۰	۳۶ چه موقع می‌توانید بدون میله‌های خطا باشید؟
۱۶۰	برای متغیر رسته‌ای
۱۶۰	هنگام ارائه داده‌های خام
۱۶۰	گروه‌های بزرگ از نقاط داده‌ها
۱۶۱	وقتی خطاها کوچک و ناچیز هستند
۱۶۱	خطاها در کجا قابل شناسایی نیستند
۱۶۲	۶ نقل قول اسناد و خطاها
۱۶۲	ارقام معنادار
۱۶۳	نوشتن اعداد معناری
۱۶۵	خطاها و ارقام معنار
۱۶۷	خطا با بی‌خطایی
۱۶۸	اعداد رایانه‌ای
۱۷۰	خلاصه
۱۷۰	۵۶ تمرینات
۱۷۳	فصل ۷: تسری خطا
۱۷۴	۱.۷ تسری خطاها چیست؟
۱۷۵	۲.۷ یک متغیره
۱۷۵	تغییر مقیاس
۱۷۶	لگاریتم‌ها
۱۷۷	۳.۷ چند متغیره
۱۷۸	مجموع یا تفاضل
۱۷۹	نسبت یا حاصلضرب
۱۸۱	۴.۷ متغیرهای همبسته
۱۸۲	۵.۷ استفاده یا عدم استفاده از تسری خطا
۱۸۴	۶.۷ مثال: فاصله بین دو نقطه
۱۸۵	۷.۷ مشتق فرمول تسری خطا برای یک متغیره
۱۸۷	۸.۷ مشتق فرمول تسری خطا برای چند متغیره
۱۹۰	۹.۷ تمرینات

۱۹۱	فصل ۸: خطاها در رگرسیون خطی ساده
۱۹۱	۱.۸ رابطه خطی بین دو متغیر
۱۹۲	میانگین پاسخ
۱۹۳	پاسخ صحیح و اختلال
۱۹۴	خطی سازی داده ها
۱۹۵	۲.۸ برازش خط راست
۱۹۷	۳.۸ بازه های اطمینان برای پارامترهای برازش خطی
۲۰۳	مثال
۲۰۵	۴.۸ خطای پیش بینی برای برازش
۲۰۸	۵.۸ رگرسیون از مبدأ
۲۱۰	مثال
۲۱۱	۶.۸ برازش کاه منحنی
۲۱۴	۷.۸ مشتق خطاها برای امتزای برازش
۲۱۶	۸.۸ تمرینات
۲۱۹	فصل ۹: مثال کارشده
۲۱۹	۱.۹ آزمایش
۲۲۱	۲.۹ نتایج
۲۲۱	ساشا
۲۲۵	لوشا
۲۲۸	ماشا
۲۳۱	۳.۹ مباحثه
۲۳۳	۴.۹ پاراگراف نهایی
۲۳۵	حل تمرینات
۲۴۹	پیوست A
۲۵۵	فهرست منابع

مقدمه

چرا یک مقدمه را می‌خوانید؟

مرسوم است که هر کتاب علمی، قبل از شروع مزین به یک «مقدمه» یا «پیش‌گفتار» یا «دیباجه»¹ ناهم اوقات ترکیبی از این موارد باشد. اگر خوش (یا به عبارتی بد) شانس باشید ممکن است یادداشتی از ویراستار، دیباجه‌ای به دنبال پیش‌گفتار ویرایش اول، پیش‌گفتاری برای ویرایش دوم، یک مقدمه کلی را بیابید. در اینجاها قبل از همه می‌توانید بخوانید که نویسنده چه می‌گوید. در ادامه خواهید دید که این کتاب منحصر به فرد است و از همه کتاب‌های دیگر را این موضوع بهتر است. سپس نویسنده سعی می‌کند توصیفی از جزئیات هر فصل ارائه دهد که می‌توان آن را به نوعی در فهرست مطالب پیدا کرد. در نهایت نوبت تقدیرهای اجباری از همه خانواده و دوستان و کسانی است که نویسنده مجبور به عنوان کردن کار بزرگ آنها است. گریزی نیست؛ دیباجه‌ها، پیش‌گفتارها و مقدمه‌ها، همه جا هستند. استانیسلاولم² (Lem 1972) کتابی نوشته است که به‌طور کامل شامل دیباجه‌ها است.

معمولاً مردم از همه این مقدمات عبور می‌کنند زیرا بسته‌کننده، پر از ادعا و خودشیفتگی و بی‌فایده هستند. خوب، آیا هنوز با ما هستید؟ اگر تا اینجا پیش آمده‌اید در واقع یکی از معدود کسانی هستید که مقدمه‌ها را مطالعه می‌کنید. بسیار عالی، بنابراین سعی می‌کنم به اختصار و نه بیش از حد معمول، به نکته اصلی بپردازم.

این کتاب درباره چیست؟

همان‌طور که عنوان کتاب نشان می‌دهد، این کتاب در مورد تحلیل خطا با تأکید بر کاربردهای آن در بیولوژی یا به‌طور کلی‌تر در علوم زیستی است. از زمان رونالد فیشر بزرگ³، آمار بخشی از ذات بیولوژی شده است. تعداد بسیار کمی از نتایج عددی در مطالعات بیولوژیکی و پزشکی را می‌توان بدون تأیید معناداری آماری منتشر کرد. یک راه

1. Stanislaw Lem

2. Ronald Fisher

انجام این کار ارائه p -value برای آزمون آماری -یا به‌طور کلی‌تر- احتمال بروز خطا در یک عبارت خاص است. این از مطالبی است که این کتاب به دنبال آن نیست. روش دیگری برای ارزیابی معناداری یک نتیجه، یافتن خطای ذاتی یا عدم اطمینان آن است. به اعتقاد من ارائه نتیجه عددی بدون هیچ‌گونه عدم اطمینان، بی‌معنا است. بنابراین دانستن نحوه محاسبه خطا خوب است و این آن چیزی است که این کتاب درباره آن است.

در اینجا به جنبه‌های مختلف تحلیل خطای می‌پردازیم: قدری از زمینه‌های نظری و روش‌های عملی محاسبه بازه‌های اطمینان و نیز نمایش گرافیکی میله‌های خطا و نقل قول اعداد همراه با خطا خواهیم گفت. به جای دستورالعمل محاسبات عملی، تأکید را بر شهود و درک آنها قرار می‌دهیم. این حال فرمول‌های دقیق انواع خطاها را نیز ارائه خواهیم داد. مراقب باشید: این یک مثال نامع در مورد آمار نیست؛ بلکه متمرکز بر درک عملی تحلیل عدم اطمینان است. جزئیات بیشتر را می‌توانید درست بعد از مقدمه، در فهرست مطالب بیابید.

این کتاب برای چه کسانی است؟

این کتاب برای بیولوژیست‌های علاقه‌مند به نوشتن شده است که می‌خواهند درک خود از تحلیل داده‌ها را بهبود بخشند. در عین حال که مخاطب اصلی، بیولوژیست‌ها هستند، این کتاب می‌تواند برای هر شخصی که با داده‌های عددی سروکار دارد و علاقمند به یادگیری چگونگی ارزیابی و مقایسه مقادیر اندازه‌گیری شده است، سودمند باشد. اگر انواع خطاهای مختلف را با استفاده از یک بسته نرم‌افزاری محاسبه می‌کنید، از مطمئن نیستید که آنها واقعاً چه معنایی دارند، این کتاب برای شما است. یا شاید کنج‌کاری می‌خواهید چند چیز اساسی در مورد تحلیل عدم اطمینان یاد بگیرید، این کتاب برای شما نیز مفید است.

درباره ریاضیات

علی‌رغم وجود فعالیت‌هایی در ادبیات تحقیق و نوشته‌هایی که روشی کاملاً شهودی را به کار می‌برند (مانند Matulsky 2010)، اعتقاد دارم که آمار بدون ریاضیات، بسیار سخت است. توضیحات واضح نمی‌توانند جایگزین صراحت اکید یک معادله ریاضی شوند. یک مشتق ساده می‌تواند بیان کند که فرمول داده‌شده از کجا آمده است. بنابراین در این کتاب، ریاضیات وجود دارد. خیلی پیچیده و وسیع نیست، اما ریاضیات وجود دارد. نیازی به گفتن نیست که در کاربردهای عملی، معادلات مورد نیاز هستند، بنابراین اگر نیاز به یافتن

عدم اطمینان خاصی هستید که در نرم‌افزارهای آماری متداول، ارائه نشده است، می‌توانید از معادلات به کار رفته در این کتاب استفاده کنید. به سادگی می‌توان این معادلات را در هر زبان برنامه‌نویسی یا حتی صفحه گسترده‌های رایانه‌ای کدنویسی کرد. ریاضیات موجود در این کتاب کاملاً مقدماتی است و در واقع فراتر از سطح دبیرستان نیست. بیشتر معادلات شامل مجموع‌ها و جبر ساده است. پیشرفته‌ترین عملگر به کار رفته، مشتق است.

قصد از دست دادن خوانندگان بالقوه را ندارم. این یک کتاب درسی ریاضیات نیست. من سعی می‌کنم معادلات را فقط در صورت لزوم استفاده کنم و تلاش می‌کنم همراه با توضیحات شهودی باشند. اغلب برای بیان معنای یک مفهوم یا فرمول، نتایج شبیه‌سازی رایانه‌ای ارائه می‌گردد. همچنین برای ساده‌سازی و تقریب در برخی موارد، توانایی برای اصلاحات پراخته‌ایم. امیدوارم این کار باعث شده باشد که ریاضیات این کتاب ساده‌تر شود. در خاتمه ذکر این احاطه لازم است. این کتاب در درجه اول برای بیولوژیست‌ها نوشته شده است نه برای ریاضیدانان و فیزیکدانان. بنابراین هیچ اثبات ریاضی وجود ندارد. برخی از نتایج، دقیق نبوده و به‌طور کلی یک عدم سخت‌گیری در مورد ریاضی وجود دارد. یک ریاضیدان ممکن است از محتوای این کتاب خشنود نگردد، بنابراین اگر شما یکی از آنها هستید، لطفاً از هم اکنون چشم‌ها باندازید.

سپاسگزاری

از پرفسور آنگوس لاموند^۱ تشکر می‌کنم که به دست از ابتدا تا انتهای مطالب نگارش شده را خوانده و نظرات بسیار ارزشمندی ارائه نمود. به عنوان یک بیولوژیست کمک کرد تا مخاطب (یعنی شما!) را بهتر درک کنم. همچنین در زبان انیسی که زبان اول من نیست، مرا یاری نمود.