

دانشگاه یزد

دانشکده ریاضی

گروه آمار

پایان نامه برای دریافت درجه کارشناسی ارشد

آمار

**استنباط استوار در مدل رگرسیون لجستیک**

استاد راهنما:

دکتر حجت اله ذاکرزاده

استاد مشاور:

دکتر علی دولتی

پژوهش و نگارش:

شهره مسطقی

مهر ۹۰

## عنوان: استنباط استوار در مدل رگرسیون لجستیک

### چکیده:

یکی از مدل‌های خطی تعمیم‌یافته‌ی مهم، مدل رگرسیون لجستیک است. به دلیل غیرخطی بودن معادلات برآوردیابی نسبت به پارامترها، معمولاً از روش‌های عددی تکراری برای پیدا کردن برآوردها استفاده می‌شود. برآوردهای حاصل به طور معمول در برابر مشاهدات دورافتاده (نقاط پرت) حساس هستند. برای حل این مشکل، روش‌هایی برای پیدا کردن برآوردهای استوار ارائه شده است.

در این پایان‌نامه دو روش پیدا کردن برآوردهای استوار مورد بررسی قرار می‌گیرد. یکی از روش‌ها، براساس مینیمم‌سازی بعضی از تابع‌های درست‌نمایی و روش دیگر برپایه تابع تأثیر است. از این روش‌ها برای ساختن آزمون فرض‌های استوار مربوط به پارامترهای مدل رگرسیون لجستیک، تحلیل لجستیک ممیزی استفاده می‌شود. با استفاده از داده‌های واقعی و همچنین شبیه‌سازی، مفید بودن روش‌های مطالعه شده، نشان داده می‌شود.

### **Title: Robust inference in the logistic regression model**

Logistic regression model is one of the most important general linear models. Because of the non-linearity of estimating equations, with respect to the parameters, the iterative numerical methods usually used to find the estimators. The resulting estimators are sensitive, in presence of outliers. Some robust methods are proposed in literature, for solving this problem. In this thesis two approaches for finding robust estimator are discussed. One of the methods is based on functional of likelihood

process and the other method is based on influence function. These methods will be used to construct robust hypotheses testing problems in logistic regression and discriminant logistic analysis. Real data and simulated data will be used to illustrate the usefulness of the presented methods.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

دانشگاه یزد

دانشکده ریاضی

گروه آمار

پایان نامه

برای دریافت درجه کارشناسی ارشد

آمار ریاضی

استنباط استوار در مدل رگرسیون لجستیک

استاد راهنما: دکتر حجت اله ذاکرزاده

استاد مشاور: دکتر علی دولتی

پژوهش و نگارش: شهره مسقطی

مهر ماه ۱۳۹۰

اثر معنوی این پژوهش، هدیه‌ی ناپختاری است

به ساحت مقدس ساقی لب تشنگان

حضرت ابوالفضل العباس

و پدر و مادر عزیزم



سپاس خدای عزیز را که به این کمترین توفیق کسب اندک دانشی از دریای بیکران علم راعطا فرمود. امید است که در سایه الطاف بی انتهایش سعادت عمل به آموخته‌های خویش را بیابم.

از استاد علم و اخلاق، جناب آقای دکتر حجت‌اله ذاکر زاده که بی‌شک حمایت‌ها و رهنمودهای ایشان همچون خورشیدی روشن مسیر پرتلاطم پژوهش‌های اینجانب در به‌ثمر رسیدن این پایان‌نامه بوده است، سپاسگزارم. از خداوند بزرگ توفیق علم آموزی برای سایستگان محضر ایشان را خواستارم.

از جناب آقای دکتر علی دولتی که در امر مشاوره این پروژه مساعدت نمودند، کمال تشکر را دارم و برای ایشان از خداوند سلامت و سعادت را خواهانم.

سپاسگزارم از جناب آقای دکتر عیسی محمودی که داوری این پایان‌نامه را پذیرفته‌اند، همچنین از آقای دکتر حسین فلاح زاده که در مطالعه و داوری تحقیق اینجانب قبول زحمت نموده‌اند، کمال سپاسگزاری را دارم.

سپاسگزارم از استاد بزرگوارم جناب آقای دکتر حمزه ترابی که پیش از این مصحح به‌شمار گردید در خدمتشان بوده‌ام. در پایان لازم می‌دانم از پدر و مادر عزیزم که پیوسته یاری‌میکردم بودند و هر لحظه‌ی تلاشم بافداکاری آن‌ها میسر گشته کمال تشکر را داشته باشم.





## چکیده

یکی از مدل‌های خطی تعمیم‌یافته‌ی مهم، مدل رگرسیون لجستیک است. به دلیل غیرخطی بودن معادلات برآوردیابی نسبت به پارامترها، معمولاً از روش‌های عددی تکراری برای پیدا کردن برآوردها استفاده می‌شود. برآوردهای حاصل به طور معمول در برابر مشاهدات دورافتاده (نقاط پرت) حساس هستند. برای حل این مشکل، روش‌هایی برای پیدا کردن برآوردهای استوار ارائه شده است. در این پایان‌نامه دو روش پیدا کردن برآوردهای استوار مورد بررسی قرار می‌گیرد. یکی از روش‌ها، براساس مینیمم‌سازی بعضی از تابع‌های درست‌نمایی و روش دیگر برپایه تابع تأثیر است. از این روش‌ها برای ساختن آزمون فرض‌های استوار مربوط به پارامترهای مدل رگرسیون لجستیک، تحلیل لجستیک ممیزی استفاده می‌شود. با استفاده از داده‌های واقعی و همچنین شبیه‌سازی، مفید بودن روش‌های مطالعه شده، نشان داده می‌شود.



# فهرست مطالب

۱	مفاهیم و تعاریف اولیه	۱
۱	۱.۱ مفاهیمی از آمار ریاضی	۱
۲	۱.۱.۱ روش ماکسیمم درستنمایی	۲
۲	۲.۱.۱ برآورد ماکسیمم درستنمایی	۲
۲	۳.۱.۱ انواع همگرایی	۲
۳	۴.۱.۱ انتگرال پذیری یکنواخت	۳
۴	۵.۱.۱ پیوستگی مطلق	۴
۴	۲.۱ مفاهیمی از آنالیز ریاضی	۴
۶	۳.۱ مفاهیمی از جبرخطی	۶
۸	۴.۱ مفهوم هم‌ارزی و رابطه‌ی آن با نسبت درستنمایی	۸
۹	۱.۴.۱ لم سوم له‌کام	۹
۱۰	۵.۱ نظریه ارگودیک	۱۰
۱۱	۲ برآورد استوار در مدل رگرسیون لجستیک	۱۱
۱۲	۱.۲ الگوهای رگرسیون خطی و غیرخطی	۱۲
۱۳	۱.۱.۲ برآورد پارامتر با استفاده از کمترین مربعات وزنی در رگرسیون خطی	۱۳
۱۵	۲.۱.۲ کمترین مربعات وزنی	۱۵
۱۵	۳.۱.۲ رگرسیون لجستیک	۱۵
۱۶	۴.۱.۲ برآورد پارامتر با استفاده از ماکسیمم درستنمایی	۱۶

۲۰۲	آشنایی با مفاهیم استواری	۱۷
۱۰۲.۲	میزان استواری	۲۲
۳۰۲	برآوردهای استوار	۲۶
۱۰۳.۲	مجانب‌های درست‌نمایی	۲۷
۲۰۳.۲	برآوردگر $M$ با استفاده از تابع تأثیر	۳۱
۴۰۲	برآوردهای استوار در مدل رگرسیون لجستیک با استفاده از مینیمم کردن تابع‌های	
	درست‌نمایی	۳۴
۱۰۴.۲	معرفی	۳۴
۲۰۴.۲	وجود برآوردگر	۳۶
۳۰۴.۲	برآوردگر میانه استوار در رگرسیون لجستیک	۳۹
۳	لجستیک ممیزی با استفاده از برآوردهای استوار: روش تابع تأثیر	۵۱
۱.۳	مقدمه	۵۱
۲.۳	مدل لجستیک ممیزی	۵۲
۳.۳	برآوردگر استوار در رگرسیون لجستیک	۵۳
۴.۳	لجستیک ممیزی و میزان خطا	۵۵
۵.۳	تابع تأثیر در میزان خطا	۵۶
۶.۳	کارایی‌های مجانبی طبقه‌بندی شده	۶۱
۷.۳	تحلیل ممیزی فیشر	۶۲
۴	آزمون استوار در رگرسیون لجستیک	۷۳
۱.۴	مقدمه	۷۳
۲.۴	آزمون‌ها برای فرض‌های خطی	۷۴
۳.۴	برآوردگر استوار و ویژگی‌های مجانبی آن	۷۸
۴.۴	آماره آزمون و رفتار مجانبی آن	۸۳
۵.۴	وزن‌های استفاده شده در برآوردگر وزنی بیانکو-یوهای	۸۶

۶.۴ داده‌های واقعی ..... ۸۶

۷.۴ مطالعه‌ی شبیه‌سازی ..... ۹۰

### الف برنامه‌های کامپیوتری

الف ۱. داده‌های واسو ..... ۹۳

الف ۲. برنامه محاسبه برآوردگر وزنی بیانکو-یوهای برای وزن دومربعی ..... ۹۴

الف ۳. برنامه محاسبه وزن هابر برای برآوردگر بیانکو-یوهای ..... ۱۰۳

الف ۴. برنامه محاسبه وزن کراکس و هاس بروک برای برآوردگر بیانکو-یوهای ..... ۱۰۳

الف ۵. برنامه محاسبه برآوردگرهای ماکزیمم درستنمایی و وزنی بیانکو-یوهای با وزن دو مربعی

جدول ۱.۴ ..... ۱۰۳

الف ۶. برنامه محاسبه  $p$ -مقدارهای نمودار ۲.۴ ..... ۱۰۴

الف ۷. برنامه شبیه‌سازی ..... ۱۰۵

الف ۸. برنامه شبیه‌سازی با اغتشاش ..... ۱۰۷

### ب واژه‌نامه‌ی فارسی به انگلیسی

پ واژه‌نامه‌ی انگلیسی به فارسی ..... ۱۱۷

کتاب‌نامه ..... ۱۲۱



## فهرست شکل‌ها

۵۸	.....	تحلیل ممیزی میزان خطا براساس توزیع مدل تحت اغتشاش $\varepsilon$	۱.۳
۸۸	.....	داده‌های واسو	۱.۴
۸۹	.....	حساسیت $p$ -مقدار	۲.۴





## فهرست جدول‌ها

۸۷	.....	برآورد پارامترها و $p$ -مقدار برای داده‌های واسو.	۱.۴
۹۰	.....	فراوانی مشاهده شده تعداد آزمون‌های رد شده	۲.۴
۹۱	.....	فراوانی مشاهده شده تعداد آزمون‌های رد شده با اغتشاش	۳.۴



## فهرست نمادها

$\triangleleft$ .....	هم‌ارزی دو اندازه احتمال
$\ll_{a.s}$ .....	پیوسته مطلق مجانبی
$a \vee b$ .....	ماکسیمم مقدار بین $a$ و $b$
$a \wedge b$ .....	مینیمم مقدار بین $a$ و $b$
$\xrightarrow{p}$ .....	همگرایی در احتمال
$Bin(.,.)$ .....	توزیع دو جمله‌ای
$\Phi(.)$ .....	تابع توزیع نرمال استاندارد
$\xrightarrow{d}$ .....	همگرایی در توزیع