به نام خداوند جان و خرد

انتشارات
۸۴۶

اقتصادسنجی داده‌های مقطعی و طبقه‌بندی‌شده

با نرم‌افزار STATA (تئوری و کاربرد)

ویرایش دوم

دکتر حسین محمّدی
استاد دانشگاه فردوسی مشهد
دکتر مرتضی محمّدی
عضو هیئت علمی دانشگاه حکیم سبزواری

سرشناسه: محمدی، حسین، ۱۳۵۷ -
عنوان و نام پدیدآور: اقتصادسنجی داده‌های مقطعی و طبقه‌بندی شده با نرم‌افزار STATA (تئوری و کاربرد)/ حسین محمدی، مرتضی محمدی؛ ویراستار ادبی هانیه اسدپور فعال مشهد.
وضعیت ویراست: ویراست ۲.
مشخصات نشر: مشهد: دانشگاه فردوسی مشهد، انتشارات، ۱۴۰۴.
مشخصات ظاهری: ۳۷۴ ص.: جدول.
فروست: انتشارات دانشگاه فردوسی مشهد؛ ۸۴۶.
شابک: ISBN: 978-964-386-669-3
وضعیت فهرست‌نویسی: فیپا.
یادداشت: چاپ دوم
موضوع: نرم‌افزار استتا — Stata
اقتصادسنجی -- برنامه‌های کامپیوتری — Econometrics -- Computer programs
شناسه افزوده: محمدی، مرتضی، ۱۳۵۸- — Mohammadi, Morteza, 1979-
شناسه افزوده:
شناسه افزوده: دانشگاه فردوسی مشهد، انتشارات.
رده‌بندی کنگره: HB۱۳۹
رده‌بندی دیویی: ۳۳۰/۰۱۵۱۹۵
شمارهٔ کتابشناسی ملّی: ۱۰۳۵۴۳۴۹

اقتصادسنجی داده‌های مقطعی و طبقه‌بندی‌شده با نرم‌افزار STATA

(تئوری و کاربرد) — ویرایش دوم

پدیدآورندگان: دکتر حسین محمّدی؛ دکتر مرتضی محمّدی
ویراستار ادبی: هانیه اسدپور فعال مشهد
مشخصات: وزیری، ۲۰۰ نسخه، چاپ دوم، زمستان ۱۴۰۴ (اوّل، ۱۴۰۱)
چاپ و صحافی: همیار
بها : ۶/۰۰۰/۰۰۰ ریال
حق چاپ برای انتشارات دانشگاه فردوسی مشهد محفوظ است.

انتشارات
۸۴۶

مراکز پخش:

فروشگاه و نمایشگاه کتاب پردیس: مشهد، میدان آزادی، دانشگاه فردوسی مشهد، جنب سلف یاس
تلفن: ۳۸۸۰۲۶۶۶- ۳۸۸۳۳۷۲۷ (۰۵۱)
مؤسّسهٔ کتابیران: تهران، میدان انقلاب، خیابان کارگر جنوبی، بین روانمهر و وحید نظری، بن‌بست گشتاسب، پلاک ۸ تلفن: ۶۶۴۸۴۷۱۵ (۰۲۱)
مؤسّسهٔ دانشیران: تهران، خیابان انقلاب، خیابان منیری جاوید (اردیبهشت) نبش خیابان نظری، شمارهٔ ۱۴۲
تلفکس: ۶۶۴۰۰۲۲۰ – ۶۶۴۰۰۱۴۴ (۰۲۱)

http://press.um.ac.ir Email: press@um.ac.ir

فهرست مطالب

مقدّمه ۹
متغیّرهای دوحالتی ۹
متغیّرهای ترتیبی ۹
متغیّرهای اسمی ۹
متغیّرهای سانسورشده ۱۰
متغیّرهای شمارشی ۱۰

فصل ۱. مروری بر مفاهیم آمار ۱۳
۱–۱ مقدّمه ۱۳
۱–۲ پارامتر و آماره ۱۴
۱–۳ انواع متغیّرها ۱۴
۱–۴ تابع احتمال ۱۶
۱–۴–۱– تابع احتمال گسسته ۱۶
۱–۴–۲ تابع چگالی احتمال پیوسته ۱۷
۱–۵ تابع توزیع تجمّعی ۱۷
۱–۶ هم‌بستگی بین متغیّرها ۱۸
۱–۶–۱ کوواریانس ۱۹
۱–۶–۲ ضریب هم‌بستگی ۱۹
۱–۶–۳ ضریب تعیین ۲۰
۱–۷ توزیع‌های آماری ۲۱
۱–۸ توزیع‌های گسسته ۲۱
۱–۸–۱ توزیع یکنواخت گسسته ۲۱
۱–۸–۲ توزیع برنولی ۲۳
۱–۸–۳ توزیع دوجمله ای ۲۳
۱–۸–۴ توزیع چندجمله ای ۲۴
۱–۸–۵ توزیع دوجمله‌ای منفی (پاسکال) ۲۴
۱–۸–۶ توزیع هندسی ۲۴

۱–۸–۷ توزیع فوق هندسی ۲۵
۱–۸–۸ توزیع پوآسن ۲۵
۱–۹ توزیع‌های پیوسته ۲۶
۱–۹–۱ توزیع یکنواخت پیوسته ۲۶
۱–۹–۲ توزیع نمایی ۲۷
۱–۹–۳ توزیع گاما ۲۸
۱–۹–۴ توزیع نرمال ۲۹
۱–۹–۵ توزیع نرمال استاندارد ۳۰
۱–۹–۶ توزیع کای اسکور (چی دو، کای دو، مربع کای) ۳۲
۱–۹–۷ توزیع t استیودنت ۳۴
۱–۹–۸ توزیع فیشر ۳۵
۱–۱۰ استنباط آماری ۳۶
۱–۱۰–۱ برآورد نقطه ای ۳۷
۱–۱۰–۲ برآورد فاصله‌ای و فاصلهٔ اطمینان ۴۱
۱–۱۱ معرفی نرم‌افزار STATA ۴۳
۱–۱۱–۱ محیط کلی نرم‌افزار استاتا ۴۳
۱–۱۱–۲ شیوهٔ ورود و ذخیرهٔ داده‌ها ۴۵

فصل ۲. مروری بر مدل‌های رگرسیون خطّی و روش‌های برآورد ۵۹
۲–۱ مقدّمه ۵۹
۲–۲ مروری بر مدل رگرسیون خطّی ۵۹
۲–۳ مدل‌های خطّی و غیرخطّی ۶۱
۲–۴ تفسیر ضرایب مدل رگرسیون ۶۴
۲–۵ برآورد با روش حداقل مربعات معمولی ۶۵
۲–۶ مدل‌های رگرسیون خطّی– غیرخطّی ۶۶
۲–۷ نقض فروض ۶۷
۲–۷–۱ میانگین شرطی غیر صفر برای عنصر خطای ε ۶۸
۲–۷–۲ وجود هم‌بستگی بین عناصر خطای ε و متغیّرهای توضیحی x ۶۸
۲–۸ برآوردکنندهٔ حداکثر درست‌نمایی ۶۹
۲–۸–۱ واریانس برآوردکنندهٔ ML ۷۱
۲–۹ آزمون فرضیه ۷۲
۲–۹–۱ آزمون‌های والد، نسبت درست‌نمایی و ضریب لاگرانژ ۷۳

۲-۹-۲ آزمون والد ۷۶
۲-۹-۳ آزمون نسبت درست‌نمایی ۸۰
۲-۹-۴ مقایسهٔ آزمون‌های والد و LR ۸۴
۲-۱۰ معیارهای خوبی برازش ۸۵
۲-۱۰-۱ ضریب تعیین یا R^2 در LRM ۸۵
۲-۱۰-۲ R^2 ساختگی برمبنای R^2 در LRM ۸۶
۲-۱۰-۳ معیارهای اطلاعات ۹۰
۲-۱۱ برآورد مدل رگرسیون روی متغیّر وابستهٔ ساختگی در STATA ۹۲

فصل ۳. پیامدهای دوتایی: الگوی لاجیت و پروبیت دوگانه ۱۰۷
۳-۱ مقدّمه ۱۰۷
۳-۲ مدل احتمال خطّی LPM ۱۰۸
۳-۳ مدل لاجیت ۱۱۲
۳-۴ مدل متغیّر پنهان برای متغیّرهای دوحالتی ۱۱۳
۳-۵ شناسایی ۱۱۸
۳-۵-۱ شناسایی احتمالات ۱۲۰
۳-۶ مدل C Log-log و مدل Log-log ۱۲۱
۳-۷ برآورد حداکثر درست‌نمایی ۱۲۳
۳-۸ اثرات پارامترها ۱۲۵
۳-۹ روش‌های تفسیر ۱۲۶
۳-۹-۱ روش تغییر جزئی در احتمالات پیش‌بینی شده یا روش اثرات نهایی ۱۲۶
۳-۹-۲ روش تغییر گسسته ۱۲۷
۳-۹-۳ تفسیر برمبنای نسبت‌های برتری ۱۲۹
۳-۱۰ برآورد مدل‌های دوحالتی با STATA ۱۳۱

فصل ۴. پیامدهای ترتیبی ۱۵۳
۴-۱ مقدّمه ۱۵۳
۴-۲ مدل متغیّر پنهان برای متغیّرهای ترتیبی ۱۵۴
۴-۲-۱ فروض مربوط به توزیع ۱۵۷
۴-۲-۲ احتمال مقادیر مشاهده شده ۱۵۸
۴-۳ شناسایی ۱۶۰
۴-۴ شیوهٔ برآورد مدل ۱۶۱

۴-۵ روش‌های تفسیر ۱۶۲
۴-۵-۱ روش تغییر جزئی در احتمالات پیش بینی شده یا روش اثرات نهایی ۱۶۲
۴-۵-۲ روش تغییر گسسته ۱۶۳
۴-۵-۳ مدل‌سازی نسبت برتری در مدل لاجیت ترتیبی ۱۶۴
۴-۶ فرضیهٔ رگرسیون موازی ۱۶۶
۴-۷ برخی مدل‌های مرتبط با داده‌های ترتیبی ۱۷۰
۴-۷-۱ مدل لاجیت ترتیبی تعمیم یافته ۱۷۰
۴-۷-۲ مدل رگرسیون گروهی ۱۷۰
۴-۷-۳ سایر مدل‌ها برای داده‌های ترتیبی ۱۷۱
۴-۸ برآورد مدل لاجیت ترتیبی با STATA ۱۷۲
۴-۹ برخی مطالعات کاربردی از مدل لاجیت ترتیبی ۲۰۳

فصل ۵. پیامدهای اسمی: لاجیت چندگانه و مدل‌های مرتبط ۲۰۷
۵-۱ مقدّمه ۲۰۷
۵-۲ برآورد هم‌زمان لاجیت‌های دوگانه ۲۰۸
۵-۳ مدل‌سازی لاجیت چندگانه ۲۱۰
۵-۳-۱ MNLM به‌عنوان یک مدل احتمال ۲۱۱
۵-۳-۲ MNLM به‌عنوان یک مدل نسبت برتری ۲۱۲
۵-۳-۳ MNLM به‌عنوان یک مدل انتخاب گسسته ۲۱۳
۵-۴ برآورد حداکثر درست‌نمایی ۲۱۵
۵-۵ محاسبه و آزمون سایر کنتراست‌ها ۲۱۵
۵-۶ دو آزمون مفید ۲۱۷
۵-۶-۱ آزمون اینکه یک متغیّر مستقل اثرگذار نیست (اثر صفر دارد). ۲۱۸
۵-۶-۲ آزمون ترکیب طبقات متغیّر وابسته ۲۱۹
۵-۷ تفسیر ۲۲۰
۵-۷-۱ تغییر جزئی (اثرات نهایی) ۲۲۱
۵-۷-۲ تغییر گسسته ۲۲۱
۵-۷-۳ تفسیر براساس نسبت‌های برتری ۲۲۲
۵-۸ مدل لاجیت شرطی ۲۲۳
۵-۹ استقلال گزینه‌های نامرتبط ۲۲۵
۵-۹-۱ آزمون IIA ۲۲۶
۵-۱۰ سایر مدل‌های مرتبط ۲۲۷

۵-۱۰-۱ مدل لاجیت متداخل ۲۲۷
۵-۱۰-۲ مدل پروبیت چندگانه ۲۳۰
۵-۱۰-۳ مدل لاجیت ترکیبی ۲۳۱
۵-۱۱ برآورد مدل‌های با پیامدهای اسمی در STATA ۲۵۲
۵-۱۲ مروری بر چند مقالهٔ کاربردی ۲۸۲

فصل ۶. پیامدهای محدودشده و مدل توبیت ۲۸۵
۶-۱ مقدّمه ۲۸۵
۶-۲ مسئلهٔ سانسور ۲۸۷
۶-۳ مدل توبیت برای پیامدهای سانسورشده ۲۸۹
۶-۳-۱ توزیع سانسور ۲۹۰
۶-۳-۲ مشکلات ایجادشده با سانسور ۲۹۱
۶-۴ روش‌های برآورد ۲۹۳
۶-۴-۱ برآورد به روش حداکثر درست‌نمایی ۲۹۳
۶-۴-۲ روش هکمن ۲۹۴
۶-۴-۳ نقض فروض ۲۹۶
۶-۵ روش‌های تفسیر ۳۰۲
۶-۵-۱ تغییر در پیامد پنهان ۳۰۲
۶-۵-۲ تغییر در پیامد بریده شده ۳۰۳
۶-۵-۳ تغییر در پیامد سانسورشده ۳۰۴
۶-۶ تعمیم مدل توبیت ۳۰۴
۶-۶-۱ سانسور از بالا ۳۰۵
۶-۶-۲ سانسور از بالا و پایین ۳۰۵
۶-۶-۳ مدل رگرسیون بریده (کوتاه) شده ۳۰۷
۶-۷ برآورد الگوی توبیت با STATA ۳۰۸
۶-۸ مروری بر چند مطالعهٔ کاربردی ۳۲۳

فصل ۷. مدل‌های رگرسیون برای پیامدهای شمارشی ۳۲۵
۷-۱ مقدّمه ۳۲۵
۷-۲ توزیع پوآسن ۳۲۶
۷-۳ مدل رگرسیون پوآسن ۳۲۷
۷-۳-۱ روش برآورد ۳۲۷

۷–۳–۲ روش تفسیر ۳۲۸
۷–۳–۳ الگوی پوآسن تعمیم‌یافته ۳۳۱
۷–۴ مدل رگرسیون دوجمله‌ای منفی ۳۳۱
۷–۴–۱ ناهمگنی و پدیدهٔ سرایت ۳۳۴
۷–۴–۲ برآورد مدل NBRM ۳۳۵
۷–۴–۳ آزمون پراکندگی بیش ازحد ۳۳۵
۷–۴–۴ روش تفسیر ۳۳۶
۷–۴–۵ سایر مدل‌های مرتبط ۳۳۶
۷–۵ مدل‌هایی برای شمارش‌های بریده شده ۳۳۷
۷–۵–۱ برآورد مدل ۳۳۸
۷–۵–۲ تفسیر ۳۳۹
۷–۶ مدل‌های شمارشی با صفر تغییریافته ۳۳۹
۷–۶–۱ مدل همراه با صفر ۳۳۹
۷–۶–۲ مدل‌های با صفر زیاد ۳۴۰
۷–۷ کار با داده‌های شمارشی در STATA ۳۴۳
۷–۸ مروری بر چند مطالعهٔ کاربردی ۳۶۵

منابع ۳۶۹
نمایه ۳۷۳

مقدّمه

مدل رگرسیون خطّی از پرکاربردترین روش‌های آماری در علوم اجتماعی و اقتصادی است که کتاب‌های زیادی این مدل را توصیف کرده‌اند و هزاران کاربرد از آن را در رشته‌های مختلف می‌توان یافت. در مدل رگرسیون خطّی عمدتاً فرض بر آن است که متغیّر وابسته به‌صورت پیوسته بوده و در تمام نمونه قابل‌مشاهده و اندازه‌گیری است، اما موارد متعددی وجود دارد که متغیّر وابسته در علوم اجتماعی و اقتصادی، پیوسته نبوده است و یا در برخی موارد قابل‌مشاهده نیست. در این کتاب مدل‌های رگرسیونی موردبحث قرار می‌گیرند که در آن‌ها متغیّر وابسته به‌صورت دوتایی، ترتیبی، اسمی، شمارشی، سانسور شده و یا بریده شده است. به این قبیل متغیّرها در این کتاب متغیّرهای وابستهٔ محدود و یا طبقه‌بندی‌شده[1] (CLDV) گفته می‌شود. مروری بر مطالعات علوم اجتماعی و اقتصادی نشان می‌دهد که CLDVها بسیار رایج هستند. برخی از نمونه‌های مدل‌های CLDV در مطالعات علوم اقتصادی و اجتماعی شامل موارد زیرند:

متغیّرهای دوحالتی[2]: این متغیّرها دوطبقه دارند و عمدتاً نشان‌دهندهٔ وقوع یا عدم وقوع یک اتفاق و یا نشان‌دهندهٔ وجود یا عدم وجود یک خصوصیت‌اند. برای مثال، آیا یک کشاورز، محصول خود را بیمه کرده است یا خیر؟ آیا یک شهروند در آخرین انتخابات رأی داده است؟ آیا یک دانش‌آموز دبیرستان تصمیم به ورود به دانشگاه دارد؟ آیا یک مصرف‌کننده همان برند قدیمی را می‌خرد یا یک برند جدید را امتحان می‌کند؟ آیا یک فرد به سؤال پرسش‌نامه پاسخ می‌دهد؟ و نظایر آن.

متغیّرهای ترتیبی[3]: این متغیّرها دارای طبقاتی هستند که قابل‌رتبه‌بندی‌اند. در یکسری مطالعات عمدتاً از پاسخ دهندگان خواسته می‌شود تا موافقت یا نظر خود را با استفاده از گزینه‌های کاملاً موافق، موافق، مخالف یا کاملاً مخالف رتبه‌بندی کنند. در برخی دیگر از مطالعات، فراوانی وقوع یک اتفاق را با استفاده از طبقات غالباً، گاهی اوقات، به ندرت و هرگز رتبه‌بندی می‌کنند. مهارت‌های تحصیلی نیز می‌توانند برحسب بالاترین مدرک تحصیلی اخذشده به دکتری، کارشناسی ارشد، کارشناسی، دیپلم، زیردیپلم طبقه‌بندی شوند. همهٔ این موارد مثال‌هایی از متغیّرهایی هستند که به‌صورت مرتب یا ترتیبی بیان شده‌اند.

متغیّرهای اسمی[4]: زمانی که چند پیامد وجود دارد که نمی‌توان آن‌ها را رتبه‌بندی کرد، از متغیّرهای اسمی یا چندگانه استفاده می‌شود. برای مثال، مشاغل می‌توانند به‌صورت تجاری، کارگری، کارمندی و حرفه‌ای

1. Categorical and limited dependent variables (CLDV)
2. Binary variables
3. Ordered variables
4. Nomial variables

تقسیم‌بندی شوند. همچنین وضعیت ازدواج را می‌توان به شکل مجرّد، متأهّل، مطلّقه و همسر فوت شده طبقه‌بندی کرد. در مطالعات مربوط به ترجیحات برند یا نام تجاری نیز می‌توان طبقات را در گروه‌های غیرترتیبی بررسی کرد. همهٔ این موارد از مثال‌هایی هستند که در آن‌ها متغیّر موردبررسی به‌صورت اسمی یا چندگانه است و هیچ نظم و ترتیبی بین طبقات برقرار نیست.

متغیّرهای سانسورشده[1]: وقتی مقدار یک متغیّر در طول یک محدودهٔ مشخص از دامنهٔ آن متغیّر غیرمشخص باشد، با متغیّرهای سانسورشده مواجه هستیم. مثال مشخصی از این مورد، مخارج روی کالاهای بادوام است؛ زیرا افرادی که درآمد قابل‌تصرّف آن‌ها کمتر از قیمت ارزان ترین کالای بادوام باشد، الزاماً مخارجشان روی کالاهای بادوامِ مذکور صفر خواهد بود یا برای مثال، دستمزد ساعتی کارگران توسط نرخ حداقل دستمزد تعیین شده توسط اتّحادیه‌ها محدود خواهد شد. همچنین متغیّرهایی که به‌صورت درصدی بیان می‌شوند، مانند درصد خانه‌های تخریب شده در یک زلزله، بین اعداد صفر تا ۱۰۰ درصد هستند و ازاین‌رو، در اعداد زیر صفر و بالای صد، سانسور شده‌اند. سانسورسازی می‌تواند به دلایل روش‌شناسی مطالعه نیز صورت پذیرد.

متغیّرهای شمارشی[2]: این متغیّرها نشان‌دهندهٔ تعداد دفعاتی‌اند که یک واقعه اتفاق می‌افتد. برای مثال، تعداد دفعاتی که یک فرد در سال به پزشک مراجعه می‌کند یا تعداد مشاغلی که یک فرد دارد. تعداد اعتصابات در سال، تعداد انتشارات علمی یک محقّق در سال، تعداد اعتراضات سیاسی در جامعه، تعداد فرزندان یک خانواده، تعداد سال‌های تحصیل، تعداد دفعات خرید و نظایر آن همگی جزء متغیّرهای شمارشی به حساب می‌آیند.

قابل ذکر است که سطح و شیوهٔ اندازه‌گیری یک متغیّر همیشه مشخص و بدون ابهام نیست. به‌همین‌دلیل، ممکن است با شیوهٔ اندازه‌گیری متغیّرها در برخی مثال‌هایی که در قبل مورداشاره قرار گرفت، موافق نباشید. شیوهٔ اندازه‌گیری و درجه‌بندی متغیّرهای تحت بررسی به موارد مختلفی ازجمله تئوری موضوع، ماهیت متغیّر، هدف تحقیق، فرضیات تحقیق و نظایر آن بستگی دارد. به‌عنوان نمونه، متغیّر آموزش و تحصیلات را می‌توان به‌عنوان یک متغیّر دوحالتی در نظر گرفت که در آن افراد با آموزش دبیرستان و بالاتر عدد ۱ و افرادی با تحصیلات پایین‌تر عدد صفر به خود بگیرند یا اینکه می‌توان آموزش را به‌عنوان یک متغیّر ترتیبی در نظر گرفت که نشان‌دهندهٔ بالاترین مدرک تحصیلی اخذشده است و شامل طبقات بی‌سواد، کمتر از دیپلم، دیپلم، کارشناسی، کارشناسی ارشد و دکتری است. حتّی می‌توان متغیّر تحصیلات را به‌عنوان یک متغیّر شمارشی در نظر گرفت که در آن متغیّر وابسته یا تحصیلات، نشان‌دهندهٔ تعداد سال‌های تحصیل فرد است. هرکدام از طبقه بندی‌های مورداشاره برای متغیّر آموزش می‌تواند بسته به هدف تحقیق و اطلاعات موجود، منطقی و مناسب باشد. وقتی طبقات متغیّر وابسته مشخص شد، لازم است

1. Censored variable
2. Count variable

از مدل مناسب برای برآورد استفاده کرد. چنانچه در مدل مورداستفاده، سطوح متغیّر وابسته نامناسب باشد، برآوردها تورش‌دار، ناسازگار و نامناسب خواهند بود. خوشبختانه، مدل‌های زیادی وجود دارند که برای CLDVsها مورداستفاده قرار می‌گیرند. مدل‌های لاجیت و پروبیت دوگانه (دوحالتی) برای پیامدهای دوحالتی مناسب هستند و مدل‌های لاجیت و پروبیت ترتیبی در مواردی مناسب‌اند که متغیّر وابسته ماهیت ترتیبی و بیش از دو طبقه داشته باشد. مدل لاجیت چندگانه[1] برای پیامدهای اسمی[2] مناسب است و مدل توبیت برای پیامدهای سانسورشده طراحی شده است. علاوه‌بر این، مدل‌های متعددی چون مدل‌های رگرسیون پوآسن و دوجمله‌ای منفی می‌توانند برای پیامدهای شمارشی مورداستفاده قرار گیرند. این مدل‌ها و مدل‌های مشابه در این کتاب بیشتر بررسی خواهند شد. خوشبختانه، نرم‌افزارهای خوب و قدرتمندی مانند STATA برای برآورد مدل‌های CLDVs وجود دارند که شیوهٔ کاربرد آن‌ها در کتاب برای برآورد این قبیل مدل‌ها مورداشاره قرار خواهد گرفت.

ازآنجاکه آشنایی با مفاهیم پایه‌ای آمار برای درک اصول اقتصادسنجی و تحلیل مدل‌های CLDVs بسیار حائز اهمیت است، در فصل ۱ کتاب، مروری بر مفاهیم پایه‌ای آمار شده است تا خوانندگان محترم با یادآوری موضوعات، برای درک مطالب فصل‌های آتی مشکل خاصی نداشته باشند. در فصل ۲ نیز مروری مختصر بر الگوی رگرسیون خطّی و شیوه‌های برآورد آن شده است و روش حداقل مربعات معمولی OLS و حداکثر درست‌نمایی MLE به اختصار توضیح داده شده و مقایسه‌ای بین مدل‌های خطّی و غیرخطّی انجام شده است. فصل ۳ مربوط به مدل‌های با پیامد دوحالتی است که در آن مدل‌های لاجیت، پروبیت و C LOG-LOG و LOG-LOG معرفی شده‌اند. در فصل ۴ پیامدهای ترتیبی شامل مدل‌های لاجیت ترتیبی و سایر مدل‌های مرتبط مورداشاره قرار گرفته‌اند و به آزمون‌های مهم در این خصوص اشاره شده است. فصل ۵ به مدل‌های با پیامد اسمی اختصاص دارد و مدل‌های لاجیت چندگانه، لاجیت شرطی، لاجیت متداخل، لاجیت ترکیبی و پروبیت چندگانه به اختصار در این فصل معرفی شده‌اند. در فصل ۶ پیامدهای محدودشده و مدل توبیت به همراه کاربردهای آن مورداشاره قرار گرفته و انواع مدل‌های توبیت به همراه روش‌های برآورد آن بررسی شده‌اند. درنهایت، فصل ۷ نیز به پیامدهای شمارشی اختصاص دارد و انواع مدل‌های مرتبط با این پیامدها که کاربردهای گسترده‌ای در علوم مختلف دارند، مورداشاره قرار گرفته‌اند.

در پایان هر فصل کاربردهای مطالب در نرم‌افزار STATA و شیوهٔ برآورد مدل و انجام آزمون‌ها در این نرم‌افزار به تفصیل و به‌صورت خودآموز مورداشاره قرار گرفته است تا محقّقان محترم بتوانند به‌راحتی به برآورد و آزمون‌های لازم در ارتباط با مطالب هر فصل بپردازند. با ذکر چند مثال کاربردی و دراختیارگذاشتن داده‌ها و مسیر داده‌ها برای محقّقان، امکان تمرین بیشتر و تفسیر نتایج به‌صورت گام‌به‌گام برای علاقه‌مندان فراهم شده است. در برخی فصل‌ها نیز در انتهای فصل، به چند مطالعهٔ کاربردی مرتبط با

1. Multinomial logit
2. Nominal outcomes

موضوع فصل اشاره شده است تا بر کاربردی بودن مطالب تأکید شود و اهمیت مدل‌های تحت بررسی بیشتر مشخص گردد.

مطالب این کتاب نتیجهٔ تدریس بیش از ده سال اقتصادسنجی در مقاطع مختلف تحصیلی در دانشگاه‌های مختلف ازجمله دانشگاه فردوسی مشهد است که امیدواریم مورداستفادهٔ علاقه‌مندان قرار گیرد. قطعاً مطالب اشاره شده خالی از اشکال نیست و امیدواریم خوانندگان محترم ایرادات احتمالی را برای نویسندگان ارسال کنند تا در ویرایش‌های بعدی کتاب اصلاح شود. در پایان از همهٔ عزیزانی که در تدوین مطالب، ویراستاری علمی و ادبی، داوری و درنهایت چاپ کتاب همکاری لازم را داشته‌اند، صمیمانه قدردانی می‌شود.

۱

مروری بر مفاهیم آمار

۱-۱ مقدّمه

واژهٔ آمار دو مفهوم کاملاً متفاوت دارد: در اصطلاح عام، آمار عبارت است از اعداد و ارقام معناداری که بیان کنندهٔ اطلاعاتی خاص مانند جمعیت یک کشور، تعداد خانوارهای یک شهر یا درآمد سرانه است، اما مفهوم تخصّصی آن یعنی علم آمار، مجموعه‌ای از روش‌هاست که برای جمع‌آوری، طبقه‌بندی، تلخیص، تجزیه وتحلیل و تفسیر اطلاعات مورداستفاده قرار می‌گیرد. ازآنجاکه یادآوری مباحث پایهٔ آماری، پیش نیاز اصلی مباحث اقتصادسنجی است، در این فصل برخی از مفاهیم پایهٔ آماری که در فصل‌های آتی بیشتر استفاده شده است، به‌طور مختصر مورداشاره قرار گرفته است. برخی موضوعات آماری ازقبیل انواع توزیع‌های پیوسته و گسسته برای درک بسیاری از مفاهیم اقتصادسنجی موردنیاز است و به‌همین‌دلیل به انواع این توزیع‌ها در این فصل به اختصار اشاره شده است. برخی موضوعات نیز ازقبیل آزمون فرضیه‌های آماری و تحلیل‌های آمار استنباطی، کاربرد بسیاری در اقتصادسنجی دارند که به دلیل گسترده بودن موضوعات مربوط، در این فصل چندان موردتأکید قرار نگرفته‌اند و برای توضیحات بیشتر درخصوص هرکدام از این مفاهیم، باید مثال‌های متعدد از آن‌ها را با کمک کتاب‌های مرجع در این حوزه بررسی کرد تا ضمن تسلّط بر مفاهیم مربوط، بتوان درک بهتری از موضوعات اقتصادسنجی نیز داشت. در انتهای فصل حاضر نیز پس از معرفی نرم‌افزار STATA برخی از روش‌های تحلیل آماری و برآورد رگرسیون خطّی با کمک این نرم‌افزار مورداشاره قرار گرفته است.

۱-۲ پارامتر و آماره

اطلاعات آماری، از یک جامعهٔ آماری حاصل می‌شود. جامعه[1] مجموعه‌ای از عناصر موردنظر است که حداقل یک صفت مشترک (صفت مشخصه) داشته باشند. جامعهٔ آماری، محدود یا نامحدود است. جامعهٔ آماریِ محدود جامعه‌ای است که عناصر آن، مجموعه‌ای محدود و پایان‌پذیر باشد، ولی جامعهٔ آماریِ نامحدود جامعه‌ای است که عناصر آن مجموعه‌ای نامحدود یا بی‌پایان باشد. جمع‌آوری اطلاعات از جامعه (آمارگیری) به محدود یا نامحدود بودن جامعه بستگی دارد. اگر جامعه محدود باشد، اغلب می‌توان از سرشماری (جمع‌آوری اطلاعات موردنظر از تمام عناصر جامعه) استفاده کرد، اما درصورتی که جامعه نامحدود باشد، اغلب سرشماری انجام نمی‌شود؛ زیرا یا دسترسی به تمام عناصر جامعه ممکن نیست و یا سرشماری بسیار پرهزینه و وقت‌گیر است. بنابراین پارامتر (شاخص موردنظر) همواره در جامعه مجهول است، هرچند در هرلحظه مقدار ثابتی دارد. در چنین شرایطی، جمع‌آوری اطلاعات با نمونه‌گیری انجام می‌شود. نمونه[2] تعداد محدودی از عناصر جامعه است که تمام ویژگی‌های جامعه را دارد. اگر شاخصی از جامعه با آمارگیری از تمام عناصر آن (سرشماری) جامعه به دست آید، به آن پارامتر[3] گویند. برای مثال، اگر میانگین قد دانش‌آموزان یک کشور با استفاده از اندازه‌گیری قد تمام دانش‌آموزان آن کشور حاصل شود، پارامتر جامعه به دست‌آمده است. اما اگر شاخصی از جامعه با آمارگیری از بخشی از عناصر آن (نمونه‌گیری) به دست آید، به آن آماره[4] (برآوردکننده یا تخمین زننده) گفته می‌شود.

طبیعی است که چون نمونه به‌صورت تصادفی از نقاط مختلف جامعه اختیار می‌شود، مقدار آماره از یک نمونه به نمونهٔ دیگر تغییر می‌کند و بنابراین یک متغیّر تصادفی است. ازاین‌رو، اگر بخواهیم شاخص‌های به‌دست‌آمده از نمونه‌گیری (آماره‌ها) را به پارامتر جامعه تعمیم دهیم، باید آن‌ها را آزمون کنیم تا نخست، بتوانیم میزان خطا در هریک از آماره‌ها را محاسبه کنیم و دوم، بتوانیم از بین آن‌ها بهترین شاخصی را پیدا کنیم که بتواند تقریب نزدیکی از پارامتر جامعه باشد. در آمار استنباطی، ابتدا شاخص‌های نمونه (آماره‌ها) ازطریق نمونه‌گیری از جامعه محاسبه می‌گردند و سپس به کمک برآورد (تخمین) و آزمون، نتایج حاصل از نمونه (آماره‌ها) به پارامترهای جامعه تعمیم داده می‌شوند. به‌طورکلی در مباحث آماری، زمانی که سخن از استنباط و استنتاج باشد، آن را آمار استنباطی می‌خوانند.

۱-۳ انواع متغیّرها

صفت[5]، کمّیت یا کیفیتی است که متعلّق به عناصر جامعهٔ آماری بوده و به دو بخش صفت ثابت (مشخصه،

1. Population
2. Sample
3. Parameter
4. Statistic
5. Attribute

مشترک) و صفت متغیّر تقسیم می‌شود. صفت ثابت، خصوصیتی است که میان افراد جامعه مشترک بوده و افراد جامعه را از جوامع دیگر متمایز می‌کند، مانند صفت دانش‌آموز بودن برای جامعهٔ دانش‌آموزان یک کشور که آن‌ها را مثلاً از دانشجویان متمایز می‌کند. اما صفت متغیّر، خصوصیتی است که اندازه‌گیری آن در یک جامعهٔ آماری، اعضای جامعه را از هم متمایز می‌کند، مانند قد دانش‌آموزان، حقوق کارمندان و نظایر آن که در مباحث آماری، این صفت، متغیّر[1] نامیده می‌شود. متغیّرها براین اساس که قابل‌اندازه‌گیری باشند یا نباشند، به‌صورت زیر دسته‌بندی می‌شوند:

۱- **متغیّر کمّی (قابل اندازه‌گیری یا شمارش):** متغیّرهایی که قابل‌اندازه‌گیری یا شمارش هستند و همواره می‌توان عددی (اعشاری یا صحیح) را به آن‌ها نسبت داد، متغیّرهای کمّی نامیده می‌شوند؛ مانند قد افراد یک شهر، شدت زلزله، میزان آلودگی هوا و تعداد تخلّفات رانندگی.

متغیّرهای کمّی براین اساس که بتوانند هر مقداری را در یک فاصلهٔ محدود (بین دو عدد دلخواه) اختیار کنند یا خیر، به دو دسته تقسیم می‌شوند: متغیّرهای کمّی پیوسته و متغیّرهای کمّی گسسته.

أ. متغیّرهای کمّی پیوسته: یک متغیّر کمّی را پیوسته گویند، درصورتی که هرگاه دو مقدار a و b را اختیار کرد، بتواند هر مقدار بین آن دو را نیز اختیار کند. به‌عبارت‌دیگر، عدد اعشاری برای آن مفهوم داشته باشد. برای مثال، اگر وزن افراد یک کلاس بین ۵۰ تا ۶۰ کیلوگرم باشد، آنگاه وزن یک فرد می‌تواند هر عددی بین ۵۰ تا ۶۰ کیلوگرم (مانند ۵۲/۵ یا ۵۳/۷۵ یا ۵۳) باشد. بنابراین، وزن یک متغیّر کمّی پیوسته است و تمام متغیّرهای کمّیِ قابل‌اندازه‌گیری مانند قد، وزن، درجه حرارت، سن و نظایر آن که می‌توانند تمام مقادیر یک بازه $[a\ ,\ b]$ را اختیار کنند، متغیّرهای کمّی پیوسته هستند.

ب. متغیّرهای کمّی گسسته: یک متغیّر کمّی را گسسته گویند، درصورتی که هرگاه دو مقدار a و b را اختیار کرد، نتواند هر مقدار بین آن دو را اختیار کند. به‌عبارت‌دیگر، عدد اعشاری برای آن مفهوم نداشته باشد. آن دسته از متغیّرهای کمّی که پیوسته نباشند، گسسته هستند. برای مثال، اگر تعداد مراجعه به پزشک برای یک فرد در طول ماه ۲ بار تا ۵ بار باشد، ممکن نیست برای یک فرد این مراجعه مثلاً ۳/۵ بار اتفاق افتاده باشد. پس تمام متغیّرهای کمّی قابل‌شمارش مانند تعداد افراد خانواده، تعداد تخلّفات رانندگی و تعداد مراجعه به پزشک و نظایر آن که نمی‌توانند تمام مقادیر بازهٔ $[b, a]$ را اختیار کنند، متغیّر کمّی گسسته هستند.

۲- **متغیّر کیفی (غیرقابل اندازه‌گیری):** متغیّرهایی که قابل‌اندازه‌گیری یا شمارش نیستند و فقط ممکن است به نوع یا ترتیب خاصی تعلّق داشته باشند، متغیّرهای کیفی نامیده می‌شوند. برای مثال، گروه خونی (A, B, AB, O)، نوع کشت (دیم یا آبی)، کیفیت کالا (سالم یا معیوب)، شغل (آزاد یا دولتی)، مراحل زندگی (کودکی، نوجوانی، جوانی، میان‌سالی و...)، مراحل تحصیل (دبستان، راهنمایی،

1. Variable

دبیرستان،...)، سطح درآمد (پردرآمد، متوسط درآمد، کم درآمد) و نظایر آن که قابل‌شمارش نیستند، جزء متغیّرهای کیفی به‌شمار می‌آیند. متغیّرهای کیفی براین اساس که یک ترتیب طبیعی بین آن‌ها وجود داشته باشد یا خیر، به دو دسته تقسیم می‌شوند: متغیّرهای کیفی ترتیبی و متغیّرهای کیفی اسمی.

أ. متغیّرهای کیفی ترتیبی: در این متغیّرها یک نوع ترتیب طبیعی بین طبقات وجود دارد؛ مانند «بالا، وسط، پایین»، «خوب، متوسط، ضعیف»، «پردرآمد، متوسط درآمد، کم درآمد».

ب. متغیّرهای کیفی اسمی: در این نوع متغیّرها هیچ نوع ترتیبی بین طبقات وجود ندارد و هیچ مقایسه‌ای بین متغیّرها امکان‌پذیر نیست؛ مانند گروه خونی (A, B ,AB, O) یا نوع شغل (آزاد یا دولتی).

۱-۴ تابع احتمال[1]

تابعی که به کمک آن می‌توان احتمال وقوع وضعیت‌های مختلف یک متغیّر تصادفی را به دست آورد، تابع احتمال نامیده می‌شود.

به‌عبارت‌دیگر، تابع احتمال f(x) تابعی است که:

أ. دامنهٔ آن (x)، مقادیر ممکن از یک متغیّر تصادفی دلخواه است.

ب. حوزه یا برد آن (f(x))، احتمالات مربوط به مقادیر متغیّر تصادفی است.

قابل ذکر است که همواره مقادیر تابع احتمال f(x) باید مثبت باشد، یعنی ($f(x) \geq 0$) و همچنین مجموع احتمالات یک متغیّر تصادفی همواره برابر ۱ است. در نتیجه ازآنجاکه حوزهٔ f(x) همان احتمالات ممکن برای یک متغیّر تصادفی است، باید مجموع مقادیر f(x) به‌ازای مقادیر ممکن برای متغیّر x در دو حالت گسسته و پیوسته، برابر ۱ شود. به‌عبارت‌دیگر:

$$X \text{ گسسته}: \begin{cases} f(x) \geq 0 \\ \sum_{\forall x} f(x) = 1 \end{cases} \qquad X \text{ پیوسته}: \begin{cases} f(x) \geq 0 \\ \int f(x)dx = 1 \end{cases}$$

۱-۴-۱- تابع احتمال گسسته[2]

درصورتی که x متغیّر تصادفی گسسته باشد، آنگاه f(x) با داشتن شرایط زیر، تابع احتمال متغیّر x خواهد بود:

أ. $\forall x: \ 0 \leq f(x) = p(x) \leq 1$ ، یعنی تابع f(x) به‌ازای هر مقدار از x برابر با p(x) (احتمال x) است و مقداری بین ۰ و ۱ دارد.

ب. $\sum f(x) = \sum p(x) = 1$، یعنی مجموع مقادیر f(x) برابر با مجموع مقادیر احتمال p(x) و برابر ۱ است.

x	x_1	...	x_n	
$P(X=x)=f(x)$	$0 \leq f(x_1) \leq 1$	...	$0 \leq f(x_n) \leq 1$	$\sum f(x) = \sum P(x) = 1$

1. Probability function
2. Discrete probability function