

## تمرین سری هفتم: استتا

موعد تحویل: سه‌شنبه ۱۶ اردیبهشت ۱۳۹۹، ساعت ۱۲:۳۰

نکته: برای پاسخ این تمرین لازم است علاوه بر فایلی که به صورت PDF که پاسخ تمرینات را دارد فایل کد استتا و لوگ‌فایل را نیز تحویل دهید. لطفاً این سه مورد را به صورت یک فایل فشرده در سامانه درس‌افزار آپلود نمایید.

این تمرین بر اساس داده‌های بودجه خانوار در سال‌های ۱۳۸۴ تا ۱۳۹۷ و هزینه‌کرد خانوارهای شهری و روستایی بر روی پوشاک و کفش طراحی شده است. در فایلی که در اختیار شما قرار گرفته است فولدر raw شامل داده‌های خام می‌باشد. در اسامی فایل‌ها R برای خانوارهای روستایی و U برای خانوارهای شهری است. هم‌چنین سال نیز در اسم فایل مشخص است. به عنوان مثال فایل D\_R84p3s03 مربوط به مخارج خانوارهای روستایی در سال ۸۴ روی پوشاک و کفش است. توجه کنید که هزینه‌کرد خانوار در یک ماه گذشته مورد سوال بوده است یعنی اعداد هزینه‌کرد میزان هزینه را در یک ماه گذشته نشان می‌دهد و برای به‌دست آوردن هزینه سالیانه لازم است این اعداد را ۱۲ برابر بکنید. فایل CPI.xlsx شاخص قیمت کل را برای خانوارهای شهری و روستایی به طور جداگانه و برای سال پایه ۹۰ نشان می‌دهد. فایل weights.dta نیز وزن هر خانوار را به همراه یک سری مشخصات دیگر در سطح خانوار از جمله شهری و روستایی بودن، درصد مردها، درصد باسوادها، میانگین سنی و تعداد اعضای خانوار نشان می‌دهد. (لیبل متغیرها را ملاحظه بفرمایید.) وزن خانوار در نمونه تعداد خانوارهایی را در کشور نشان می‌دهد که مشابه این خانوار هستند. بنابراین برای به‌دست آوردن مشخصات در سطح کشور لازم است حتماً از میانگین وزنی و ... استفاده کنید. هم‌چنین پرسش‌نامه بودجه خانوار سال ۹۷ نیز به عنوان نمونه در اختیار شما قرار گرفته است. پرسش‌نامه در بخش پوشاک برای سال‌های قبل نیز مشابه این سال است.

۱. فایل D\_R97p3s03 را در استتا باز و تمیز کرده و در یک فولدر جداگانه و با عنوان مناسب ذخیره کنید.

برای تمیز کردن لازم است حداقل اسامی متغیرها را تغییر دهید و متغیرهای عددی را از رشته به عدد تبدیل کنید.

۲. متغیر جدیدی بسازید که نشان می‌دهد هر خانوار در این داده چند بار تکرار شده است. این متغیر چه چیزی را

نشان می‌دهد و چرا همواره برابر یک نیست؟ بیشترین مقدار این متغیر جدید چند است؟

۳. کاری که در مرحله «۱» انجام دادید را برای تمام فایل‌های هزینه‌کرد (یعنی ۲۸ فایل) انجام دهید. (طبیعتاً لازم

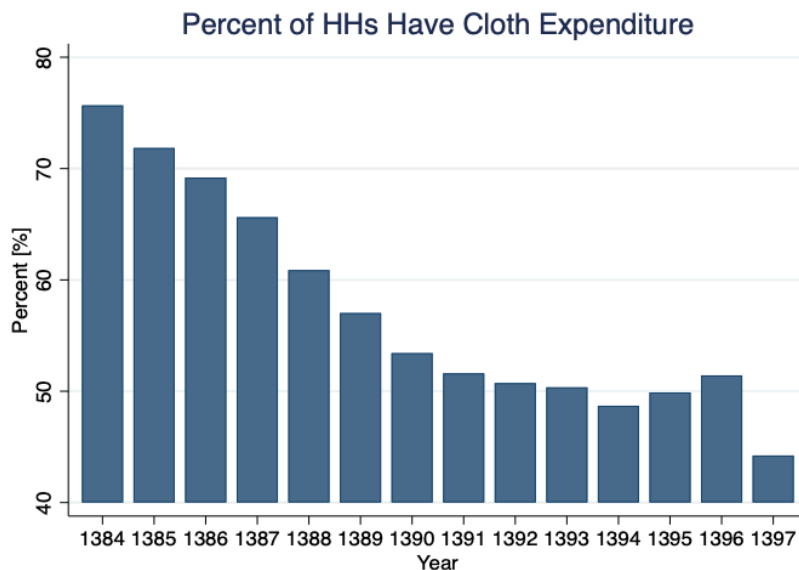
است از دستورات حلقه استفاده کنید تا مجبور نباشید ۲۸ بار یک کد را تکرار کنید.)

۴. فایل جدیدی بسازید که شامل هزینه‌کرد تمام خانوارهای موجود در داده (اعم از شهری و روستایی و برای تمام سال‌ها) روی پوشاک و کفش است. لازم است مجموع هزینه‌ها برای یک خانوار در هر سال تجمیع شده باشند یعنی هر خانوار (در هر سال) تنها یک بار در داده موجود باشد و یکی از ستون‌ها هزینه‌کرد اسمی سالانه این خانوار را روی پوشاک و کفش نشان دهد.

۵. با استفاده از فایل CPI.xlsx میزان مخارج را حقیقی کنید. هم‌چنین با استفاده از داده تعداد اعضای خانوار در داده weights هزینه‌کرد را به سرانه تبدیل کنید.

۶. داده weights تمام خانوارهای بودجه خانوار را نشان می‌دهد ولی داده هزینه‌کرد روی پوشاک مربوط به خانوارهایی است که در ماه گذشته هزینه‌کرد روی پوشاک داشته‌اند. پاسخ دهید چند درصد خانوارها در هر سال هزینه‌کرد غیر صفر روی پوشاک داشته‌اند؟ نموداری مانند شکل ۱ (با جزئیات موجود در شکل) رسم کنید که این مورد را نشان می‌دهد. (یادآوری: به وزن خانوارها توجه کنید!)

این تمرین ادامه دارد ...



شکل ۱: درصد خانوارهایی که روی پوشاک هزینه کرده‌اند