

# بررسی عدم توازن بین هزینه و درآمد خانوار در ایران

تاریخ دریافت: ۸۶/۴/۱۱

تاریخ پذیرش: ۸۶/۱۱/۲۵

فرشید جمشیدی\*

زهره فلاح محسن‌خانی\*\*

طی چهل سال که از اجرای طرح هزینه و درآمد خانوار در ایران می‌گذرد، در تمامی این سالها، متوسط درآمد خانوارها کمتر از هزینه بوده است. هرچند ممکن است در برخی سالها به علت پاره‌ای از مشکلات اقتصادی، این پدیده قابل توجیه باشد، اما استمرار آن در یک دوره طولانی گویای این واقعیت است که اختلاف، ناشی از بحرانهای اقتصادی موجود در سیکل‌های اقتصادی

\* فرشید جمشیدی؛ کارشناس پژوهشکده آمار.

E.mail: f\_jamshidi@srtc.ac.ir

\*\* زهره فلاح محسن‌خانی؛ عضو هیأت علمی پژوهشکده آمار.

E. mail: zohrehf@srtc.ac.ir

نیست؛ بلکه تداعی کننده کم پوششی طرح آمارگیری یا کم گویی درآمد در دسته خاصی از خانوارها است.

در این مطالعه پس از بررسی و تحلیل نتایج آمارگیری از هزینه و درآمد خانوار در سالهای ۸۰-۱۳۷۶ مشخص شد که این مشکل منحصر به یک دسته خاص از خانوارها نیست، بنابراین برای رفع این مشکل باید پرسشهای مربوط به بخش درآمد در پرسشنامه آمارگیری از هزینه و درآمد خانوار بگونه ای طراحی شود که پوشش بهتر درآمد حاصل شود.

### کلید واژه‌ها:

توازن، هزینه و درآمد خانوار، طرح آمارگیری، تجزیه واریانس، روش دانکن

## مقدمه

بیشتر بودن هزینه از درآمد مسئله‌ای است که گاهی می‌تواند اتفاق بیفتد. در چنین مواقعی خانوار از پس‌انداز خود برای جبران هزینه‌هایی که درآمدی برای آنها وجود نداشته است، استفاده می‌کند؛ ولی در صورت تداوم این مسئله خانوار یا باید از هزینه‌های خود بکاهد یا منبع درآمد جدیدی برای خود کسب کند. با این توضیح وجود هزینه‌های بیشتر از درآمد در دراز مدت برای یک خانوار توجیه‌پذیر نیست. در بعد کلان نیز نمی‌توان متوسط هزینه‌های خانوارهای جامعه را در طولانی مدت بیشتر از متوسط درآمدهای آنها تصور نمود. در یک مقطع خاص زمانی به دلایلی نظیر تورم، رکود اقتصادی، گرانی ناگهانی اقلام ضروری و نظایر آن، وجود این پدیده می‌تواند توجیه اقتصادی داشته باشد، اما در آمارگیری از هزینه و درآمد خانوار که هر ساله به صورت نمونه‌گیری انجام شده است؛ تقریباً هر ساله این پدیده مشاهده شده که به یقین یا نشان‌دهنده کم‌گویی درآمد از سوی خانوارها یا کم‌پوششی درآمد در طرح آمارگیری است.

یکی از علل کم‌گویی درآمد، عدم صحت پاسخ به پرسش‌شوندگانی است که به علی‌قصد مخفی نگه‌داشتن درآمد واقعی خود را دارند. چنانچه به توان گروه‌هایی از خانوارها را یافت که در پاسخگویی به پرسش‌های مربوط به درآمد صادق نیستند شاید با تغییر رویه آمارگیری از این گروه‌ها به توان، درصد قابل توجهی از پدیده کم برآوردی درآمد را جبران کرد.

در صورتی که نتوان چنین خانوارهایی را از سایر خانوارها جدا نمود، می‌توان دلیل بروز این پدیده را وجود نقص در پرسش‌های مربوط به درآمد طرح آمارگیری هزینه و درآمد دانست؛ به عبارتی وقتی اختلاف درآمد از هزینه در خانوارهایی با مشخصه‌های خاص منفی نشود و این مشکل در تمام جامعه آمارگیری ملاحظه شود باید از رویکرد دیگری که آن اصلاح پرسش‌های مربوط به درآمد برای کل جامعه آماری است، استفاده شود. به عبارتی هدف این مقاله پاسخ به این پرسش است که آیا در ایران خانوارهایی با مشخصه‌های خاص، تمایل به ابراز درآمد حقیقی خود ندارند یا مشکل کم‌گویی درآمد به علت عدم اطلاع‌گیری صحیح بخش درآمد در آمارگیری هزینه و درآمد است. برای انجام بررسی‌های فوق از نتایج طرح

آمارگیری از هزینه و درآمد خانوارهای شهری<sup>۱</sup> و روستایی<sup>۲</sup> مرکز آمار ایران طی سالهای ۱۳۷۶ تا ۱۳۸۰ استفاده شده است. این اطلاعات شامل مشخصات خانوار از قبیل تعداد افراد، تعداد افراد شاغل، تعداد افراد باسواد، تعداد افراد دارای درآمد و مشخصات سرپرست خانوار؛ از قبیل وضع سواد، مدرک تحصیلی، وضع فعالیت، کد شغل، کد فعالیت محل کار و وضع شغلی است. همچنین این اطلاعات شامل داده‌هایی مربوط به هزینه و درآمد خانوار از قبیل: هزینه‌های خالص، ناخالص، خالص خوراکی، خالص غیر خوراکی و درآمد بوده که هزینه‌های خوراکی، غیر خوراکی و درآمد به تفکیک اجزای آنها نیز موجود است. مرحله اول در تحلیل این داده‌ها آن است که مشخصه‌هایی از خانوار که آنها را از نظر اختلاف منفی هزینه و درآمد بهتر دسته‌بندی می‌کنند، انتخاب شود و در مرحله دوم سطوحی از مشخصه، که خانوارهای موجود در آن سطوح دارای بیشترین اختلاف منفی هزینه و درآمد هستند استخراج می‌شود. پس از مرحله دوم، هدف اصلی، پاسخ به این پرسش است که خانوارهایی که در سطوح استخراج شده در مرحله دوم قرار گرفته‌اند، چه کسری از اختلافات منفی هزینه و درآمد را پوشش می‌دهند؟ اگر سهم کوچکی از کل خانوارها- که دارای اختلاف منفی هزینه و درآمد هستند- سهم بزرگی از کل اختلافات منفی هزینه و درآمد را پوشش دهد، می‌توان گفت که با اصلاح اطلاع‌گیری از آنها، مشکل اختلاف منفی هزینه و درآمد در جامعه حل و یا تا حد قابل توجهی برطرف می‌شود. اما اگر برای پوشش دادن سهم بزرگی از کل اختلافات منفی هزینه و درآمد، سطوحی از مشخصه انتخاب شده استخراج شوند که سهم بزرگی از خانوارها در آن سطوح قرار بگیرند، مشکل حل نشده است؛ یعنی اگر جواب پرسش بالا مثبت باشد، برای رفع مشکل وجود اختلافات منفی در کل جامعه باید بر روی گروه‌های خاصی از خانوارها تمرکز شود و اگر جواب منفی باشد، مشکل وجود اختلافات منفی مربوط به گروه‌های خاصی (که می‌توان بر اساس اطلاعات موجود آنها را مشخص کرد) نیست؛ بلکه رویه آمارگیری بطور کل باید اصلاح شود.

۱. آمارگیری از هزینه و درآمد خانوارهای شهری، تهران، مرکز آمار ایران.

۲. آمارگیری از هزینه و درآمد خانوارهای روستایی، تهران، مرکز آمار ایران.

بدین منظور در بخش دوم مقاله داده‌های آمارگیری از هزینه و درآمد خانوار مرکز آمار ایران طی سال‌های ۸۰-۱۳۷۶ بررسی و تحلیل شده است.

## بررسی و تحلیل نتایج آمارگیری از هزینه و درآمد خانوار طی سال‌های ۸۰-۱۳۷۶

مرحله اول: تعیین مشخصه‌های مناسب برای دسته‌بندی خانوارها از نظر اختلاف منفی هزینه و درآمد:

در این مرحله از ایده تجزیه واریانس<sup>۱</sup> استفاده می‌شود، دیدگاه تجزیه واریانس این است که<sup>۳</sup> اگر جامعه‌ای با متغیر  $x$  به  $k$  گروه تقسیم شده باشد، بطوریکه  $x_{ij}$ ، متغیر مشاهده شده فرد  $j$  ام گروه  $i$  ام،  $j = 1, \dots, n_i$ ،  $i = 1, \dots, k$  و  $n_i$  تعداد افراد واقع در گروه  $i$  ام باشد، مجموع مربعات اختلاف متغیر هر فرد از میانگین جامعه به نام تغییرات کل (SST) به صورت زیر به دو قسمت تجزیه می‌شود<sup>۴</sup>:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^k n_i (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2$$

که در آن:

<sup>۱</sup>. Analysis of Variance

<sup>۲</sup>. Kotz, Samuel, *Encyclopedia of Statistical Sciences*, (John Wiley and Sons Inc USA, 2005), Second Edition, Volume 1, pp.133-141.

<sup>۳</sup>. Neter, J., Kutner, M and Wasserman, W, *Applied Linear Statistical Models*. (Mc Graw Hill, 1996).

<sup>۴</sup>. Montgomery, D.C. *Design and Analysis of Experiments*. John Wiley & Sons. Inc, USA, 2000).

$$\bar{x}_i = \text{متوسط متغیر } x \text{ در گروه } i \text{ ام}$$

$$\bar{x}_{..} = \text{متوسط متغیر } x \text{ بین کل افراد}$$

جمله نخست سمت راست تساوی، تغییرات بین گروهها ( $SS_{bg}$ ) و جمله دوم سمت راست تساوی، تغییرات داخل گروهها ( $SS_{wg}$ ) نام دارند. هر چه تغییرات بین گروهها سهم بیشتری از تغییرات کل را به خود اختصاص دهد، بدان معنی است که گروههای در نظر گرفته شده برای جامعه، اختلافات بیشتری نسبت به هم دارند. نسبت  $R^2 = \frac{SS_{bg}}{SS_T}$  عددی

حقیقی بین صفر و یک است<sup>۱</sup> که هر چه به یک نزدیکتر باشد، بیانگر وجود اختلاف بیشتر بین گروهها و هر چه به صفر نزدیکتر باشد، بیانگر همانندی بیشتر گروهها است. از طرفی برای آزمون معنی داری اختلاف بین گروهها از کمیت دیگری استفاده می شود<sup>۲</sup> که عبارت است از،  $F = \frac{SS_{bg} / K - 1}{SS_{wg} / N - K}$  که در آن  $N$  تعداد افراد جامعه است. اگر فرضیاتی نظیر

نرمال بودن متغیر و یکسان بودن واریانس گروهها برقرار باشد،  $F$  دارای توزیع  $F_{(k-1, N-K)}$  است و اگر  $F > F_{(\alpha; k-1, N-K)}$  باشد، فرض یکسان بودن گروهها در سطح  $\alpha$  رد می شود<sup>۳</sup>.

برای تعیین مشخصه‌های مناسب می توان از ایده تجزیه واریانس کمک گرفت<sup>۴</sup>، به این ترتیب که خانوارهایی که اختلاف هزینه و درآمد آنها منفی است به عنوان «جامعه آماری» و سطوح مشخصه‌ها که بر اساس قسمت یکم (خصوصیات اجتماعی اعضای خانوار) پرسشنامه هزینه و درآمد خانوار قابل حصول است؛ به عنوان «گروههای جامعه آماری» در نظر گرفته می شوند؛ برای مثال چنانچه مشخصه مورد نظر وضع شغلی سرپرست خانوار باشد، تمامی خانوارهایی که اختلاف هزینه و درآمد آنها منفی است به شش گروه تقسیم می شوند که

<sup>1</sup>. Krishnaiah, P. R., Mudholkar, G. S., and Subbaiah, P. *In Analysis of Variance*, Vol 1, (Handbook of Statistics, North-Holland, New York, 1980).

<sup>2</sup>. Lindman, H. R. *Analysis of Variance in Complex Experimental Designs*, (San Francisco: W. H. Freeman & Co, 1974).

<sup>3</sup>. Christensen, R. *Analysis of Variance, Design and Regression: Applied Statistical Methods*, (Chapman & Hall, USA, 1998).

<sup>4</sup>. Cortina, J. M. and Hossein Nouri, "Effect Size for ANOVA Designs", *Sage Publications. Quantitative Applications in the Social Science Series*, No 129, (2000).

عبارت است از خانوارهایی با سرپرست شاغل، بیکار، دارای درآمد بدون کار، دانش‌آموز، خانه‌دار و سایر. به این ترتیب از بین دو مشخصه، مشخصه‌ای مناسب‌تر است که  $R^2$  ی مربوط به تجزیه واریانس بر اساس گروه‌بندی ایجاد شده توسط آن مشخصه بیشتر باشد.  $R^2$  ی مربوط به مشخصه‌های قابل تعریف برای خانوارها با استفاده از اطلاعات حاصل از آمارگیری‌های هزینه و درآمد خانوارهای شهری طی سالهای ۱۳۷۶ تا ۱۳۸۰، محاسبه شده و در جدول (۱) خلاصه شده‌اند.

جدول ۱.  $R^2$  مربوط به مشخصه‌های خانوار شهری

نام مشخصه	$R^2$	نام مشخصه	$R^2$	نام مشخصه	$R^2$	نام مشخصه	$R^2$
تعداد افراد خانوار	۰/۰۰۲۸۰۵	تعداد افراد دارای درآمد خانوار	۰/۰۰۱۹۹۸	مدرک تحصیلی سرپرست خانوار	۰/۰۱۵۹۶۹	شغل سرپرست خانوار	۰/۰۲۴۱۳۴
تعداد افراد شاغل خانوار	۰/۰۰۲۰۷۱	تعداد افراد باسواد خانوار	۰/۰۰۶۷۹۳	وضع شغلی سرپرست خانوار	۰/۰۰۹۳۲۸	فعالیت اصلی محل کار سرپرست خانوار	۰/۰۱۰۸۴۹

تعداد مشاهدات = ۲۵۹۹۷ خانوار

$R^2$  های محاسبه شده، گویای این مطلب هستند که هیچ کدام از مشخصه‌ها نمی‌توانند آنطور که انتظار می‌رود خانوارها را از نظر اختلاف منفی هزینه و درآمد دسته‌بندی کنند با این وجود مشخصه‌های شغل، فعالیت اصلی محل کار و مدرک تحصیلی سرپرست خانوار نسبت به سایر مشخصه‌ها از تفاوت‌های بیشتری بین سطوح خود برخوردار هستند و اگر بتوان مشخصه‌ای مناسب را تعریف نمود، آن مشخصه از «تلفیق» مشخصه‌های فوق حاصل می‌شود. منظور از «تلفیق» مشخصه این است که با کنار هم قرار دادن سطوح دو مشخصه، مشخصه جدیدی تعریف شود که تعداد سطوح آن برابر حاصلضرب تعداد سطوح در مشخصه اصلی است. برای مثال اگر مشخصه A و B به ترتیب دارای سطوح  $\{a_1, a_2\}$  و  $\{b_1, b_2, b_3\}$

باشند، مشخصه تلفیق شده این دو با یکدیگر که با AB نشان داده می‌شود، دارای شش سطح؛ بدین ترتیب می‌باشد:  $\{a_1b_1, a_1b_2, a_1b_3, a_2b_1, a_2b_2, a_2b_3\}$ .  
 به این ترتیب از تلفیق دو به دوی سه مشخصه مذکور، ۳ مشخصه تلفیق شده حاصل می‌شود، اما به دلیل اینکه مشخصه‌های شغل و فعالیت اصلی محل کار سرپرست خانوار دارای تعداد سطوح زیادی هستند، مشخصه‌های تلفیق شده این دو با یکدیگر و هر یک با مشخصه مدرک تحصیلی سرپرست خانوار، دارای تعداد سطوح بسیار زیادی هستند و به تبع آن تعداد خانوارهای موجود در آنها بعضاً صفر، یک و حداکثر تعداد اندکی خانوار خواهد بود و در نتیجه اعتبار نتایج حاصل از این تلفیقها به دلیل تکرارهای کم در هر یک از سطوح، دچار نقصان می‌شود. بنابراین به جای در نظر گرفتن شغل و فعالیت‌های جزئی که با کدهای چهار رقمی مشخص می‌شوند از شغل و فعالیت‌های اصلی که کد یک رقمی دارند، استفاده شده است. جدول (۲)،  $R^2$  ی مربوط به مشخصه‌های دو به دو تلفیق شده سه مشخصه شغل، فعالیت و مدرک تحصیلی را در خانوارهای شهری نشان می‌دهد.

جدول ۲.  $R^2$  مشخصه‌های تلفیق شده از سه مشخصه شغل،

فعالیت و مدرک تحصیلی در خانوار شهری

مشخصه	شغل و فعالیت	فعالیت و مدرک تحصیلی	شغل و مدرک تحصیلی
$R^2$	۰/۰۳۹۱۵	۰/۱۴۰۳۸	۰/۰۳۱۷۵۳

تعداد مشاهدات = ۲۵۹۹۷ خانوار

از بین سه مشخصه تلفیق شده، مشخصه فعالیت و مدرک تحصیلی سرپرست خانوار با  $R^2$  ی تقریباً ۰/۱۴ به عنوان مناسب‌ترین مشخصه‌ای که بتواند خانوارها را از نظر اختلاف منفی هزینه و درآمد دسته‌بندی نماید، انتخاب می‌شود.

مشابه با بررسیهای ذکر شده که برای خانوارهای شهری انجام شده است، داده‌های مربوط به خانوارهای روستایی نیز مورد بررسی قرار گرفت و نتایج حاصل از آن، مشخصه‌های

تعداد افراد و تعداد افراد باسواد خانوار، شغل و وضع شغلی سرپرست خانوار را با بیشترین  $R^2$  مطابق با جدول (۳)، انتخاب می‌کند.

جدول ۳.  $R^2$  مربوط به مشخصه‌های خانوارهای روستایی<sup>۱</sup>

$R^2$	نام	$R^2$	نام	$R^2$	نام	$R^2$	نام
۰/۰۱۴۴۵۲	فعالیت محل کار	۰/۰۰۶۱۵۷	مدرک تحصیلی	۰/۰۰۳۲۴۶	تعداد افراد	۰/۰۳۶۹۷۳	تعداد افراد خانوار
۴۴۸۹۹	سرپرست خانوار	۱۷۲۴۵	سرپرست خانوار	۴۹۹۸۶	بدرآمد خانوار	۴۰۴۱۹	
۰/۰۲۶۳۶	وضع شغلی	۰/۰۲۷۳۸۵	شغل سرپرست خانوار	۰/۰۳۲۸۴۶	تعداد افراد	۰/۰۱۰۹۷	تعداد افراد
۴۴۸۹۹	سرپرست خانوار	۴۴۸۹۹	سرپرست خانوار	۴۵۷۱۴	باسواد خانوار	۴۵۴۸۷۰	شاغل خانوار

از تلفیق دوبه‌دوی چهار مشخصه مذکور، شش مشخصه تلفیق شده تعریف، و  $R^2$  ی مربوط به آنها مطابق با جدول (۴)، محاسبه شده است.

<sup>۱</sup> اختلاف در تعداد مشاهدات به این دلیل است که خانوارهایی که در سطح مشخصه آنها کمتر از پنج خانوار وجود دارد، حذف شده‌اند.

جدول ۴.  $R^2$  مربوط به تلفیق دوبه‌دوی مشخصه‌های تعداد افراد و تعداد افراد باسواد خانوار، شغل و وضع شغلی سرپرست خانوار در خانوارهای روستایی

نام مشخصه‌ی تلفیق شده	$R^2$	تعداد مشاهدات (خانوار)
تعداد افراد و تعداد افراد باسواد خانوار	۰/۰۳۲۰۴۵	۳۵۵۹۸
تعداد افراد و شغل سرپرست خانوار	۰/۰۴۷۱۷۹	۳۴۸۳۶
تعداد افراد باسواد و شغل سرپرست خانوار	۰/۰۴۸۲۳۸	۳۲۳۵۴
تعداد افراد و وضع شغلی سرپرست خانوار	۰/۰۵۴۱۱۰	۳۴۸۶۸
شغل و وضع شغلی سرپرست خانوار	۰/۰۳۵۰۹۱	۳۴۸۷۷
تعداد افراد با سواد و وضع شغلی سرپرست	۰/۰۵۵۰۰۲	۳۲۳۵۸

بدین ترتیب از بین مشخصه‌های تلفیق شده مذکور، مشخصه تعداد افراد باسواد و وضع شغلی سرپرست خانوار با  $R^2$  ی تقریباً ۰/۰۵۵ انتخاب می‌شود.

#### مرحله دوم؛ دسته‌بندی سطوح مشخصه انتخاب شده

به منظور دست یافتن به خانوارهایی با تمایز فاحش در اختلاف منفی هزینه و درآمد نسبت به خانوارهای دیگر، لازم است سطوح مشخصه تلفیق شده مرحله اول را دسته بندی نمود. این دسته‌بندی باید به گونه‌ای انجام شود که گروههای موجود در یک دسته اختلاف کم و گروههای موجود در دسته‌های متفاوت، اختلاف معنی‌داری داشته باشند. انجام این دسته‌بندی را به کمک روشهایی مانند: LSD<sup>۱</sup>، دانته<sup>۲</sup>،

<sup>۱</sup>. Least Significant Difference

<sup>۲</sup>. Dunnett

دانکن<sup>۱</sup>، توکی<sup>۲</sup> و ... می‌توان انجام داد.

هر یک از روشها، دارای مزایا و معایبی هستند و بر اساس اهداف و داده‌های مختلف ممکن است روش مناسب متفاوت باشد. اما از بین روشهای دسته‌بندی، روش دانکن روشی است که تخصیص گروهها را به راحتی انجام می‌دهد. اما در دیگر روشها، ابتدا آزمون معنی‌داری اختلاف بین دویه‌دوی گروهها انجام می‌شود، سپس کاربر با توجه به معنی‌دار بودن یا نبودن هر یک از آزمونها، دسته‌بندی را خود انجام می‌دهد مسلماً در صورتی که تعداد گروهها زیاد باشد، روش دانکن روش مطلوب‌تر و بعضی مواقع تنها روش عملی محسوب می‌شود. در این تحقیق بنا به دلایل فوق از روش دسته‌بندی دانکن استفاده شده است.<sup>۳</sup> در خانوارهای شهری دسته‌بندی دانکن سطوح مشخصه فعالیت و مدرک تحصیلی سرپرست با ۸۸ سطح را به پنج گروه تقسیم می‌کند. با مرتب کردن گروهها از نظر متوسط اختلاف منفی، گروه اول که سطوح داخل آن دارای بیشترین اختلافات منفی هستند، تنها دارای یک سطح و گروه دوم شامل چهارده سطح است. بنابراین به‌وسیله دسته‌بندی دانکن می‌توان پانزده سطح از مجموع ۸۸ سطح این مشخصه تلفیق شده را به عنوان سطوحی که خانوارهای واقع در آنها بیشترین اختلاف منفی هزینه و درآمد را دارا هستند و بطور معنی‌داری با خانوارهای واقع در سطوح دیگر تفاوت دارند، انتخاب کرد. پرسشی که در اینجا مطرح می‌شود این است که این خانوارها چند درصد از کل خانوارهای موجود را تشکیل داده و چه کسری از کل اختلافات منفی در آنها نهفته شده است؟

پانزده سطر اول جدول (۵)، سطوح انتخاب شده توسط دسته‌بندی دانکن به همراه تعداد خانوار مشاهده شده در نمونه از آن سطوح و جمع و متوسط اختلاف هزینه و درآمد خانوارهای نمونه در آن سطوح را نمایش می‌دهد. همچنین در این جدول نسبت تجمعی تعداد خانوار به کل و نسبت تجمعی جمع اختلاف هزینه و درآمد به کل نیز نشان داده شده

<sup>۱</sup>. Duncan

<sup>۲</sup>. Speed, T. P. W. "Tukey's Contribution to Analysis of Variance", *The Annals of Statistics*, Vol 30, No. 6, (2002).

<sup>۳</sup>. حسن ورمزیار، زهره فلاح محسن‌خانی و فرشید جمشیدی، *ارائه راهکارهای مناسب تولید آمارهای درآمد خانوار*، تهران، پژوهشکده آمار، ۱۳۸۳.

است که بیان‌کننده آن است که چنانچه نقطه برش در انتخاب سطوح، کد فعالیت و کد مدرک مربوط به آن سطر باشد، چند درصد از خانوارهای انتخاب شده چه کسری از جمع اختلافات را پوشش می‌دهند.

جدول ۵. تعداد خانوار، جمع و متوسط اختلاف هزینه و درآمد خانوارهای شهری موجود در سطوح انتخاب شده به وسیله دسته‌بندی دانکن به همراه نسبت تجمعی

نسبت تجمعی جمع اختلاف به کل	نسبت تجمعی تعداد خانوار به کل	تعداد خانوار	جمع اختلاف هزینه و درآمد	متوسط اختلاف هزینه و درآمد	کد مدرک تحصیلی	کد فعالیت اصلی محل کار
٪۱/۲۳	٪۰/۰۲	۶	-۴۱۶۸۱۱۹۴۸۷	-۶۹۴۶۸۶۵۸۱	۱	۷
٪۱/۴۲	٪۰/۰۸	۱۴	-۶۶۳۶۵۰۲۱۱/۹	-۴۷۴۰۳۵۸۶/۶	۲	۷
٪۱/۶۷	٪۰/۱۶	۲۲	-۸۴۰۹۵۵۳۶۷/۱	-۳۸۲۲۵۲۴۴	۳	۵
٪۲/۰۵	٪۰/۳۰	۳۵	-۱۲۸۲۷۰۵۳۶۴	-۳۶۶۴۸۷۴۴/۷	۳	۶
٪۲/۹۳	٪۰/۶۵	۹۳	-۳۰۱۲۶۶۰۲۸۴	-۳۲۳۹۴۱۹۶/۶	۷	۷
٪۳/۱۸	٪۰/۷۶	۲۸	-۸۳۷۹۶۹۶۲۸/۲	-۲۹۹۲۷۴۸۶/۷	۴	۷
٪۳/۳۰	٪۰/۸۲	۱۴	-۴۱۴۳۴۸۴۸۰/۷	-۲۹۵۹۶۶۳۲۰	۵	۷
٪۴/۲۲	٪۱/۲۵	۱۱۳	-۳۱۴۱۴۵۴۳۶۷	-۲۷۸۰۰۴۸۱/۱	۵	۶
٪۴/۳۰	٪۱/۲۸	۹	-۲۴۱۱۷۰۱۵۲/۶	-۲۶۷۹۶۶۸۳/۶	۹	۷
٪۴/۴۳	٪۱/۳۵	۱۸	-۴۴۵۷۸۷۳۱۷	-۲۴۷۶۵۹۶۲/۱	۶	۷
٪۵/۵۲	٪۱/۹۳	۱۵۱	-۳۷۱۸۰۲۱۲۸۳	-۲۴۶۲۲۶۵۷/۵	۸	۷
٪۵/۵۶	٪۱/۹۵	۵	-۱۲۰۴۶۴۹۵۲/۸	-۲۴۰۹۲۹۹۰/۶	۳	۹
٪۵/۷۱	٪۲/۰۴	۲۲	-۵۱۹۵۶۱۶۶۴	-۲۳۶۱۶۴۳۹/۳	۰	۵
٪۶/۰۶	٪۲/۲۳	۵۱	-۱۱۸۶۰۵۵۱۸۵	-۲۳۳۵۵۹۸۴	۹	۶
٪۶/۸۴	٪۲/۶۸	۱۱۵	-۲۶۶۴۵۹۴۵۶۱	-۲۳۱۷۰۳۸۷/۵	۴	۶
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
٪۷/۹	٪۷/۲۸	۱۲۶	-۱۲۳۷۳۶۲۲۱۸	-۹۸۲۰۳۳۵/۱	۱	۰

همانطور که در جدول (۵) مشاهده می‌شود، با انتخاب این پانزده سطح، خانوارهایی انتخاب شده‌اند که تقریباً ۲/۷٪ از کل تعداد خانوارهای نمونه هستند که تنها ۶/۸٪ از کل اختلافات منفی هزینه و درآمد را پوشش می‌دهند. همچنین اگر برای مثال پوشش ۸۰٪ از کل اختلافات منفی هزینه و درآمد مدنظر باشد، لازم است تعدادی از سطوح این مشخصه (تا کد مدرک، فعالیت «۰ و ۱») که ۷۲/۸٪ از کل خانوارها از آن سطوح هستند، انتخاب شود. پس جواب پرسش مطرح شده این است: نمی‌توان سطوحی از مشخصه فعالیت و مدرک تحصیلی را انتخاب نمود که با وجود درصد نسبتاً کمی از خانوارها در آن سطوح، کسر قابل توجهی از جمع اختلافات منفی هزینه و درآمد، مربوط به آن خانوارها باشد.

با انجام بررسی مشابه برای خانوارهای روستایی به وسیله دسته‌بندی دانکن سطوح مشخصه «وضع شغلی سرپرست و تعداد افراد باسواد خانوار» به شانزده گروه تقسیم شده‌اند. پس از مرتب کردن گروه که از نظر اختلاف هزینه و درآمد، گروه اول، دوم و سوم، هر یک دو سطح و گروه چهارم، ۲۱ سطح از کل ۵۴ سطح این مشخصه را دارا هستند. بنابراین سطوحی که در سه گروه اول دسته‌بندی دانکن قرار دارند، انتخاب شده‌اند. هرچند انتظار می‌رود جمعاً شش سطح در این سه گروه وجود داشته باشد، اما به دلیل اینکه گروه‌بندی در دسته‌بندی دانکن دارای مرز مشخص نیست و ممکن است سطوحی در بیش از یک گروه واقع شوند، در سه گروه اول انتخاب شده، چهار سطح قرار گرفته‌اند. چهار سطر اول جدول (۶) مشابه با جدول (۵)، اطلاعات مربوط به این گروه‌ها را نمایش می‌دهد. همانطور که در این جدول مشاهده می‌شود، با انتخاب چهار سطح قرار گرفته در سه گروه اول، تنها ۰/۷٪ از جمع اختلافات منفی هزینه و درآمد پوشش داده می‌شود.

جدول ۶. تعداد خانوار، جمع و متوسط اختلاف هزینه و درآمد خانوارهای روستایی موجود در سطوح انتخاب شده با استفاده از دسته‌بندی دانکن به همراه نسبت تجمعی

کد وضع شغلی	تعداد افراد باسواد	متوسط اختلاف هزینه و درآمد	جمع اختلاف هزینه و درآمد	تعداد خانوار	نسبت تجمعی تعداد خانوار به کل	نسبت تجمعی جمع اختلاف به کل
۲	۱۳	-۴۰۹۱۶۲۳۵/۳	-۲۴۵۴۹۷۴۱۱/۹	۶	۰/۰۰۰۱۸۵	۰/۰۰۰۷۲۲
۱	۱۰	-۳۶۷۸۰۴۳۲/۶	-۶۶۲۰۴۷۷۸۷/۳	۱۸	۰/۰۰۰۷۴۲	۰/۰۰۴۶۶۸
۱	۹	-۳۰۰۷۰۴۹۲/۸	-۱۴۱۳۳۱۳۱۶۲	۴۷	۰/۰۰۲۱۹۴	۰/۰۰۶۸۲۲
۱	۱۱	-۲۳۹۵۰۹۴۹/۳	-۱۶۷۶۵۶۶۴۵/۱	۷	۰/۰۰۲۴۱۱	۰/۰۰۷۳۱۵
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
۳	۲	-۸۶۵۴۱۸۹/۹	-۵۸۵۸۸۸۶۵۵۴	۶۷۷	۰/۶۹۷۸۸	۰/۷۹۳۸۴۱

مانند جدول (۵)، جدول (۶) نیز گویای این مطلب است که نه تنها چهار سطح انتخاب شده توسط دسته‌بندی دانکن نمی‌توانند خانوارهایی را با پوشش بالا از کل اختلافات منفی هزینه و درآمد مشخص نمایند بلکه برای رسیدن به پوششی معادل ۸۰٪ لازم است ۷۰٪ از خانوارها انتخاب شوند.

### نتیجه‌گیری

مشخص شد که حتی با استخراج مشخصه‌هایی از خانوار که سطوح آنها از نظر اختلاف منفی هزینه و درآمد تفاوت معنی‌دار دارند، نمی‌توان خانوارهایی را مشخص نمود که سهم اندکی از کل تعداد خانوارها و سهم عمده‌ای از کل اختلافات منفی هزینه و درآمد را به خود اختصاص دهند. در واقع این نتیجه‌گیری گویای این مطلب است که گرچه کم گزارش شدن درآمد مشکلی است که در بعضی از خانوارها نسبت به بقیه بیشتر مشهود است اما نمی‌توان گفت که معضلی است منحصر به یک دسته‌ی خاص از خانوارها. بنابراین برای رفع مشکل اختلاف منفی هزینه و درآمد نمی‌توان تمرکز را روی قشر خاصی از جامعه معطوف

کرد. این که پدیده کمتر بودن درآمدها از هزینه‌ها، منحصر به اقشار خاصی نیست؛ می‌تواند منعکس‌کننده این واقعیت باشد که طرح آمارگیری از هزینه و درآمد خانوار که هر ساله توسط مرکز آمار ایران اجرا می‌شود در بخش درآمد خود نمی‌تواند تمامی درآمدها را پوشش دهد، به عبارتی به علت مطلوب نبودن پرسشهای مربوط به بخش درآمد این طرح قادر به دریافت اطلاعات واقعی درآمد افراد نمی‌باشد. بنابراین به نظر می‌رسد عملی‌ترین راه‌حل برای رفع این مشکل، اصلاح قسمت درآمد پرسشنامه طرح جاری آمارگیری هزینه و درآمد خانوار است.

## پی‌نوشتها:

۱. آمارگیری از هزینه و درآمد خانوارهای شهری، تهران، مرکز آمار ایران، (۸۰-۱۳۷۶).
۲. آمارگیری از هزینه و درآمد خانوارهای روستایی، تهران، مرکز آمار ایران، (۸۰-۱۳۷۶).
۳. ورمزیار، حسن؛ فلاح محسن‌خانی، زهره و جمشیدی، فرشید. ارائه راهکارهای مناسب تولید آمارهای درآمد خانوار، تهران: پژوهشکده آمار، ۱۳۸۳.
4. Christensen, R. *Analysis of Variance, Design and Regression: Applied Statistical Methods*. USA: Chapman & Hall., 1998.
5. Cortina, J. M. and Hossein Nouri. "Effect Size for ANOVA Designs", *Sage Publications. Quantitative Applications in the Social Science Series*, No. 129, (2000).
6. Kotz, Samuel. *Encyclopedia of Statistical Sciences*. John Wiley and Sons Inc, USA, Second Edition, Vol. 1, 2005.
7. Krishnaiah, P. R., Mudholkar, G. S., and Subbaiah, P. *In Analysis of Variance*, Vol 1, Handbook of Statistics, North-Holland, NewYork., 1980.
8. Lindman, H. R. *Analysis of Variance in Complex Experimental Designs*, San Francisco: W. H. Freeman & Co., 1974.
9. Montgomery, D.C. *Desing and Analysis of Experiments*. John Wiley & Sons. Inc, USA., 2000.
10. Nester, J., Kutner, M and Wasserman, W. *Applied Linear Statistical Models*. Mc Graw Hill., 1996.
11. Speed, T. P. W. "Tukey's Contribution to Analysis of Variance"., *The Annals of Statistics*, Vol 30, No. 6, (2002).